ΕΠΙΛΕΓΜΕΝΕΣ ΠΤΥΧΙΑΚΕΣ ΚΑΙ ΔΙΠΛΩΜΑΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

STUDBOOK

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ + ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ΤΟΜΟΣ 19
2022 - 23

STUDENT BOOK

Εκδίδεται μία φορά το χρόνο από το:

Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών,
Πανεπιστημιούπολη, 15784 Αθήνα

Επιμέλεια έκδοσης:

Επιτροπή Ερευνητικών και Αναπτυξιακών Δραστηριοτήτων
Θ. Θεοχάρης (υπεύθυνος έκδοσης), Καθηγητής, Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Η. Μανωλάκος, Καθηγητής, Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Γραφιστική επιμέλεια - Επιμέλεια κειμένων:

Λ. Χαλάτση, ΕΤΕΠ, Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Εικόνα εξωφύλλου και εσωφύλλων: Colorful software coding with computer in abstract background (variations). Όλες οι εικόνες δημιουργήθηκαν από το Midjourney Bot.
Το Midjourney (https://www.midjourney.com) είναι ένα πρόγραμμα τεχνητής νοημοσύνης που δημιουργήθηκε και φιλοξενείται από ένα ανεξάρτητο ερευνητικό εργαστήριο με έδρα το Σαν Φρανσίσκο Midjourney, Inc. Το Midjourney δημιουργεί εικόνες από περιγραφές φυσικής γλώσσας, που ονομάζονται «prompts», παρόμοιες με το DALL-E και το Stable Diffusion του OpenAI.

ISSN 1792-8826

Περιεχόμενα

Πρόλογος

Ο τόμος αυτός περιλαμβάνει περιλήψεις επιλεγμένων διπλωματικών και πτυχιακών εργασιών που εκπονήθηκαν στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών κατά το διάστημα **01/01/2021 - 31/12/2021 και 01/01/2022 - 31/12/2022**. Πρόκειται για τον διπλό **19ο τόμο** στη σειρά αυτή. Στόχος του θεσμού είναι η ενθάρρυνση της δημιουργικής προσπάθειας και η προβολή των πρωτότυπων εργασιών των φοιτητών του Τμήματος.

Η έκδοση αυτή είναι ψηφιακή, έχει δικό της ISSN και αναρτάται στην επίσημη ιστοσελίδα του Τμήματος ώστε να έχει μεγάλη προσβασιμότητα. Για το στόχο αυτό, σημαντική ήταν η συμβολή της Λήδας Χαλάτση που επιμελήθηκε και φέτος την ψηφιακή έκδοση και πέτυχε μια ελκυστική ποιότητα παρουσίασης, ενώ βελτίωσε και την ομοιογένεια των κειμένων.

Η στάθμη των επιλεγμένων εργασιών είναι υψηλή και κάποιες από αυτές έχουν είτε δημοσιευθεί είτε υποβληθεί για δημοσίευση.

Θα θέλαμε να ευχαριστήσουμε τους φοιτητές για το χρόνο που αφιέρωσαν για να παρουσιάσουν τη δουλειά τους στα πλαίσια αυτού του θεσμού και να τους συγχαρούμε για την ποιότητα των εργασιών τους. Ελπίζουμε η διαδικασία αυτή να προσέφερε και στους ίδιους μια εμπειρία που θα τους βοηθήσει στη συνέχεια των σπουδών τους ή της επαγγελματικής τους σταδιοδρομίας.

Η Επιτροπή Ερευνητικών και Αναπτυξιακών Δραστηριοτήτων

Θ. Θεοχάρης (υπεύθυνος έκδοσης), Η. Μανωλάκος

Αθήνα, Απρίλιος 2023

Πτυχιακές Εργασίες

Protection of Sensitive Data: Creating, Analyzing and Testing Protocols of Differential Privacy

Nikolaos G. Galanis

ABSTRACT

The problem of preserving privacy while extracting information during data analysis, has been an everlasting one. Specifically, during the big data era, user details can be easily compromised by a malicious handler, something considered both as a security, and as a privacy issue.

An ideal answer to this problem is finding a balance that would benefit the users and their privacy, as well as researchers who try to collect and analyze data for scientific reasons, while allowing no other party to access the data. The optimal fix to the subject is Differential Privacy: a promise, made by the data handler to the user, that their privacy will not be affected, by allowing their data to be used in any analysis, while the output data statistics should be accurate enough for any researcher to extract useful information from them.

The goal of this thesis was to examine and compare previously created mechanisms for D.P from companies like IBM, Google and tools like ARX, in order to determine if they succeed to keep that promise. Moreover, a result of this thesis was the creation of our own mechanism that serves the purpose of achieving Local D.P., a form of Differential Privacy that is nowadays widely used in machine learning algorithms. Even though the scientific community has come up with lots of similar algorithms, it was discovered that when the number of users used in such a protocol is limited, the accuracy error is extremely large. Thus, a new protocol was constructed and tested in comparison with pre-existing ones, alongside with an easy-to-use library that allows the user to use a plethora of protocols. During this thesis a lot of evaluation of the methods was used, in order to prove the usability and the efficiency of Differential Privacy.

Keywords: Differential Privacy, Security, User data, Data Privacy, Noisy Data, Aggregation of Data

ADVISOR

Chatzikokolakis Kostas, Associate Professor NKUA

1 INTRODUCTION

In general, when we consider the problem of privacy, we refer to the protection of the disclosure of sensitive information of individuals, when a collection of data about these individuals (dataset) is made publicly available.

There are many proposed solutions to the problem, with the main two categories that have been widely spread being privacy via anonymization (which is demonstrated in techniques like k-anonymity), and privacy via randomization. We are going to focus on the latter, given the fact that the first method has been proven to be prone to many different attacks (deanonymization attacks).

The most effective and up-to-date method for applying privacy into a dataset, is via randomization. The method used is called Differential Privacy and is based on injecting noise into the users' data.

Differential Privacy is actually a promise made by the data handlers, to the participants of a study:" You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies/ datasets/ info resources are available"

The goal is to make the data widely available for analysis, while protecting the users. However, is it possible to learn nothing about an individual, while gathering useful information about a population? This is actually what D.P. is trying to achieve.

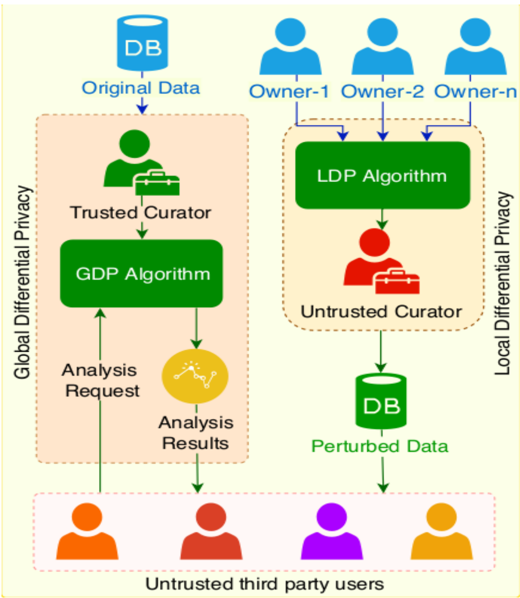
A more formal definition of D.P. is the following:

A randomized algorithm M with domain $N^{|x|}$ is (ε, δ)-differentially private, if for all $S \in Range(n)$ and for all $x, y \backslash in\ N$^{|x|} s.t. $||x-y||_1 \leq 1$

$$Pr[M(x) \in S] \leq e^{\epsilon} Pr[M(y) \in S] + \delta$$

where the probability space is over the coin flips of the mechanisms M. If δ= 0, we say that M is ε-differentially private.

As mentioned during the definition, due to the room that is left for its interpretation, there can be many forms of Differential Privacy. There are two major fields recognized, the Global D.P. and the Local D.P.. Their major difference is the curator of the data. In the Global model, the curator must be trusted, as he collects the non-private data and has to pass them through a D.P. algorithm. On the other hand, in the Local model, the curator may as well be untrusted, since the users perturb their data on their own, using a specific protocol. The key differences of the two forms are shown in the Figure below.



Another difference between the two models is the amount of noise added. With the absence of a trusted curator, the users themselves must add a significant amount of noise into their data, in order to preserve their privacy. This of course results in a need of many users (several thousands), in order for the L.D.P. protocols to function correctly and accurately. In this thesis, we are going to examine both models, by quoting their definitions, observing already existing algorithms, and creating our own L.D.P. protocol.

As every new step in Computer Science, Differential Privacy has some issues that are yet to be solved, and some others not covered by its definition. One major problem is the behavior of the protocols when the number of users is limited. The definition of D.P. is based on the alteration of the data in order not to reveal sensitive information. Thus, if a small number of users are involved in those protocols, the accuracy of the results might be way off the standards that we set, in order to satisfy the epsilon requirements of the user. Another (unsolvable) issue, that mainly lies on the basis of surveys, is the possibility that conclusions drawn from a survey may reflect statistical information about an individual. For example, if a survey about the correlation of smoking and dental problems is conducted, someone that has specific dental problems might be deemed as a smoker, despite keeping his privacy about the fact that he is smoking, during the survey. That is something that D.P. does not promise: unconditional freedom from distinguishing. This is not however a violation of the definition of D.P., as the survey teaches us that specific private attributes correlate with public observable attributes, since this correlation would be observed independent of the presence or absence of the individual in the survey. There are several more issues as the ones covered above, however we are not going to focus on those, rather on the advantages of D.P.

2 EXAMINATION OF PREVIOUSLY-EXISTING GLOBAL DP LIBRARIES

The first goal of this Thesis is to examine previously existing programming libraries and APIs that provide the application of Differential Privacy to a dataset. This has been achieved by many companies, such as Google and IBM, but also from research programs like ARX that study the benefits of data privacy. We separate those implementations regarding their output. The possible outputs of a mechanism that adds D.P. to a dataset can be:

- An answer to a query, in a private manner.
- An anonymized dataset that meets the criteria of D.P.

In the first category, we can distinguish libraries such as Google's and IBM's, that have functions which if applied on a dataset, and given a specific query, can return a single answer. In the second one, we can find libraries such as the ARX tool, that given a dataset and a group of privacy settings (such as the amount of noise to be inserted), produces an anonymized version of a dataset, that has obviously reduced information in comparison to the original one, but is usable by the final user.

We are going to conduct all our testings using noise generated by the Laplace Mechanism, thus we must first define its theoretical behavior. The ℓ1 sensitivity of a query ƒ is defined as following:

$$\Delta f = \max_{\{||x-y||_1 = 1\}} \left||f(x) - f(y)|\right|_1$$

Where x, y ϵ $N^{|X|}$.

This quantity shows the effect by which a single participant's data can change in the worst case during the query ƒ, and thus, the uncertainty that we must insert to the response in order to protect them.

The Laplace Distribution with a scale b, is the distribution with probability density function:

$$Lap(x|b) = \frac{1}{2b} exp\left(-\frac{|x|}{b}\right)$$

whose variance is $\sigma^2 = 2b^2$, and is actually a symmetric version of the exponential distribution.

In order to be of use in our definition, the scale of the noise will be calibrated to the sensitivity of the query ƒ, divided by epsilon. Thus, the noise used will be drawn from

$$Lap\left(\frac{\Delta f}{\epsilon}\right)$$

When applying D.P. mechanisms to our data, we provide the privacy settings of our choice (epsilon variable) and obtain an answer to each of our questions. Thus, in our testings, our goal is to determine the accuracy of the answers, given a specific ε, or some other settings, and comparing them to the true answer, using some metrics. Those metrics are different for each query type. In this section we are going to focus on two types of queries: statistical, and histograms.

In each one of our following testings, we are going to run the query many times. As we already know, D.P. relies on probabilistic algorithms that can sometimes produce extreme results. This may be rare, but we want our testings and conclusions to be accurate. So, we are going to run each query 100 times, and return as a result the mean value of those runs.

In the case of statistical queries, their answer is usually a real number, so in order to check their alteration with the true answer, we are going to consider the absolute difference between the truth and the query answer.

The analytical conduction of the experiments is presented in the full Thesis paper.

3 A LIBRARY FOR LOCAL DIFFERENTIAL PRIVACY

3.1 Definition of Local D.P

As we mentioned in previous chapters, there are two major forms of Differential Privacy. Having analyzed and tested the first one, Global D.P., it is now time to examine Local D.P., by explaining some possible protocols, as well as building our own.

In Local D.P., there is a significant difference compared to Global DP: there is no trusted curator between the data and the users, as they just want to send their data, while already being anonymized. Thus, an algorithm must perturb the data before sending it to the untrusted curator, who will then transmit it to the analysts.

In order to achieve that goal, the user must randomize the value before making it public (i.e., sending it to the untrusted curator). Then, the curator which collects the data (we will reference to him as aggregator moving forward), collects the data and tries to retrieve their original values, with a goal of producing the most accurate results possible. Thus, each LDP algorithm has the following steps:

- Each user encodes, and then perturbs the private value that he wants to make public
- Each user sends out the result of the perturbation process, with that being only the final value, as they keep the intermediate results for themselves
- The untrusted data curator collects each user’s value and implements some kind of aggregation in order to retrieve the stats that he wants from the data given to him.

In comparison with Global D.P., the Local model has advantages, as well as disadvantages. Its main advantages are:

- The user is not forced to trust the data curator, as only the perturbed value is reported
- Simpler implementation of the algorithms, due to the district steps taken by both sides.

while the main disadvantages are the following:

- The noise added should be larger than the Global model, in order to satisfy the definition, thus the number of people in the dataset should be significant for accurate results to be produced.
- Because this is not always possible, many real-world applications use extremely high values of epsilon compared to what we got used to during our testing in the Global models.

During this Thesis, concern was raised for the main disadvantage of L.D.P., and thus we will present a new protocol aiming to reduce the need for many users, while still covering the definition. However, the definition for L.D.P. is quite different than the Global model one's.

We can say that an algorithm A satisfies ε-Local Differential Privacy, if and only if for any input v_1, v_2, we have

$$\forall y \in Range(A)\colon\ r[A(v_1) = y] \leq e^{\epsilon} * Pr[A(v_2) = y]$$

where $Range(A)$ denoted the set of all possible outputs of the algorithm A.

Apart from R.R., several L.D.P. protocols have been implemented during the years, with many of them being widely used by companies in order to protect users' data. One of the most famous protocols is RAPPOR [15], created by Google, and being currently used in the Chrome browser for the company to provide useful info to its users without compromising their privacy. Also, Apple has created its own protocol of L.D.P., and utilize it in their products.

However, we are not going to focus on those protocols moving forward, than the ones presented in [10], a paper which introduces many algorithms for L.D.P., each one with different perturbation techniques and suitable for different circumstances.

3.2 Previously existing Local D.P. protocols

During this chapter we gave a definition of each algorithm, implement it using Python, and compare the accuracy results produced by those protocols, just like during our testings of the G.D.P. models. Each protocol has two parts: the users and the aggregator. For the users we must each time define the following functions:

- *Encode*(): Encodes the true value that the user wants to report
- *Perturb*(): Perturbs the encoded value, in order to produce the random value that will be reported

For the aggregator we must each time define the *Aggregate*() function, that collects the reported random values of the users and produces the results according to the model.

The following protocols are presented in [10], and are called ”pure” protocols, because of the way they aggregate the data produced by the user. For each one of them, we should define a *Support*() function, that indicates for each value of the possible outcomes, the reported values that are supported. Thus, with the notation $\sum_j Support(y_j)$ we mean the sum of all the supported values of the y_{th} element of the dataset.

Also, for a protocol to be pure, two probabilities must be defined, p_* and q_*, where the first notes the probability that the true value is supported by an element y, and the second one the probability of another value is supported by the element y. The protocol is pure if and only if $p_* > q_*$

If a protocol is pure, the estimation of the total reported values for an element of the dataset i, is the following:

$$\text{Estimation} = \frac{\sum_j 1_{support(y^j)}(i) - nq*}{p^* - q^*}$$

where j denotes the j_{th} user reporting their value, and n the total size of the vector of the reported values.

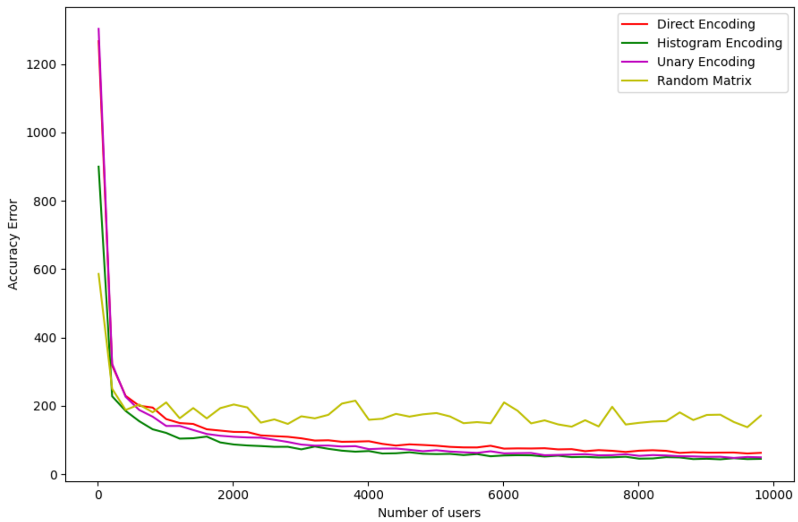
The following pure and non-pure protocols were analyzed and tested in the original thesis:

- Basic RAPPOR (non-pure)
- Extended RAPPOR (non-pure)
- Random Matrix Projection (non-pure)
- Direct Encoding (pure)
- Histogram Encoding (pure)
- Unary Encoding (pure)

The experiment that we will conduct is the accuracy error depending on the number of users used during the survey covered by the protocol. In the definition of L.D.P. the observation of the need of lots of users was made, and it is now time that we examine it. We are going to use a fixed epsilon value, one that our protocols behave similarly for (at least the pure ones, in which we will focus our research moving forward). Our epsilon value that we are going ot used will be fixed and equal to 1.5.

We are going to run the protocols and compare them using the Manhattan Distance. Additionally, we are each time going to divide the result of the metric with the number of users participated, as the simple error is going to increase when the users increase. Hence, this division is going to give us the error depending on the size of our domain. The results are shown in Figure 4.5.

The results confirm the allegations made after explaining the definition of L.D.P.: When the number of participants in a survey is low, the error produced is very high. Every protocol has similar behavior, as we can see that for fewer than 1000 users the relative error is even 6 times larger than for more than 1000 users. Actually, as we can see from the graph, the turning point is around 2000 users: the relative error drops and stabilizes after this number of participants.



That fact triggered our thoughts, on what could possibly be done in order to reduce that problem. The thoughts made on this subject are analyzed in the next chapter, by creating a new L.D.P. protocol, sensitive to the distance between the true and the reported values.

3.3 A distance sensitive Local D.P. encoding protocol

In the previous section we discussed various LDP Protocols that function extremely well for numerical values (histogram type), with accuracy that is completely acceptable. However, when the number of users is limited, (e.g., under 1000), we made the observation that the accuracy error is extremely large. This is mainly due to the fact that the probability of an item to be chosen is independent of the distance between the true and the selected value.

Thus, for the needs of this Thesis, a new L.D.P. protocol was constructed. The idea proposed is to have the probability of choosing an element of the domain to depend on the distance from the true value. This could prove very helpful for histogram values but does not make any sense for categorical values. From now on, we are going to focus on histogram values.

Based on our idea, the probabilities' distribution, in comparison with the distribution of the D.E. protocol, will look like the one in the following figures.

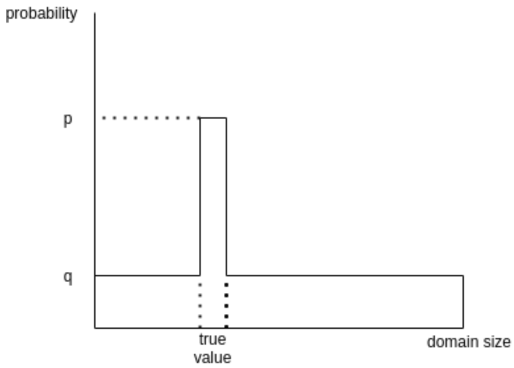


Figure 1: Direct Encoding protocol's probabilistic distribution

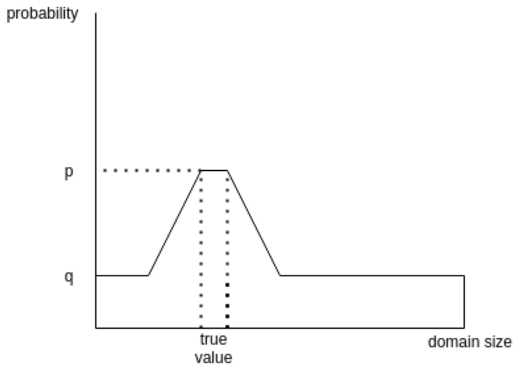


Figure 2: Distance Sensitive protocol's Probabilistic distribution

Thus, there is an area around the true value that has high probability to be selected. The width of this area (from now on θ) is defined by the epsilon setting that the user wants to use. The idea for having a specific area and not decreasing our probabilities as low as it goes when we diverge from the true value, is based on our need to be able to serve low epsilon values, as the first would be a big no for when the users want to use a high privacy setting.

The mathematical background, structure, epsilon requirements and extreme cases of the protocol are presented in detail in the full Thesis paper.

We are now going to define our protocol, by determining the 3 basic operations for an LDP protocol, the encoding, the perturbation and the aggregation methods. We are going to use the following symbols:

- D: The protocol's domain. In this set, we have each i for $1 \leq i \leq |D|$ as each item in the domain D
- a: The quantity that we computed in the previous section
- θ: The constant used in the previous section, which denotes the area around the true value that the probabilities will be higher than others.

Encoding: The encoding procedure is trivial. Just like the Wang paper, we are just going to set:

$$Encode(v) = v$$

for each value v of the domain. The values are going to be randomized during the perturbation step.

Perturbation: Given the previous section, the randomization during the perturbation step is define as following:

$$Pr[Perturb(x) = i] = \begin{cases} p = a & \text{if } i = x \\ q = \frac{a}{|c|(|c|+1)} & c = \min(\theta, |i - x|), \text{ otherwise} \end{cases}$$

Aggregation: The aggregation step was the most tricky during the building of the protocol. A similar approach to the aggregation of pure protocols was chosen, but with a few changes. After several different tries, the optimal aggregation found, was the following: the protocol supports only the reported values corresponding to the true one, thus Support(v) = v. However, the p* quantity is the sum of all the probabilities inside the area:

$$p^* = \sum_{x \in (-\theta, \theta)} p(x)$$

Finally, the q$*$ quantity is the probability of choosing an element from outside the θ area, thus, equal to s. Hence, the estimation generated for a value v of the possible answers in the domain is defined as following:

$$\text{Estimation} = \frac{\sum_j 1_{support(v^j)}(i) - nq^*}{p^* - q^*}$$

The most difficult part of the implementation of our protocol consists of creating the probabilistic distribution for each element of the domain, depending on the true value. This can prove to be costly, if we have a large domain or if we are in the case of the extreme x values.

However, we do not need to compute every single probability, as it is clear from the definition that they are independent from the true value: they only depend on a (and on the domain size in case of an extreme x value). The quantity $|i - x|$ can only take values in the range of $[1, \theta]$, thus constant for every possible true value. Moreover, for the domain values outside of the area, the probability is fixed and equal to a $\frac{a}{\theta(\theta+1)}$. Hence, the probabilities can be computed in advance, either by each user, or given to the protocol by the aggregator.

The protocol has been implemented using Python and can be found in the GitHub repository of this Thesis. Moving forward, we are going to use this implementation in order to conduct some testings to ensure the protocol's functionality.

The most important is how our protocol behaves for an increasing number of users: we must check if it produces better accuracy error than the competitors. This is our next testing, where we are going to set $\varepsilon = ln(20)$, in order for the conditions to be favorable for each one of the protocols. The results are shown in the Figure 3.

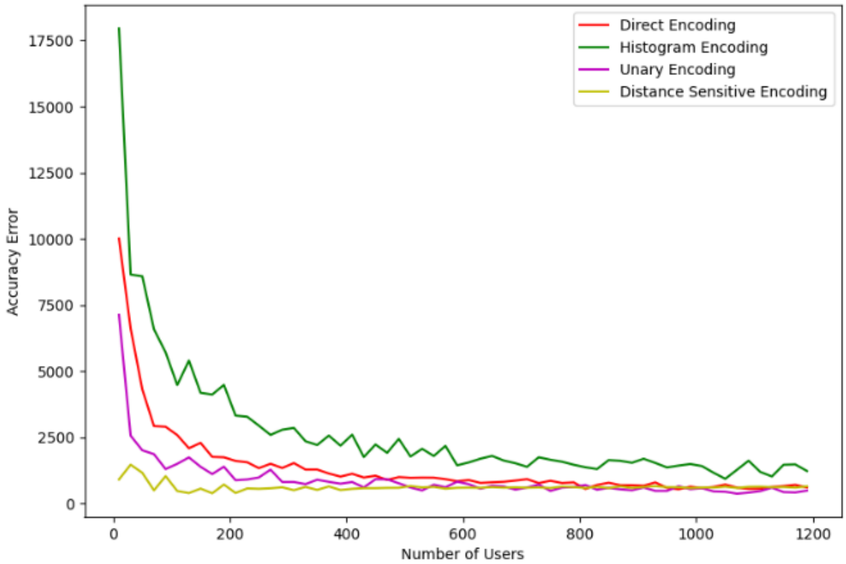


Figure 3: Increasing users measurements for D.S. protocol compared by Kantorovich Distance

For the specific epsilon setting, our protocol produces extremely good accuracy error for a small number of users., beating by a lot the U.E. protocol. The comparisons have been made using the Kantorovich metric, the most characteristic of them all, as it considers the distance between the real answers and the projections, exactly what our protocol is designed to do.

In general, the D.S. protocol succeeds when the number of the participants in a survey is extremely low, and functions similarly with the other protocols for an increasing number of users. The downside is that it does not always takes full advantage of the epsilon setting, as explained in a previous section. However, the results are more than satisfying. Hence, this is a fully functioning protocol that can be used for the application of L.D.P., especially in a situation when few people take part in the survey. The protocol will be further tested in more extreme cases, but this is beyond the scope of this Thesis.

4 CONCLUSIONS AND FUTURE WORK

The goal of this thesis was to analyze the importance of protecting sensitive data and doing so in an efficient way. After its elaboration, it is clear that Differential Privacy is a secure and efficient way for data anonymization. Having two forms, the Global and the Local, it can cover many different scenarios, including Machine Learning applications.

Differential Privacy is the future of Data Protection and Anonymization, as its results cannot be compromised, due to the random noise that the algorithms introduce. Unlike previous methods, such as k-anonymity, there is not yet an attack that can reduce the privacy created by D.P. algorithms, which makes this technique ideal.

Despite the use of random noise, the data is still useful, as the mathematical ideas behind the aggregation were built with the mindset of eliminating this noise using data normalization.

Having explored many different applications, algorithms and protocols we can safely say that when it comes down to Global D.P., IBM's diffprivlib is a state-of-the-art library that produces extremely good results. Its use is quite simple as a Python API is provided, thus can be safely added to any numerical dataset.

When someone wants to apply L.D.P. during a survey, the pure protocols analyzed and tested are suitable for high efficiency combined with good protections of the members. With simple algorithms, they do not require a trusted curator in order to perform, hence users can perturb their data, and then safely report it. However, when the number of users is small, the Distance Sensitive Protocol created for the needs of this Thesis is the best option, as the other protocols produce extreme noise in order to maintain the privacy levels. On the contrary, the D.S. protocol takes into account the distance between the true value and the one being reported when creating its probabilistic space, thus lowering the error produced.

Our plans for future work are centered around the D.S. protocol. We would like to perfect its aggregation method, as it may produce satisfying results, but with a different approach it can maybe become even better. Moreover, we would like to perform more demanding experiments for extreme cases of dataset sizes, domain sizes and theta values.

Finally, similar testings like the ones introduced in this Thesis can be performed in other D.P. libraries, as the accuracy measurements is a good indicator if someone wants to rank those libraries.

REFERENCES

[1] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. now Publishers Inc.

[2] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. Theory of Cryptography, 265–284.

[3] Holohan, N., Braghin, S., Mac Aonghusa, P., & Levacher, K. (2019, July 4). Diffprivlib: The IBM Differential Privacy Library. arXiv.org.

[4] Li, N., Qardaji, W., & Su, D. (2012). On sampling, anonymization, and differential privacy or,k-anonymization meets differential privacy. Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security - ASIACCS '12.

[5] Bild, R., Kuhn, K. A., & Prasser, F. (2018). SafePub: A Truthful Data Anonymization Algorithm With Strong Privacy Guarantees. Proceedings on Privacy Enhancing Technologies, 2018(1), 67–87.

[6] Christofides, T. C. (2003). A generalized randomized response technique. Metrika, 57(2), 195–200.

[7] Chatzikokolakis, K., Palamidessi, C., & Stronati, M. (2015). Location privacy via geo-indistinguishability. ACM SIGLOG News, 2(3), 46–69.

[8] Jain, P., Gyanchandani, M., & Khare, N. (2018). Differential privacy: its technological prescriptive using big data. Journal of Big Data, 5(1).

[9] Bebensee, B. (2019, July 27). Local Differential Privacy: a tutorial. arXiv.org.

[10] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2017. Locally differentially private protocols for frequency estimation. In Proceedings of the 26th USENIX Conference on Security Symposium (SEC'17). USENIX Association, USA, 729–745.

[11] Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., & Palamidessi, C. (2013). Broadening the Scope of Differential Privacy Using Metrics. Privacy Enhancing Technologies, 82–102.

[12] Chamikara, M.A.P. & Bertok, P. & Khalil, Ibrahim & Liu, D. & Camtepe, Seyit. (2019). Local Differential Privacy for Deep Learning.

[13] Chatzikokolakis, K., Fernandes, N., & Palamidessi, C. (2020). Refinement Orders for Quantitative Information Flow and Differential Privacy. Journal of Cybersecurity and Privacy, 1(1), 40–77.

[14] Bassily, R., & Smith, A. (2015). Local, Private, Efficient Protocols for Succinct Histograms. Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing.

[15] Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). RAPPOR. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.

[16] "Surgery Charges Across the U.S.", https://data.world/dmikebishop/surgery-charges-across-the-u-s.

[17] "NBA Salaries", https://data.world/datadavis/nba-salaries

Learning-Augmented Algorithms for Online TSP on the Line

Konstantinos Lakis

ABSTRACT

We study the online Traveling Salesman Problem (TSP) on the line augmented with machine-learned predictions. In the classical problem, there is a stream of requests released over time along the real line. The goal is to minimize the makespan of the algorithm. We distinguish between the open variant and the closed one, in which we additionally require the algorithm to return to the origin after serving all requests. The state of the art is a 1.64-competitive algorithm and a 2.04-competitive algorithm for the closed and open variants, respectively. In both cases, a tight lower bound is known.

In both variants, our primary prediction model involves predicted positions of the requests. We introduce algorithms that (i) obtain a tight 1.5 competitive ratio for the closed variant and a 1.66 competitive ratio for the open variant in the case of perfect predictions, (ii) are robust against unbounded prediction error, and (iii) are smooth, i.e., their performance degrades gracefully as the prediction error increases.

Moreover, we further investigate the learning-augmented setting in the open variant by additionally considering a prediction for the last request served by the optimal offline algorithm. Our algorithm for this enhanced setting obtains a 1.33 competitive ratio with perfect predictions while also being smooth and robust, beating the lower bound of 1.44 we show for our original prediction setting for the open variant. Also, we provide a lower bound of 1.25 for this enhanced setting.

ADVISOR

Panagiotis Stamatopoulos, Professor NKUA

The main part of this thesis was carried out during the student's stay at the Max Planck Institute for Informatics as a research intern, where he was advised by postdoctoral researcher Dr. Themistoklis Gouleakis and collaborated with Ph.D. student Golnoosh Shahkarami.

The result of this work (preprint available at https://arxiv.org/abs/2206.00655) was accepted for presentation at the AAAI 2023 conference.

1 INTRODUCTION

The Traveling Salesman Problem (TSP) is one of the most fundamental and widely studied problems in computer science, both in its offline version (Lawler 1985), where the input is known in advance, and the online version (Ausiello et al. 2001) where it arrives sequentially. In this paper, we consider the online Traveling Salesman Problem (TSP) on the real line. This version of the problem arises in real-world scenarios such as one dimensional delivery/collection tasks. Such tasks include the operation of elevator systems, robotic screwing/welding, parcel collection from massive storage facilities and cargo collection along shorelines (Ascheuer et al. 1999; Psaraftis et al. 1990). An illustrative example, which was also described in (Chen et al. 2019), is the following. Consider a robot that is used in a row of storage shelves in an intelligent warehouse of a large shipping company. This robot is tasked with moving left and right on the aisle to collect the items ordered by customers. However, these orders arrive online, meaning that the release time and location of the item are only revealed at the time of order. The goal is to route the robot such that it collects and returns all the items as soon as possible, minimizing the makespan as it is called. This practically interesting task is indeed captured by the theoretical framework of online TSP on the line. The offline version of the problem can be solved in quadratic time (Bjelde et al. 2021), therefore the difficulty lies in the late availability of the input, such as the locations of the requests. However, the great availability of data as well as the improved computer processing power and machine learning algorithms can make it possible for predictions to be made on these locations (e.g combining information from historical data, events that may affect demand of stored items, etc). In a line of work that started a few years ago (Lykouris and Vassilvtiskii 2018) and sparked a huge interest (Purohit, Svitkina, and Kumar 2018; Antoniadis et al. 2020a; Gollapudi and Panigrahi 2019; Wang and Li 2020; Angelopoulos et al. 2020; Wei 2020; Rohatgi 2020), it has been demonstrated that such prior knowledge about the input of an online algorithm has the potential to achieve improved performance (i.e competitive ratio) compared to known algorithms (or even lower bounds) that do not use (resp. assume the absence of) any kind of prediction. Therefore, it is natural to consider ways to utilize this information in this problem using a so-called learning-augmented approach.

The input to our online algorithm consists of a set of requests, each associated with a position on the real line as well as a release time. An algorithm for this problem faces the task of controlling an agent that starts at the origin and can move with at most unit speed. The agent may serve a request at any time after it is released. The algorithm's objective is to minimize the makespan, which is the total time spent by the agent before serving all requests. We have two different variants of the problem, depending on whether the agent is required to return to the origin after serving all the requests or not. This requirement exists in the *closed* variant, while it does not in the *open* variant. The makespan in the closed variant is the time it takes the agent to serve the requests *and* return to the origin.

We quantify the performance of an online algorithm by its *competitive ratio*, i.e., the maximum ratio of the algorithm's *cost* to that of an optimal *offline* algorithm OPT, over all possible inputs. We say that an algorithm with a competitive ratio of c is c-competitive. Under this scope, the online TSP on the line has been extensively studied and there have been decisive results regarding lower and upper bounds on the competitive ratio for both variants of the problem. Namely, a tight bound of ≈ 1.64 was given for the closed variant, while the corresponding value for the open variant was proven to be ≈ 2.04 (Ausiello et al. 2001; Bjelde et al. 2021). However, no previous work exists for the learning-augmented setting of the problem, and neither does for the learning-augmented setting of any other restriction of the online TSP on general metric spaces. We address this shortcoming in this work.

1.1 Our setup

First of all, to define our prediction model and algorithms, it is necessary to know the number of requests n[1]. This setting shows up in various real world scenarios. For example, in the case of item collection from a horizontal/vertical storage facility, the capacity of the receiving vehicle, which awaits the successful collection of all items in order to deliver them to customers, dictates the number of items to be collected. We note that since n is known, we can assume that each prediction corresponds to a specific request determined by a given labeling, which is shared by both sets (requests and predictions). Under this assumption, we define the LOCATIONS prediction model. In this model, the predictions are estimates for the positions of the requests. The error η increases along with the maximum distance of a predicted location to the actual location of the identically labeled request and is normalized by the length of the smallest interval containing the entire movement of the optimal algorithm. We also define an enhanced prediction model for the open variant named LOCATIONS + FINAL (LF in short) that additionally specifies a request which is predicted to be served last by OPT. In this model, we additionally consider the error metric δ, which increases with the distance of the predicted request to the request actually served last by OPT. We also normalize δ in the same way as η. These models and their respective errors are defined formally in the Preliminaries section.

Properties of learning-augmented algorithms. In the following we formalize the consistency, robustness and smoothness properties. We say that an algorithm is:

1 Since this (slightly) modifies the original problem definition, the previous competitive ratio lower bound results for the classical problem do not necessarily hold for our setup even without predictions. We show in the full version of the paper that the lower bound of 1.64 still holds for the closed variant and that a tight lower bound of 2 holds for the open variant.

1) α-*consistent*, if it is α-*competitive* with no prediction error.
2) β-*robust*, if it is β-*competitive* with any prediction error.
3) γ-*smooth* for a continuous function $\gamma(err)$, if it is $\gamma(err)$-*competitive*, where err is the prediction error.

Note that err could potentially be a tuple of different errors.

In general, if c is the best competitive ratio achievable without predictions, it is desirable to have $\alpha < c$, $\beta \leq k \cdot c$ for some constant k and also the function γ should increase from α to β along with the error err. We note that c,α,β and the outputs of γ may be functions of the input and not constant in that regard.

1.2 Our contributions

Throughout this paper, we give upper and lower bounds for our three different settings (closed variant-$LOCATIONS$, open variant-$LOCATIONS$, and open variant-$LOCATIONS$ + $FINAL$). These settings are deterministic, i.e. the algorithms do not have access to random numbers. We do not consider any randomized settings in this work. The lower bounds refer to the case of perfect predictions and are established via different $attack$ strategies. That is, we describe the actions of an adversary ADV, who can control only the release times of the requests and has the goal of *maximizing* the competitive ratio of any algorithm ALG. We emphasize that ADV is given the power to observe ALG's actions and act accordingly. In more detail, ADV does not need to specify the release times in advance, but can release a request at time t, taking the actions of ALG until time t into account. This is, in fact, the most powerful kind of adversary. The upper bounds are established via our algorithms and are defined for every value of the error(s). Recall that η and δ refer to the two types of error we consider. Our algorithms and attack strategies are intuitively described in their respective sections. We now present the main ideas and our results.

Closed variant under *LOCATIONS*. We will start by intuitively describing our algorithm for this setting and then continue with our lower bound. We design the algorithm $FARFIRST$. The main idea is that we first focus entirely on serving the requests on the side with the furthest extreme, switching to the other side when all such requests are served. When serving the requests on one side, we prioritize them by order of decreasing amplitude. The intuition is that we have the least possible amount of leftover work for our second departure from the origin, which limits the ways in which an adversary may attack us. We obtain the theorem below. More details are given in the ”Closed Variant” section.

Theorem 1: *FARFIRST is* $min\{f(\eta),3\}$*-competitive, where* $f(\eta)$ *is the following function.*

$$f(n) = \frac{3(1+\eta)}{2}$$

We emphasize that for $\eta = 0$, this competitive ratio remarkably matches our lower bound of 1.5, making *FARFIRST* optimal.

Our lower bound for this setting is accomplished via an attack strategy that is analogous to a cunning magician’s trick. Suppose that the magician keeps a coin inside one of their hands. They then ask a pedestrian to make a guess for which hand contains the coin. If the pedestrian succeeds, they get to keep the coin. However, the magician can always make it so that the pedestrian fails, for example by having a coin up each of their sleeves and producing the one not chosen by the pedestrian. One can draw an analogy from this trick to our attack strategy, which is described in the ”Closed Variant” section in more detail.

Theorem 2: *For any* $\epsilon > 0$*, no algorithm can be* $(1.5 - \epsilon)$*-competitive for closed online TSP on the line under the* $LOCATIONS$ *prediction model.*

Open variant under *LOCATIONS*. The algorithm we present for this setting is named *NEARFIRST*. This algorithm first serves the requests on the side opposite to the one *FARFIRST* would choose. Another divergence from *FARFIRST* that should be noted is that for the side focused on second, *NEARFIRST* prioritizes requests that are predicted to be *closer* to the origin, since there is no requirement to return to it, thus avoiding unnecessary backtracking. More details about the algorithm and a proof sketch of the following theorem are given in the relevant section further in the paper, in the section ”Open Variant”.

Theorem 3: *NEARFIRST is* $min\{f(\eta),3\}$*-competitive, for the following function* $f(\eta)$*.*

$$f(\eta) = \begin{cases} 1 + \dfrac{2(1+\eta)}{3-2\eta}, & for\ \eta < \dfrac{2}{3} \\ 3, & for\ \eta \geq \dfrac{2}{3} \end{cases}$$

As in the previous setting, we utilize the "magician's trick" in order to design a similar attack strategy. We describe exactly how this is done in the corresponding section for the open variant under LOCATIONS. This leads to the establishment of a lower bound, as stated below.

Theorem 4: *For any $\epsilon > 0$, no algorithm can be $(1.44$-ε$)$-competitive for open online TSP on the line under the $LOCATIONS$ prediction model.*

Open variant under *LOCATIONS+FINAL*. Our algorithmic approach to this setting is again similar to the one implemented in $NEARFIRST$. The difference is that instead of choosing the side with the near extreme first, we choose the side whose extreme is further away from the predicted endpoint of OPT. We name this algorithm $PIV\ OT$, to emphasize that the prediction for the last request acts as a pivot for the algorithm to decide the first side it will serve. A theorem about $PIV\ OT$ is presented below, for which a proof sketch has been given in the corresponding section.

Theorem 5: *$PIVOT$ is $min\{f(\eta,\delta),3\}$-competitive, for the function $f(\eta,\delta)$ below.*

$$f(\eta, \delta) = \begin{cases} 1 + \dfrac{1+2(\delta+3\eta)}{3-2(\delta+2\eta)}, & 3-2(\delta+2\eta) > 0 \\ 3, & 3-2(\delta+2\eta) \leq 0 \end{cases}$$

For this setting, we reuse the attack strategy initially designed for the closed variant. The only difference is that we add another request at the origin with a release time of 4. We explain how we derive the following theorem in the corresponding section.

Theorem 6: *For any $\epsilon > 0$, no algorithm can be $(1.25 - \epsilon)$-competitive for open online TSP on the line under the LF prediction model.*

1.3 Related work

Online TSP: The online TSP for a general class of metric spaces has been studied in (Ausiello et al. 2001), where the authors show lower bounds of 2 for the open variant and 1.64 for the closed variant. These bounds are actually shown on the real line. Additionally, a 2.5-competitive algorithm and a 2-competitive algorithm are given for the general open and closed variants respectively. A stronger lower bound of 2.04 was shown for the open variant in (Bjelde et al. 2021), where both bounds are also matched in the real line. For the restriction of the closed online TSP to the non-negative part of the real line, (Blom et al. 2001) give a tight 1.5-competitive algorithm. By imposing a fairness restriction on the adversary, they also obtain a 1.28-competitive algorithm. In (Jaillet and Wagner 2006), the authors introduce the ”online TSP with disclosure dates”, where each request may also be communicated to the algorithm before it is released. The authors show improvements to the competitive ratio of various previous algorithms as a function of the difference between disclosure and release dates.

Learning-Augmented algorithms: Learning-Augmented algorithms have received significant attention since the seminal work of (Lykouris and Vassilvtiskii 2018), where they investigated the online caching problem with predictions. Based on that model, (Purohit, Svitkina, and Kumar 2018) proposed algorithms for the ski-rental problem as well as non-clairvoyant scheduling. Subsequently, (Gollapudi and Panigrahi 2019), (Wang and Li 2020), and (Angelopoulos et al. 2020) improved the initial ski-rental problem. The latter also proposed algorithms with predictions for the list update and bin packing problem and demonstrated how to show lower bounds for algorithms with predictions. Several works, including (Rohatgi 2020), (Antoniadis et al. 2020a), and (Wei 2020), improved the initial results regarding the caching problem.

The scheduling problems with machine-learned advice have been extensively studied in the literature. In (Moseley et al. 2020), the makespan minimization problem with restricted assignments was considered, while (Mitzenmacher 2020) used predicted job processing times in different scheduling scenarios. The works of (Bamas et al. 2020) and (Antoniadis, Ganje, and Shahkarami 2021) focused on the online speed scaling problem using predictions for workloads and release times/deadlines, respectively.

There is literature on classical data structures. Examples include the indexing problem, (Kraska et al. 2018), bloom filters, (Mitzenmacher 2018). Further learning-augmented approaches on online selection and matching problems (Antoniadis et al. 2020b; Dutting et al. 2021) and a more general framework of online primal-dual algorithms (Bamas, Maggiori, and¨ Svensson 2020) also emerged, and there is a survey (Mitzenmacher and Vassilvitskii 2020).

Independent Work: Compared to the problem considered in this paper, a more general one, the online metric TSP, as well as a more restricted version in the half-line, have been studied in (Bernardini et al. 2022) under a different setting, concurrently to our work. We note that only the closed variant is considered in (Bernardini et al. 2022). Since the prediction model is different (predictions for the positions as well as release times of the requests are given) and a different error definition is used, the results are incomparable.

2 PRELIMINARIES

The problem definition. In the online TSP on the line, an algorithm controls an agent that can move on the real line with at most unit speed. We have a set $Q = \{q_1,...,q_n\}$ of n requests. We emphasize that for this problem definition, the algorithm receives the value n as input. Each request q has an associated position and release time. To simplify notation, whenever a numerical value is expected from a request q (for a calculation, finding the minimum of a set etc.) the term q will refer to the *position* of the request. Whenever we need the release time of a request, we shall use $rel(q)$. Additionally, the algorithm receives as input a set $P = \{p_1,...,p_n\}$ of predictions regarding the positions of the requests. That is, each p_i attempts to approximate q_i. We assume without loss of generality that Q always contains a request q_0 at the origin with release time 0 and that P contains a perfect prediction $p_0 = 0$ for this request[2].

We use t to quantify time. To describe the position of the agent of an algorithm ALG at time $t \geq 0$, we use $pos_{ALG}(t)$. We may omit this subscript when ALG is clear from context. We can assume without loss of generality that $pos(0) = 0$. The speed limitation of the agent is given formally via $|pos(t') - pos(t)| \leq |t' - t|$, $\forall t,t' \geq 0$. A request q is considered served at time t if $\exists\, t' : pos(t') = q, rel(q) \leq t' \leq t$, i.e., the agent has moved to the request no earlier than it is released. We will say that a request q is *outstanding* at time t, if ALG has not served it by time t, even if $rel(q) > t$, i.e. q has not been released yet. Let t_{serve} denote the first point in time when all requests have been served by the agent. Also, let $|ALG|$ denote the makespan of an algorithm ALG, for either of the two variants. Then, for the open variant $|ALG| = t_{serve}$ while for the closed one $|ALG| = min\{t : pos(t) = 0, t \geq t_{serve}\}$. For any sensible algorithm, this is equivalent to $t_{serve} + |pos(t_{serve})|$, since the algorithm knows the number of requests and will immediately return to the origin after serving the last one. The objective is to minimize the value $|ALG|$, utilizing the predictions.

Notation. We define $L = min(Q)$ and $R = max(Q)$. Recall that Q contains a request at the origin and thus $L \leq 0$ and $R \geq 0$. We refer to each of these requests as an *extreme* request. If $|L| > |R|$, we define $Far = L, Near = R$. Otherwise, $Far = R, Near = L$. That is, Far is the request with the largest distance from the origin out of all requests. Then, $Near$ is simply the other extreme. We will also refer to the value $|q|$ as q's *amplitude*. We will say that a prediction p is (un)released/outstanding/served if the associated request q is (un)released/outstanding/served.

The *LOCATIONS* prediction model. We now introduce the $LOCATIONS$ prediction model. Let $q_1,...,q_n$ be a labeling of the requests in Q. The predictions consist of the values $p_1,...,p_n$, where each p_i attempts to predict the position of q_i.

Error definition for the *LOCATIONS* prediction model. To give an intuition for the metric we will introduce, let us first describe what it means for a prediction to

[2] This can be seen to be without loss of generality by considering a "handler" algorithm ALG_0 which adds this request/prediction pair to *any* input and copies the actions of any of our algorithms ALG for the modified input. We observe that $|OPT|$ is unchanged and $|ALG_0| = |ALG|$.

be bad. In any well-posed definition, the further p_i is from q_i, the worse it should be graded. However, we must also take into account the "scale" of the problem, meaning the length of the interval $[L,R]$ that must be traveled by any algorithm, including OPT. The larger this interval, the more lenient our penalty for p_i should be. Therefore, we define the error as

$$\eta[Q,P] = \frac{max_i\{|q_i - p_i|\}}{|L| + |R|}$$

Additionally, we define $M = \eta \cdot (|L| + |R|)$.

An important lemma for the *LOCATIONS* prediction model. We now present a lemma about this prediction model that will be used widely in our proofs and that contains intuitive value.

Lemma 1: *Let $L_P = min(P), R_P = max(P)$. Then, $|L_P| \geq |R_P|$ implies $|L| \geq |R| - 2M$, and $|R_P| \geq |L_P|$ implies $|R| \geq |L| - 2M$.*

Enhanced prediction model for the open variant. Motivated by the performance of our algorithm under the $LOCATIONS$ prediction model, we enhance it with a prediction f' which attempts to guess the label f of a request on which OPT may finish. We name this new model LF (short for $LOCATIONS + FINAL$). The error η is unchanged. We also introduce a new error metric δ. Let $q_{f'}$ be the request associated with the prediction $p_{f'}$. We then choose q_f to be a request on which OPT may finish that minimizes the distance to $q_{f'}$. We then define the new error as

$$\delta[Q, q_f, q_{f'}] = \frac{|q_{f'} - q_f|}{|L| + |R|}$$

Similarly to before, we define $\Delta = \delta \cdot (|L| + |R|)$.

3 CLOSED VARIANT

In this section, we consider the closed variant under the $LOCATIONS$ prediction model. We provide the $FARFIRST$ algorithm, which obtains a competitive ratio of

1.5 with perfect predictions and is also smooth and robust. Additionally, we give an attack strategy that implies a lower bound of 1.5 for the competitive ratio of any algorithm in this setting, making *FARFIRST* optimal. The formal proofs are deferred to the full version of the paper.

The *FARFIRST* algorithm. Before giving the algorithm, we define the *FARFIRST* ordering on the predictions of an input. For simplicity, we assume that the furthest prediction from the origin is positive. Let $r_1,...,r_a$ be the positive predictions in descending order of amplitude and $l_1,...,l_b$ be the negative predictions ordered in the same way. The *FARFIRST* ordering is $r_1,...,r_a,l_1,...,l_b$. Any predictions on the origin are placed in the end. Ties are broken via an arbitrary label ordering.

We present the algorithm through an update function used whenever a request is released. This update function returns the plan of moves to be executed until the next release of a request. Note that $ext(side,set)$ returns the extreme element of the input set in the side specified, where $side = true$ means the right side. Also, the $\oplus$ symbol is used to join moves one after another. When all the moves are executed, the agent waits for the next release. This only happens when waiting on a prediction.

Algorithm 1: $FARFIRST$ update function.

Input : Current position pos, set O of unserved released requests, first unreleased prediction p in $FARFIRST$ ordering or 0 if none exist, the side $farSide$ with the furthest prediction from the origin.
Output: A series of (unit speed) moves to carry out until the next request is released.
$posSide \leftarrow (pos > 0)$;
$pSide \leftarrow (p > 0)$;
if $pos = 0$ **then** $posSide \leftarrow farSide$;
if $p = 0$ **then** $pSide \leftarrow \overline{posSide}$;
return $move(ext(posSide, O \cup \{pos\})) \oplus move(ext(pSide, O \cup \{p\})) \oplus move(p)$;

In order to give some further intuition on *FARFIRST*, we first give the definition of a *phase*.

Definition 1: *A phase of an algorithm ALG is a time interval* $[t_s,t_e]$ *such that* $pos_{ALG}(t_s) = 0$, $pos_{ALG}(t_e) = 0$ *and* $pos_{ALG}(t') \neq 0$, $\forall\, t' \in (t_s,t_e)$. *That is, ALG starts and ends a phase at the origin and does not cross the origin at any other time during the phase.*

In the following, when we refer to the *far* side, we mean the side with the furthest *prediction* from the origin. The *near* side is the one opposite to that. We see that $FARFIRST$ works in at most three phases. The first phase ends when all predictions on the far side have been released and the agent has managed to return to the origin with no released and outstanding request on the far side. During this phase, any request on the far side is served as long as $FARFIRST$ does not move closer to the origin than the far side's extreme unreleased prediction. Note that some *surprise* requests may appear, i.e., far side requests that were predicted to lie on the near side. These requests are also served in this phase. The second phase lasts while at least one prediction is unreleased. During this phase, the agent serves any request released on the near side, using the predictions as guidance, similarly to the first phase. Requests released on the far side are ignored during this phase. Note that no surprises can occur here, since all far side predictions were released during the first phase. A third phase may exist if some requests were released on the far side during the second phase. These requests' amplitudes are bounded by M, since they were predicted to be positioned on the near side. This simple algorithm is consistent, smooth and robust, as implied by the following theorem.

Theorem 1: *$FARFIRST$ is $min\{f(\eta),3\}$-competitive, where $f(\eta)$ is the following function.*

$$f(n) = \frac{3(1+\eta)}{2}$$

We now give a proof sketch that covers the intuition behind our formal proof. The 3-robustness is seen using an absolute worst case scenario in which $FARFIRST$ is $|OPT|$ units away from the origin at time $|OPT|$ (due to the unit speed limitation), and all the requests to serve are on the opposite side. For the consistency and smoothness, we note that $|OPT| \geq 2(|Near|+|Far|)$. It is therefore sufficient to prove that $|FARFIRST|-|OPT| \leq |Near|+|Far|+3\eta\cdot(|Near|+|Far|) = |Near|+|Far|+ 3M$.

We refer to the left hand side as the *delay* of $FARFIRST$. We now see why this bound holds intuitively. We first describe a worst case scenario. In this scenario, OPT first serves the near side completely, and then does the same for the far side, without stopping. Let t_e denote the end time of the first phase. We see that $t_e \leq |OPT| + M$, because $FARFIRST$ follows the fastest possible route serving the requests on the far side, except for a possible delay of M attributable to a misleading prediction. Note that in this worst case, all requests on the near side must have been released by t_e. Therefore, $FARFIRST$ accumulates an extra delay of at most 2 times the maximum amplitude of these requests. By Lemma 1, this value is at most $|Near|$ + $|Far|$ + $2M$. There are also other possibilities than this worst case, but they also can incur a delay of at most $|Near|$ + $|Far|$ + $3M$, because $|OPT|$ and $|FARFIRST|$ both increase when such cases occur.

A 1.5-attack. We now describe an attack strategy that imposes a lower bound of 1.5 on the competitive ratio of any algorithm ALG. For the sake of exposition, we assume that there is a request on every real number in the interval $[-1,1]$. This is approximated by a limiting process in the formal proof. These requests are released in two phases. The first phase lasts while $pos_{ALG}(t)$ has not exited the interval $[L_U(t), R_U(t)]$, where $L_U(t), R_U(t)$ are the leftmost and rightmost unreleased requests respectively at time t. During this phase, any request with distance d from the origin is released at time $2 - d$. Note that OPT could start serving requests in either side immediately and without stopping during the first phase. This phase ends when $pos_{ALG}(t)$ first exits the aforementioned interval. Assuming without loss of generality that the interval is exited from the left side (which corresponds to choosing the left hand in the magician analogy given in a previous section), the unreleased requests on the left side have their release time delayed to $4 - d$ while the requests on the right side are released as in the first phase (which corresponds to the magician producing the coin in the right hand). Note that OPT can finish by $t = 4$ by moving to 1, then to -1 and then back to the origin with full speed. At the start of the second phase, ALG can either wait for the delayed requests on the side it chose or travel some extra distance to first serve the other side. It turns out that $|ALG|$ can be shown to be arbitrarily close to 6 via the limiting process we mentioned, yielding the theorem below.

Theorem 2: *For any $\epsilon > 0$, no algorithm can be $(1.5 - \epsilon)$-competitive for closed online TSP on the line under the $LOCATIONS$ prediction model.*

4 OPEN VARIANT

In this section, we consider the open variant. We have two prediction models for this variant. The first one is the *LOCATIONS* prediction model and the second is the enhanced *LOCATIONS* + *FINAL* model (*LF* in short). For both settings, we give algorithms and lower bounds.

4.1 The *LOCATIONS* prediction model

Under the *LOCATIONS* prediction model, we design the *NEARFIRST* algorithm, which achieves a competitive ratio of 1.66 with perfect predictions and is also smooth and robust. We complement this result with a lower bound of 1.44 using a similar attack strategy to the one used for the closed variant.

The *NEARFIRST* algorithm. As we mentioned in the introduction, *NEARFIRST* is similar to *FARFIRST* and actually slightly simpler. In essence, *NEARFIRST* simply picks a direction in which it will serve the requests. Then, it just serves the requests either from left to right or from right to left, using the predictions as guidance. The pseudocode for *NEARFIRST* is given below.

Algorithm 2: $NEARFIRST$ update function.

Input : Current position pos, set O of unserved released requests, set P of predictions.
Output: A series of (unit speed) moves to carry out until the next request is released.
$P' \leftarrow$ the unreleased predictions in P;
if P' *is empty* **then**
  if $pos < \frac{max(O)+min(O)}{2}$ **then return** $move(min(O)) \oplus move(max(O))$;
  else return $move(max(O)) \oplus move(min(O))$;
end
if $|min(P)| < |max(P)|$ **then return** $move(min(P' \cup O)) \oplus move(min(P'))$;
else return $move(max(P' \cup O)) \oplus move(max(P'))$;

Recall that $move(x) \oplus move(y)$ is used to indicate a move to x followed by a move to y. We present the following theorem regarding the competitive ratio of *NEARFIRST*.

Theorem 3: *NEARFIRST is* $min\{f(\eta),3\}$*-competitive, for the following function* $f(\eta)$.

$$f(\eta)=\begin{cases}1+\dfrac{2(1+\eta)}{3-2\eta}, & for\ \eta<\dfrac{2}{3}\\ 3, & for\ \eta\geq\dfrac{2}{3}\end{cases}$$

We now give an intuitive proof sketch for this theorem. As in the case of $FARFIRST$, the 3-robustness holds because at time $|OPT|$, $NEARFIRST$ has ”leftover work” of at most $2|OPT|$ time units (to return to the origin and then copy OPT). For the consistency/smoothness, we draw our attention to the request q_f served last by OPT. For the following, we assume that $NEARFIRST$ serves the requests left to right. Let $d = |q_f - R|$. We will show that the delay of $NEARFIRST$ is bounded by $M + d$. Let t_{q_f} be the time when $NEARFIRST$ has served all requests to the left of q_f, including q_f. It turns out that $t_{q_f} \leq |OPT| + M$, because $NEARFIRST$ serves this subset of requests as fast as possible, except for a possible delay of M due to a misleading prediction. Then, in this worst case, $NEARFIRST$ accumulates an extra delay of at most d, proving our claim.

Finally, we bound OPT from below as a function of d. We see that OPT can either serve the requests L,R,q_f in the order L,R,q_f or in the order R,L,q_f. The worst case is the latter, where we see that $|OPT| \geq 2|R| + |L| + (|L| + |R| - d) = 3|R| + 2|L| - d$. Since $d \leq |L| + |R|$, we obtain

$$\frac{|NEARFIRST\ |}{|OPT\ |}=1+\frac{|NEARFIRST\ |-|OPT\ |}{|OPT\ |}\leq 1+\frac{M+|L|+|R|}{2|R|+|L|}$$

Because $NEARFIRST$ considers L the near extreme due to the predictions, by Lemma 1 we find that $|R| \geq \frac{1-2\eta}{2}(|L| + |R|)$, which in turn proves our bound.

A 1.44-attack. The logic of our attack is the same as that used for the attack described in the section for the closed variant. There are two technical differences. The first phase here ends when $pos_{ALG}(t)$ first exits the interval $[3L_U(t)+2, 3R_U(t)-2]$, where $L_U(t), R_U(t)$ are the leftmost and rightmost unreleased request respectively at time t. The other difference lies in the release times of the second phase. We again delay the release times of the requests in the side chosen by ALG, i.e. the side from which the interval was exited. But now, each request with distance d from the origin has its release time delayed to $2 + d$ instead of $4 - d$. Note that OPT can finish by $t = 3$ by first going to the side not chosen by ALG. However, $|ALG|$ can be seen to be arbitrarily close to $4+\frac{1}{3}$, yielding the theorem below.

Theorem 4: *For any $\epsilon > 0$, no algorithm can be $(1.44-\varepsilon)$-competitive for open online TSP on the line under the $LOCATIONS$ prediction model.*

4.2 The *LOCATIONS+FINAL* prediction model

In our final setting we consider the open variant under the LF prediction model. We give the $PIVOT$ algorithm, which is 1.33-competitive with perfect predictions and is also smooth and robust. We also reuse the attack strategy described for the closed variant to achieve a lower bound of 1.25.

The $PIVOT$ algorithm.

The final algorithm we present works in the same way as $NEARFIRST$, except for the order in which it focuses on the two sides of the origin. Instead of heading to the near extreme first, $PIVOT$ prioritizes the side whose extreme is further away from the predicted endpoint of OPT, which is provided by the LF prediction model. The pseudocode for $PIVOT$ is given below. Note that $P_{f'}$ refers to the element in P with label f'.

Algorithm 3: $PIVOT$ update function.

Input : Current position pos, set O of unserved released requests, set P of predictions, label f' of OPT's predicted endpoint.
Output: A series of (unit speed) moves to carry out until the next request is released.

```
P' ← the unreleased predictions in P;
if P' is empty then
    if pos < (max(O)+min(O))/2 then return move(min(O)) ⊕ move(max(O)) ;
    else return move(max(O)) ⊕ move(min(O)) ;
end
if P_f' > (max(P)+min(P))/2 then return move(min(P' ∪ O)) ⊕ move(min(P')) ;
else return move(max(P' ∪ O)) ⊕ move(max(P')) ;
```

As for the previous algorithms, we show a theorem that pertains to $PIVOT$'s competitive ratio for different values of the η and δ errors.

Theorem 5: *$PIVOT$ is $\min\{f(\eta,\delta),3\}$-competitive, for the function $f(\eta,\delta)$ below.*

$$f(\eta,\delta)=\begin{cases}1+\dfrac{1+2(\delta+3\eta)}{3-2(\delta+2\eta)}, & 3-2(\delta+2\eta)>0\\ 3, & 3-2(\delta+2\eta)\le 0\end{cases}$$

We provide a proof sketch of this theorem. The proof is very similar to the one used for $NEARFIRST$'s competitive ratio. In fact, the robustness is shown in exactly the same way. For the consistency/smoothness, the delay is bounded by $M+d$ in the same way, where d is the distance of the last request q_f served by OPT to the extreme served second by $PIVOT$. The same lower bounds for $|OPT|$ hold as well. We additionally bound d as a function of the error-dependent values Δ and M. When there is no error, we can bound d to be at most $\frac{|L|+|R|}{2}$ instead of $|L|+|R|$, which gives a better competitive ratio than that of $NEARFIRST$. An important distinction is that we do not make use of Lemma 1, since the algorithm does not consider the values $|L|$ and $|R|$.

A 1.25-attack. The attack strategy we employ in the current setting is almost the same as the one used for the closed variant. The only difference is that a special request is placed at the origin with a release time of 4. This request is also the last request served by OPT (and thus the optimal solution of the open variant also works for the closed variant) and ALG is informed of this by the LF prediction model. The idea of the proof is that if ALG were to finish before $t=5$, then another algorithm ALG' could solve the closed variant of this input in less than 6 time units, contradicting our first lower bound of 1.5 for the closed variant. This attack strategy implies the following theorem.

Theorem 6: *For any $\epsilon>0$, no algorithm can be $(1.25-\epsilon)$-competitive for open online TSP on the line under the LF prediction model.*

5 CONCLUSION

We have examined the online TSP on the line and provided lower bounds as well as algorithms for three different learning-augmented settings. An immediate extension of our results would be to bridge the gap between the lower and upper bounds we have shown for the open variant. Also, it would be interesting to establish error-dependent lower bounds and/or optimal consistency-robustness tradeoffs. Moreover, an improvement would be to remove the assumption of knowing the number of requests n. A technique that could perhaps allow an algorithm to achieve that is to periodically make sure that the algorithm terminates in case no new requests appear. Another interesting direction is for more general versions of online TSP to be investigated, like the case of trees. Finally, we believe that the combination of learning-augmented techniques along with randomization would lead to much better results, and therefore suggest this direction as future work.

REFERENCES

[1] Angelopoulos, S.; Durr, C.; Jin, S.; Kamali, S.; and Renault, M. P. 2020. Online Computation with Untrusted Advice. In ITCS.

[2] Antoniadis, A.; Coester, C.; Elias, M.; Polak, A.; and Simon, B. 2020a. Online metric algorithms with untrusted predictions.´ In ICML.

[3] Antoniadis, A.; Ganje, P. J.; and Shahkarami, G. 2021. A Novel Prediction Setup for Online Speed-Scaling. CoRR.

[4] Antoniadis, A.; Gouleakis, T.; Kleer, P.; and Kolev, P. 2020b. Secretary and Online Matching Problems with Machine Learned Advice. In NeurIPS.

[5] Ascheuer, N.; Grotschel, M.; Krumke, S. O.; and Rambau, J. 1999. Combinatorial Online Optimization. In¨ Operations Research Proceedings.

[6] Ausiello, G.; Feuerstein, E.; Leonardi, S.; Stougie, L.; and Talamo, M. 2001. Algorithms for the On-Line Travelling Salesman. Algorithmica.

[7] Bamas, E.; Maggiori, A.; Rohwedder, L.; and Svensson, O. 2020. Learning Augmented Energy Minimization via Speed Scaling.´ In NeurIPS.

[8] Bamas, E.; Maggiori, A.; and Svensson, O. 2020. The Primal-Dual method for Learning Augmented Algorithms. In´ NeurIPS.

[9] Bernardini, G.; Lindermayr, A.; Marchetti-Spaccamela, A.; Megow, N.; Stougie, L.; and Sweering, M. 2022. A Universal Error Measure for Input Predictions Applied to Online Graph Problems. CoRR.

[10] Bjelde, A.; Hackfeld, J.; Disser, Y.; Hansknecht, C.; Lipmann, M.; Meißner, J.; SchlOter, M.; Schewior, K.; and Stougie, L.¨ 2021. Tight Bounds for Online TSP on the Line. ACM Trans. Algorithms.

[11] Blom, M.; Krumke, S. O.; De Paepe, W. E.; and Stougie, L. 2001. The Online TSP Against Fair Adversaries. INFORMS journal on computing.

[12] Chen, P.-C.; Demaine, E. D.; Liao, C.-S.; and Wei, H.-T. 2019. Waiting is not easy but worth it: the online TSP on the line revisited. CoRR.

[13] Dutting, P.; Lattanzi, S.; Leme, R. P.; and Vassilvitskii, S. 2021. Secretaries with Advice. In¨ EC.

[14] Gollapudi, S.; and Panigrahi, D. 2019. Online Algorithms for Rent-Or-Buy with Expert Advice. In ICML.

[15] Jaillet, P.; and Wagner, M. R. 2006. Online Routing Problems: Value of Advanced Information as Improved Competitive Ratios. Transportation Science.

[16] Kraska, T.; Beutel, A.; Chi, E. H.; Dean, J.; and Polyzotis, N. 2018. The Case for Learned Index Structures. In SIGMOD.

[17] Lawler, E. 1985. The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization. Wiley-Interscience series in discrete mathematics and optimization. John Wiley & Sons.

[18] Lykouris, T.; and Vassilvtiskii, S. 2018. Competitive caching with machine learned advice. In ICML.

[19] Mitzenmacher, M. 2018. A Model for Learned Bloom Filters and Optimizing by Sandwiching. In NeurIPS.

[20] Mitzenmacher, M. 2020. Scheduling with Predictions and the Price of Misprediction. In ITCS.

[21] Mitzenmacher, M.; and Vassilvitskii, S. 2020. Algorithms with Predictions. In Beyond the Worst-Case Analysis of Algorithms. Cambridge University Press.

[22] Moseley, B.; Vassilvitskii, S.; Lattanzi, S.; and Lavastida, T. 2020. Online Scheduling via Learned Weights. In SODA.

[23] Psaraftis, H. N.; Solomon, M. M.; Magnanti, T. L.; and Kim, T.-U. 1990. Routing and Scheduling on a Shoreline with Release Times. Management Science.

[24] Purohit, M.; Svitkina, Z.; and Kumar, R. 2018. Improving Online Algorithms via ML Predictions. In NeurIPS.

[25] Rohatgi, D. 2020. Near-optimal bounds for online caching with machine learned advice. In SODA.

[26] Wang, S.; and Li, J. 2020. Online Algorithms for Multi-shop Ski Rental with Machine Learned Predictions. In AAMAS.

[27] Wei, A. 2020. Better and Simpler Learning-Augmented Online Caching. In APPROX/RANDOM.

Facial Inpainting Methods for Robust Face Recognition

Vasileios-Marios P. Panagakis

ABSTRACT

The human face is probably the most well-studied object in computer vision, enabling a wide range of applications such as affective computing, biometrics and video analytics, to mention a few examples. For such applications to operate correctly, noise-free facial images are required. This request is far from easy to satisfy in real-world conditions. Occlusions such as eyeglasses, sunglasses and face masks, cause serious corruption to the face images and hinder the performance of face-related applications. This thesis aims to restore severely occluded face images to a non-occluded form to facilitate their identification. To achieve that, we investigate several inpainting models and evaluate them on the face recognition task.

Subject Area: Image Processing, Computer Vision, Deep Learning

Keywords: Image Inpainting, Face Occlusions, Machine Learning, Neural Networks, Robust Principal Component Analysis, Face Recognition

ADVISOR

Yannis Panagakis, Associate Professor NKUA

1 INTRODUCTION

The human face is probably the most well-studied object in computer vision, enabling a wide range of applications such as affective computing, biometrics and video analytics, to mention a few examples. For these applications to be functional complex face recognition systems have been built. Face recognition [12] is a method of identifying or verifying the identity of an individual using their face. Thus, these systems are used to identify people in photos, video, or in real-time. Face recognition systems use computer algorithms to pick out specific, distinctive details about a person's face. These details, such as the distance between the eyes or the shape of the chin, are then converted into a mathematical representation and compared to data on other faces collected in a face recognition database. As expected, face recognition systems vary in their ability to identify people under challenging conditions such as poor lighting, low-quality image resolution, and suboptimal angle of view. However, external conditions are not the only ones that may affect the quality of the face recognition process. Especially, in a real-world scenario, occlusions such as eyeglasses, sunglasses and face masks often hide big parts of human faces causing serious corruption to the face images and making the face recognition task particularly challenging. Although some algorithms can handle face recognition with occlusion, they still suffer from performance degradation due to occlusion's extent. Therefore, the removal of occlusions in face images is a very important, yet challenging task.

This thesis aims to restore severely occluded face images to a non-occluded form to facilitate their identification. To achieve that, we investigate several inpainting models and evaluate them on the face recognition task. The models are based on two principal face inpainting methodologies. The first, supervised method, known as Generative Landmark Guided Face Inpainting (or LaFIn) exploits some of the most innovative and state-of-the-art tools, in the machine learning field, the deep neural networks. LaFIn's architecture benefits from the integration of facial landmarks and accomplishes the desired face restoration. The second, unsupervised method known as Principal Component Pursuit using Side Information, Features and Missing Values (or PCPSFM) is a variation of the famous Robust Principal Component Analysis (RPCA) method. PCPSFM utilizes domain-dependent prior knowledge and manages to recover a low-rank matrix $\mathbf{L_0}$, containing the inpainted face. At the same time, it isolates the occlusions in a separate, sparse matrix $\mathbf{S_0}$.

The contributions of this thesis can be summarized as follows:

- A new face image dataset is created containing only cropped and aligned, en face images, which cover large gender, age and ethnicity variations. The dataset embeds both clear and occluded face images. It is pointed out how occlusions of different shapes, sizes and realistic colors can be generated around significant facial landmark locations.
- Furthermore, we demonstrate the applicability of both supervised and unsupervised face inpainting methodologies with both quantitative and qualitative evaluation. Thus, we not only manage to generate highly realistic inpainted images but also, achieve high accuracy results on the face recognition task.
- We introduce our own novel evaluator, VGGFace2 Classifier, to validate the accuracy of our inpainting models for the face recognition task. In fact, VGGFace2 Classifier outperforms long-established evaluators in the aforementioned task.

1.1 Related Work

Deep learning-based methods are a major group of supervised methods, which deal with the image inpainting problem. The context encoder [5], which is treated as a pioneer deep-learning method for image completion, introduces an encoder-decoder network trained with an adversarial loss [6]. After that, plenty of follow-ups have been proposed to improve the performance from various aspects. For instance, the scheme in [7] employs both global and local discriminators to accomplish the task. Deep face inpainting methods are another significant group of supervised methods. Specific to face completion, the authors of [8] construct a loss, which takes care of the gap in semantic segmentation (face parsing), between the inpainted face images and the ground truth ones, expecting to achieve better preservation of the face structure. However, this work often suffers from color inconsistency and is unable to process, effectively faces with large poses. Facial landmarks are better to act as guidance, thanks to their neatness, sufficiency, and robustness to reflect the structure of the face. Many works, such as [9, 10] have successfully applied landmarks to the task of face generation.

Concerning the unsupervised inpainting methods, the traditional methods of the category consist of two main representative branches, the diffusion-based and patch-based approaches. Diffusion-based approaches [1, 2] propagate low-level features around the occluded areas, in an iterative way. However, these methods are only effective in reforming regions without structure. At the same time, patch-based methods [3, 4] attempt to copy similar blocks from either the same image or a set of images to the target regions. The computational cost of calculating the similarity between their blocks is expensive, even though some works like [3] have been proposed, to accelerate the procedure. Another group of unsupervised inpainting methods is called the non-blind inpainting methods. Those techniques fill in the occluded part of an image using the pixels around the missing region. Exemplar-based techniques that cheaply and effectively generate new texture by sampling and copying color values from the source are widely used. In paper [19], a non-blind inpainting method suggests a unified scheme to determine the fill order of the target region, using an exemplar-based texture synthesis technique. The confidence value of each pixel and image isophotes are combined to determine the priority of filling. [20] presents an image inpainting technique to remove occluded pixels when the occlusion is small.

2 LAFIN: GENERATIVE LANDMARK GUIDED FACE INPAINTING

Generative Landmark Guided Face Inpaintor (LaFIn) [11], is a deep network built to carry out the face inpainting problem. LaFIn comprises a Landmark Prediction

Module and an Image Inpainting Module, trained on the CelebA and Celeba-HQ datasets [20].

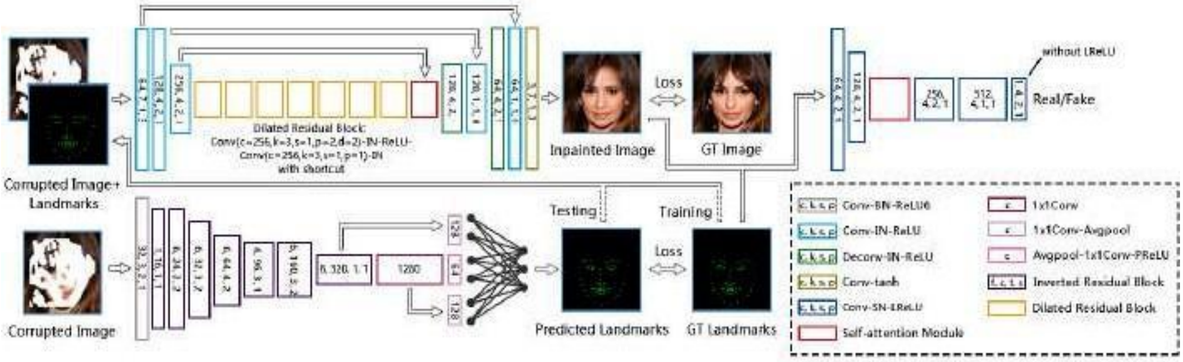


Figure 1: LaFIn architecture. At first, the Landmark prediction module estimates the landmarks and then the Inpainting module applies them to the corrupted image [11]

2.1 Landmark Prediction Module

The goal of the Landmark Prediction Module is to retrieve a set of 68 landmarks from a corrupted face image. For the face inpainting task, we are more concerned about getting landmarks that can accurately identify the face structure and its basic attributes, like pose and expression, rather than finding the precise location of each unique face landmark. The reason behind this simplification is that most of the in-between landmarks don't offer important information, concerning the face inpainting task. LaFIn’s Landmark Prediction Module follows the same architecture as most of the preexistent landmark detectors (see [16, 17]). Specifically, it is built upon the MobileNet-V2 model, proposed in “Mobilenetv2: Inverted residuals and linear bottlenecks” [18] and focuses on feature extraction.

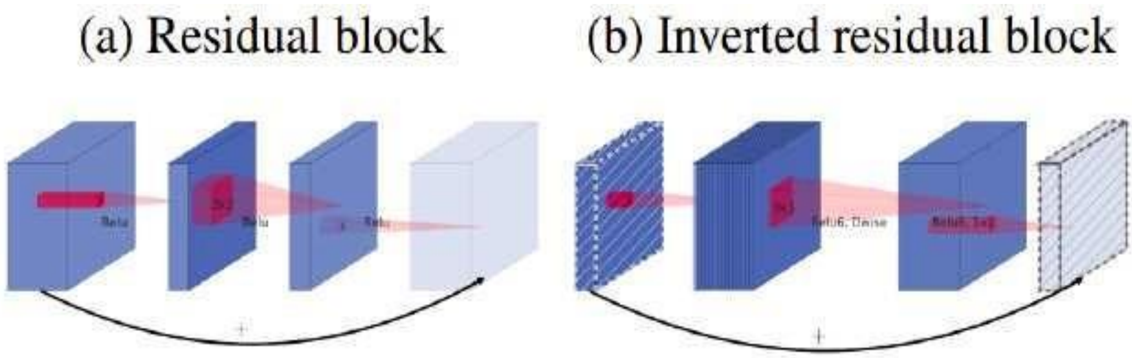


Figure 2: Evolution of residual blocks. (a) MobileNets: Residual block, (b) MobileNet-V2: Inverted residual block [18]

2.2 Image Inpainting Module

The purpose of the Image Inpainting Module is to restore faces by taking occluded images and their predicted landmarks. LaFIn's authors use the famous Generative Adversarial Network (GAN) architecture, proposed in "Generative Adversarial Nets" [6] by I. Goodfellow et al, to build the inpainting module. GAN uses two networks, called the generator and the discriminator. Based on insights from game theory, GAN's training objective can be considered as a mini-max game, where the generator needs to produce fake, realistic images conditioned on the landmarks, out of a known prior distribution, to fool the discriminator. At the same time, the discriminator needs to distinguish between the real and the generated fake images. The networks are trained in a competitive adversarial manner. The convergence is reached when the generated results are not distinguishable from the real ones.

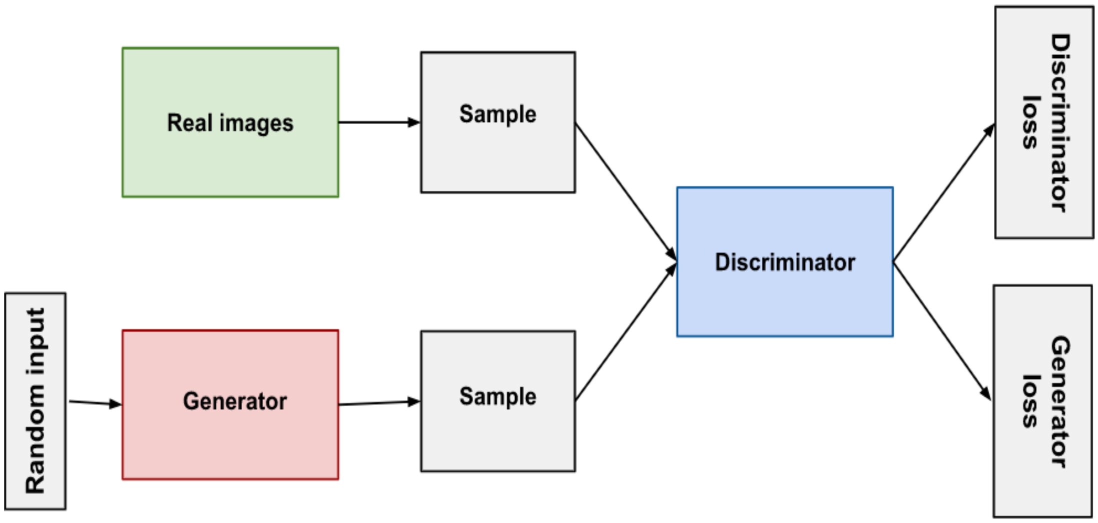


Figure 3: GAN schema

3 ROBUST PRINCIPAL COMPONENT ANALYSIS USING SIDE INFORMATION

3.1 Problem Definition

Suppose a data matrix **M** is given, which can be decomposed as

$$\mathbf{M} = \mathbf{L_0} + \mathbf{S_0}, \quad (1)$$

where $\mathbf{L_0}$ is a low-rank matrix and $\mathbf{S_0}$ is a sparse matrix. In this case, both components are of arbitrary magnitude. Neither the low-dimensional column and row space of $\mathbf{L_0}$, nor their dimension is known. The locations of the non-zero entries of $\mathbf{S_0}$ and their values are unknown, as well. The goal is to recover accurately or even exactly the low-rank and sparse components, in an efficient manner. In the case of the face inpainting problem, matrix **M** contains aligned face images, stacked as column vectors. The face images can be both occluded and non-occluded. The recovered $\mathbf{L_0}$ matrix contains all the inpainted face images in the form of column vectors, while $\mathbf{S_0}$ matrix consists of images, depicting the occluded parts of the initial faces, stacked in a corresponding format.

3.2 Problem Variation

A new problem is set to replace the initial one aiming to deal with the case of huge corruption. Similarly to problem (1), in the new problem, the purpose is to recover a low-rank matrix $\mathbf{L_0}$ from highly corrupted measurements. The entries in $\mathbf{S_0}$ can have arbitrarily large magnitude and their support is assumed to be sparse, but unknown. E. Candes et al. showed in “Robust Principal Component Analysis?” [24], that not only this problem can be solved, but it can be solved by tractable convex optimization.

Let $||\mathbf{M}||_* := \sum_i \sigma_i(\mathbf{M})$ denote the nuclear norm of the matrix **M**, i.e., the sum of the singular values of **M**, and let $||\mathbf{M}||_1 = \sum_{ij}(|\mathbf{M}_{ij}|)$ denote the l_1-norm of **M** seen as a long vector in $R^{n1\times n2}$. The authors of [24] ended up showing that under rather weak assumptions, the Principal Component Pursuit (PCP) solves (2) and recovers exactly the low-rank $\mathbf{L_0}$ and the sparse $\mathbf{S_0}$.

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{L}||_* + \lambda||\mathbf{S}||_1 \qquad (2) \\ &\text{subject to} \quad \mathbf{M} = \mathbf{L} + \mathbf{S}. \end{aligned}$$

3.3 Robust Principal Component Analysis using Side Information, Features and Missing Values

Over the years, many variants have been proposed, trying to confront the convex PCP problem (2) more efficiently for several different applications, including background modeling from surveillance video and removing shadows or specularities from face images. Two important variants were presented in [21] and [22].

Principal Component Pursuit with Features (PCPF) method in [21] assumes that there are available orthogonal column spaces $\mathbf{U} \in \mathrm{R}^{n_1 \times d_1}$, where $d_1 \leq n_1$ and row spaces $\mathbf{V} \in \mathrm{R}^{n_2 \times d_2}$, where $d_2 \leq n_2$, with the following objective:

$$\begin{aligned} \text{minimize} \quad & ||\mathbf{H}||_* + \lambda ||\mathbf{E}||_1 \qquad (3) \\ \text{subject to} \quad & \mathbf{X} = \mathbf{U}\mathbf{H}\mathbf{V}^T + \mathbf{E}, \end{aligned}$$

where $\mathbf{H} \in \mathrm{R}^{d_1 \times d_2}$ is a bilinear mapping for the recovered low-rank matrix $\mathbf{L} \in \mathrm{R}^{d_1 \times d_2}$, with rank $r \ll min(n_1, n_2)$ and $\mathbf{E} \in \mathrm{R}^{n_1 \times n_2}$ is a sparse matrix with entries of arbitrary magnitude. The main drawback of this model is that features need to be accurate and noiseless, which is not trivial in practical scenarios.

In the case of missing data, the robust matrix recovery method [22] enhances PCP to deal with occlusions:

$$\begin{aligned} \text{minimize} \quad & ||\mathbf{L}||_* + \lambda ||\mathbf{W} \circ \mathbf{E}||_1 \qquad (4) \\ \text{subject to} \quad & \mathbf{X} = \mathbf{L} + \mathbf{E}, \end{aligned}$$

where $\mathbf{W}$ is the matrix of binary occlusion masks and $\mathbf{A} \circ \mathbf{B}$ symbolizes the element-wise multiplication of two matrices of the same dimension. The method's Jacobi-type update schemes can be implemented in parallel and hence are attractive for solving large-scale problems.

For the purposes of this thesis, we are going to experiment with the Robust Principal Component Pursuit using Side information, Features and Missing values (PCPSFM), proposed in "Side Information for Face Completion: A Robust PCA Approach'' [15]. This work introduces a novel convex program to use side information, which is a noisy approximation of the low-rank component, within the PCP framework. Moreover, the suggested method can handle missing values, while the developed optimization algorithm grants better convergence rates. Last but not least, the introduced model can use side information to exploit prior knowledge regarding the column and row spaces of the low-rank component, expanding even more the potential of the algorithm.

At first, the authors of [15] presented a PCPSM model, which uses side information with missing values. For (5) to be valid, they set as a precondition that a noisy estimate of the low-rank component of the data $\mathbf{S} \in \mathrm{R}^{n_1 \times n_2}$ must be available.

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{L}||_* + \alpha||\mathbf{L}\text{-}\mathbf{S}||_* + \lambda||\mathbf{W} \circ \mathbf{E}||_1 \qquad (5) \\ &\text{subject to} \quad \mathbf{X} = \mathbf{L} + \mathbf{E}, \end{aligned}$$

where $\alpha > 0$, $\lambda > 0$ are parameters that weigh the effects of side information and noise sparsity.

Then, they utilized their proposed PCPSM model to generalize PCPF (3), which led to the introduction of the novel PCPSFM model, using side information, features and missing values.

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{H}||_* + \alpha||\mathbf{H}\text{-}\mathbf{D}||_* + \lambda||\mathbf{W} \circ \mathbf{E}||_1 \qquad (6) \\ &\text{subject to} \quad \mathbf{X} = \mathbf{UHV}^T + \mathbf{E},\ \mathbf{D} = \mathbf{U}^\mathbf{T}\mathbf{SV}, \end{aligned}$$

where $\mathbf{H} \in \mathrm{R}^{d_1 \times d_2}$, $\mathbf{D} \in \mathrm{R}^{d_1 \times d_2}$ are bilinear mappings for the recovered low-rank matrix $\mathbf{L}$ and side information $\mathbf{S}$ respectively. The low-rank matrix $\mathbf{L}$ is recovered from the optimal solution ($\mathbf{H}^*$, $\mathbf{E}^*$) to objective (6) via $\mathbf{L} = \mathbf{UH}^*\mathbf{V}^T$

The authors chose the multi-block Alternating Direction Method of Multipliers (ADMM) to deal with the convex problem (6). ADMM operates by carrying out repeated cycles of updates until it converges. In this particular problem, a small number of iterations is of great necessity, because of the high computation cost that occurs from certain steps of the algorithm. The solution of (6) is summarized in the following Algorithm, where S_τ: R → R denotes the shrinkage operator $S_\tau(x) = sgn(x)\, max(|x| - \tau, 0)$, which naturally extends to matrices, $S_\tau(\mathbf{A})$ by applying it to matrix **A** element-wise. Similarly, $D_\tau(\mathbf{A})$ denotes the singular value thresholding operator given by $D_\tau(\mathbf{A}) = \mathbf{U}S_\tau(\Sigma)\mathbf{V}^T$, where $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ is the SVD of **A**.

.

Algorithm ADMM solver for PCPSFM

Input: Observation **X**, mask **W**, side information **S**, features **U**, **V**, parameters α, $\lambda > 0$, scaling ratio $\beta > 1$.

1: **Initialize:** $\mathbf{Z} = 0$, $\mathbf{N} = \mathbf{B} = \mathbf{H} = 0$, $\beta = \frac{1}{||X||_2}$.

2: **while** not converged **do**

3: $\mathbf{E} = S_{\lambda\mu^{-1}}(\mathbf{X} - \mathbf{UHV}^T + \frac{1}{\mu}\mathbf{Z}) \circ \mathbf{W} + (\mathbf{X} - \mathbf{UHV}^T + \frac{1}{\mu}\mathbf{Z}) \circ (\mathbf{1} - \mathbf{W})$

4: $\mathbf{H} = \mathbf{U}^T D_{\frac{1}{2\mu}}(\frac{1}{2}(\mathbf{X} - \mathbf{E} + \frac{1}{\mu}\mathbf{Z}) + \mathbf{U}(\mathbf{B} + \mathbf{U}^T\mathbf{SV} - \frac{1}{\mu}\mathbf{N})\mathbf{V}^T))\mathbf{V}$

5: $\mathbf{B} = D_{\alpha\mu^{-1}}(\mathbf{H} - \mathbf{U}^T\mathbf{SV} + \frac{1}{\mu}\mathbf{N})$

6: $\mathbf{Z} = \mathbf{Z} + \mu(\mathbf{X} - \mathbf{E} - \mathbf{UHV}^T)$

7: $\mathbf{N} = \mathbf{N} + \mu(\mathbf{H} - \mathbf{B} - \mathbf{U}^T\mathbf{SV})$

8: $\mu = \mu \times \beta$

9: **end while**

Return: $\mathbf{L} = \mathbf{UHV}^T$, $\mathbf{E}$.

4 EXPERIMENTAL EVALUATION AND DISCUSSION

For our experiments, a part of the CelebFaces Attributes Dataset (CelebA) [27] was used, following the required processing to match the needs of the face inpainting task. CelebA is a large-scale face attributes dataset with 202,599 RGB celebrity face images, including 10,177 person identities. The images in this dataset cover large pose variations and background clutter. Concerning the number of images chosen from the CelebA dataset, it’s clear we couldn’t work with all of the 202,599 images, not only due to our restricted resources but also because of the long execution time required for each one of the multiple algorithm executions. Apart from the quantity of the images, their quality is a crucial factor as well. For instance, we could not use en profile faces for the purposes of this project, where facial landmarks are not distinct, since the methods we used have to memorize as better as possible the basic structure and attributes of the human face, to effectively restore the occluded parts during the test procedure. For the reasons above, we picked out and processed 1600 images of 100 celebrity identities. The images are distributed equally, meaning there are 16 images per identity, including 1 manually occluded image.

As mentioned above, our dataset contains one occluded face per identity. So, for the implementation of this project, we applied 100 different occlusions, given that there are 100 identities located in our dataset. For the creation of the occlusions, we made use of the OpenCV python library. One of its uses is to design shapes of different sizes and colors on a given image. Thus, we tried to exploit OpenCV to create a variety of occlusion combinations of different shapes (rectangles, circles, triangles, lines) and sizes (small, medium, big), filled with realistic colors (shades of red, brown, pink) that could be detected in an actual face occlusion (e.g., a face trauma). For each one of the occluded images, our goal was to hide at least one of the significant facial landmarks, to make the inpainting task as challenging as possible for our models.

Figure 4: Sample of our dataset grouped by identity

4.1 Inpainting Results

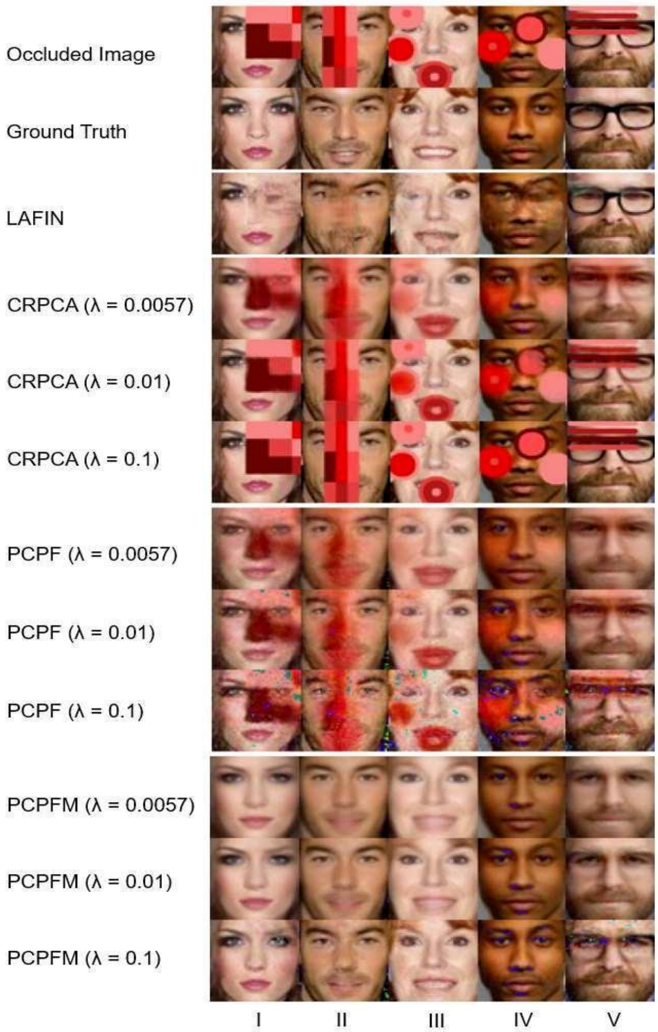


Figure 5: Image Inpainting on medium occlusions

To quantify and evaluate the quality of the inpainting results we applied a performance indicator in the form of a Reconstruction Error metric. This metric measures the distance between an inpainted image and the ground truth image and returns a value, which represents the deviation between the two images. It's clear that the smaller the error value, the greater the similarity of the images. The reconstruction error is denoted as follows

$$\mathrm{RE} = \left(\frac{||\mathbf{GT}-\mathbf{INP}||_F}{||\mathbf{GT}||_F}\right)^2 \qquad (7)$$

where **GT** is the array representation of the ground truth image and **Inp** is the array representation of an inpainted image. The array values lie in the range of [0,1].

For the evaluation of our inpainting results, we deployed the Mean Reconstruction Error (MRE), meaning we calculated the average reconstruction error for the 100 inpainted images produced by each model.

Table 1: Mean Reconstruction Error on medium occlusion dataset

Models		Mean Reconstruction Error
LAFIN		0.037
CRPCA	λ = 0.0057	0.076
	λ = 0.01	0.104
	λ = 0.1	0.127
PCPF	λ = 0.0057	0.066
	λ = 0.01	0.080
	λ = 0.1	0.115
PCPFM	λ = 0.0057	0.024
	λ = 0.01	**0.020**
	λ = 0.1	0.021

4.2 Evaluation on Face Recognition

Reaching the end of this thesis, it has become clear by now, that the purpose of the face inpainting task is the restoration of occluded faces to a non-occluded form, to facilitate their identification. Though, we mustn't be complacent, that a visually flawless inpainted image guarantees the accurate retrieval of the identity. Hence, for the purposes of this project, we deployed three different evaluators (K-Nearest Neighbors Classifier, Linear Support Vector Machine Classifier, VGGFace2 Classifier) to validate the accuracy of our inpainting models for the face recognition task. The K-Nearest Neighbors (KNN) and Linear Support Vector Machine (SVM) Classifiers are based on the homonymous supervised machine learning algorithms, that are mostly used to solve classification problems. KNN assumes that similar things exist in close proximity. At first, it stores all the initial data (train set) and then classifies the new data points (test set), based on their similarity to the initial data. This means when a new data point appears it can be easily classified into the most similar of the available categories. Linear SVM's main attribute is the creation of a line or a hyperplane, which separates the data into classes. In the SVM algorithm, the initial data (train set) are mapped as points in n-dimensional space with their values being the coordinates of their locations. The objective of SVM is to maximize the width of the gap between the two classes. The new data points (test set) are then mapped into that same space, and they join the class, which corresponds to the side of the hyperplane they fall into. During our experiments, we used the KNeighborsClassifier and SVC Classifier implementations of the Scikit-learn Python package. Moreover, we generated the face encodings using the face_recognition Python library, which constructs an array of 128 values. Face encodings are arrays of RGB values, containing specialized information about a face image.

On the other hand, VGGFace2 Classifier is an implementation of our own. Practically, we built a KNN Classifier, named after the VGGFace2 dataset. VGGFace2 is made of around 3.31 million images divided into 9131 classes, each representing a different person's identity. Thanks to its low label noise, high pose and age diversity, it has become a popular dataset suitable to train state-of-the-art deep learning models on face-related tasks. We named our Classifier after

VGGFace2, because this time, we chose to generate the face encodings differently. Specifically, we deployed a ResNet50 [23] neural network pre-trained on VGGFace2, which takes an image face as input and returns an array of 2048 values, representing the unique face attributes, namely the face encodings. Afterward, we passed the face encodings in our KNN Classifier, where we utilized the cosine similarity metric to classify our inpainted results to the predicted identities.

Table 2: Evaluation results on medium occlusions using KNN Classifier

Models		Exact accuracy (%)	Ranked-3 accuracy (%)	Ranked-5 accuracy (%)
LAFIN		66	71	80
CRPCA	$\lambda = 0.0057$	58	69	73
	$\lambda = 0.01$	68	72	77
	$\lambda = 0.1$	62	73	80
PCPF	$\lambda = 0.0057$	42	48	53
	$\lambda = 0.01$	45	55	60
	$\lambda = 0.1$	33	43	49
PCPFM	$\lambda = 0.0057$	48	62	67
	$\lambda = 0.01$	69	77	84
	$\lambda = 0.1$	**76**	**83**	**85**

Table 3: Evaluation results on medium occlusions using Linear SVM Classifier

Models		Exact accuracy (%)
LAFIN		60
CRPCA	$\lambda = 0.0057$	33
	$\lambda = 0.01$	68
	$\lambda = 0.1$	67
PCPF	$\lambda = 0.0057$	40
	$\lambda = 0.01$	44
	$\lambda = 0.1$	32
PCPFM	$\lambda = 0.0057$	61
	$\lambda = 0.01$	69
	$\lambda = 0.1$	**76**

Table 4: Evaluation results on medium occlusions using VGGFace2 Classifier

Models		Exact accuracy (%)	Ranked-3 accuracy (%)	Ranked-5 accuracy (%)
LAFIN		75	82	85
CRPCA	$\lambda = 0.0057$	54	65	72
	$\lambda = 0.01$	71	81	85
	$\lambda = 0.1$	75	82	87
PCPF	$\lambda = 0.0057$	37	54	61
	$\lambda = 0.01$	46	56	65
	$\lambda = 0.1$	43	55	61
PCPFM	$\lambda = 0.0057$	46	62	70
	$\lambda = 0.01$	67	79	84
	$\lambda = 0.1$	**76**	**84**	**88**

5 CONCLUSION AND FUTURE WORK

The purpose of this thesis was to restore several occluded face images to a non-occluded form, using different inpainting methods, which we evaluated on the face recognition task. LaFIn is a supervised method built upon a series of deep neural networks, which integrate facial landmarks and generate an inpainting result for a given occluded face image. PCPSFM, on the other hand, is an unsupervised method, based on the RPCA methodology. It is fed with both occluded and non-occluded face images and returns a low-rank matrix $\mathbf{L_0}$, containing all the inpainted face images and a sparse matrix $\mathbf{S_0}$, which consists of images, depicting the occluded parts of the initial faces. To achieve that, PCPSFM incorporates many improvements, compared to the classic RPCA method, with the most important of them being side information, features and missing values.

Through the course of this thesis, we managed to produce several successful inpainting results, and we achieved very satisfying evaluation scores on the face recognition task, always in proportion to the examined occlusion size. However, there are still many upgrades we could have included in our experiments to optimize even more the inpainting results and the evaluation accuracy if we had the appropriate resources and a loose deadline. First of all, having access to a stable machine with an integrated GPU would allow us to focus on a more complex implementation of the PCPSFM Algorithm, designed to utilize the processing power of the GPU, which would lead to a notable reduction of PCPSFM's execution time. In this context, it would be in our best interest to process and experiment with a lot more than 1600 face images, to feed our models with as much data as possible, expecting a significant improvement in the inpainting results. In fact, the face images could be of a resolution higher than 100 × 100 to expose in even more detail the restoration of the occluded face parts. Moreover, taking for granted the availability of resources in the future, it may be worth proceeding to the re-training of LaFIn network, on the same cropped and aligned set of clear face images, used as side information or features in the PCPSFM models. This way, the comparison of the two methods will be fair, counter to the circumstances of this project, where we made use of the default LaFIn version, pre-trained on the initial in-the-wild, aligned CelebA images. Last but not least, there is always a chance of finding a better evaluation method, capable of achieving higher accuracy scores for the face recognition task.

REFERENCES

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. “Image inpainting”. In 27th Annual Conference on Computer Graphics and Interactive Techniques, pages 417–424, 2000.

[2] H. Yamauchi, J. Haber, and H.-P. Seidel. “Image restoration using multiresolution texture synthesis and image inpainting”. In Computer Graphics International, pages 120–125. IEEE, 2003.

[3] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: “A randomized correspondence algorithm for structural image editing”. ACM Transactions on Graphics (ToG), 28(3):24:1–24:11, 2009.

[4] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf. “Image completion using planar structure guidance”. ACM Transactions on graphics (TOG), 33(4):129:1–129:10, 2014.

[5] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. “Context encoders: Feature learning by inpainting”. In CVPR, pages 2536–2544, 2016.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. “Generative adversarial nets”. In NeurlPS, pages 2672–2680, 2014.

[7] S. Iizuka, E. Simo-Serra, and H. Ishikawa. “Globally and locally consistent image completion”. ACM Transactions on Graphics (ToG), 36(4):107, 2017.

[8] Y. Li, S. Liu, J. Yang, and M.-H. Yang. “Generative face completion”. In CVPR, pages 3911–3919, 2017.

[9] J. Zhang, X. Zeng, Y. Pan, Y. Liu, Y. Ding, and C. Fan. “Faceswapnet: Landmark guided many-to-many face reenactment”. arXiv preprint arXiv:1905.11805, 2019.

[10] Q. Sun, L. Ma, S. Joon Oh, L. Van Gool, B. Schiele, and M. Fritz. “Natural and effective obfuscation by head inpainting”. In CVPR, pages 5050–5059, 2018.

[11] Y. Yang, X. Guo, J. Ma, L. Ma, H. Ling. “LaFIn: Generative Landmark Guided Face Inpainting”. arXiv:1911.11394. GitHub: https://github.com/YaN9-Y/lafin.

[12] Electronic Frontier Foundation, July 1990. https://www.eff.org/pages/face-recognition. [Accessed 2/9/21]

[13] A. Criminisi, P. Perez, and K. Toyama. “Region filling and object removal by exemplar-based image inpainting”. IEEE Transactions on image processing, 13(9):1200–1212, 2004.

[14] L. Cheng, J. Wang, Y. Gong, and Q. Hou. “Robust deep auto-encoder for occluded face recognition”. In Proceedings of the 23rd ACM international conference on Multimedia, pages 1099–1102. ACM, 2015.

[15] N. Xue, J. Deng, S. Cheng, Y. Panagakis, S. Zafeiriou. “Side Information for Face Completion: A Robust PCA Approach”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(10):2349–2364, 2019.

[16] A. Kumar and R. Chellappa. “Disentangling 3d pose in a dendritic CNN for unconstrained 2d face alignment". In CVPR, pages 430–439, 2018.

[17] S. Xiao, J. Feng, L. Liu, X. Nie, W. Wang, S. Yang and A. Kassim. “Recurrent 3d-2d dual learning for large-pose facial landmark detection”. In CVPR, pages 1633–1642, 2017.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In CVPR, pages 4510–4520, 2018.

[19] Z. Liu, P. Luo X. Wang, X. Tang, “Large-scale CelebFaces Attributes (CelebA) Dataset”, 29 July 2016; https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html. [Accessed 6/9/2021]

[20] O. Ronneberger, P. Fischer, T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation”. arXiv: 1505.04597.

[21] K. Chiang, C. Hsieh, and I. Dhillon, “Robust principal component analysis with side information”. In ICML, 2016.

[22] F. Shang, Y. Liu, J. Cheng, and H. Cheng, “Robust principal component analysis with missing data”. In CIKM, pages 1149–1158, 2014.

[23] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition”. In CVPR, 2016.

[24] E. J. Candes, X. Li, Y. Ma, J. Wright. “Robust Principal Component Analysis?” arXiv:0912.3599.

Οι Τεχνολογίες Πληροφορίας & Επικοινωνίας στην κοινωνική ένταξη του προσφυγικού πληθυσμού: Δημιουργία ψηφιακού εργαλείου για την πρόσβαση στις e-Services της Υπηρεσίας Ασύλου

Ελένη Πολυξένη Γ. Παπαδάκη

ΠΕΡΙΛΗΨΗ

Οι κοινωνικοπολιτικές και υγειονομικές μεταβολές σε ολόκληρο τον κόσμο, συνδυαστικά με την ταχεία τεχνολογική πρόοδο, έχουν οδηγήσει τα τελευταία χρόνια σε ραγδαία ψηφιοποίηση υπηρεσιών και διαδικασιών. Συγχρόνως, η Ελλάδα, είναι μία χώρα, η οποία αποτελεί σημείο εισόδου στην ευρωπαϊκή ήπειρο και παραδοσιακά υποδέχεται μεγάλα ρεύματα προσφυγικών ροών από τρίτες χώρες που αιτούνται άσυλο σε αυτήν. Οι υπηρεσίες που συνδέονται με το νομικό καθεστώς αυτής της πληθυσμιακής μερίδας υπάγονται στην κατηγορία των υπηρεσιών που έχουν ψηφιοποιηθεί, χωρίς ωστόσο να λαμβάνουν υπόψη τα ιδιαίτερα χαρακτηριστικά και τις ανάγκες των ανθρώπων στους οποίους απευθύνονται. Στο πλαίσιο της παρούσας πτυχιακής εργασίας, μέσω της αξιοποίησης των Τεχνολογιών Πληροφορίας και Επικοινωνίας (ΤΠΕ) δημιουργείται ένα ψηφιακό εργαλείο που καλύπτει την ανάγκη αυτή, παρέχοντας στην ομάδα στόχου τη δυνατότητα πρόσβασης σε έγκυρη πληροφόρηση και μεθόδους ανάπτυξης των απαραίτητων ψηφιακών δεξιοτήτων (Διαθέσιμο στο σύνδεσμο: https://hrcmfcathens.com/digital-tool-asylum/)

Θεματική Περιοχή: Ψηφιακό εκπαιδευτικό περιβάλλον

Λέξεις-Κλειδιά: Ψηφιακές δεξιότητες, ΤΠΕ, H5P, λογισμικό καθοδηγούμενης διδασκαλίας, προσομοίωση, πρόσφυγες

ΕΠΙΒΛΕΠΩΝ

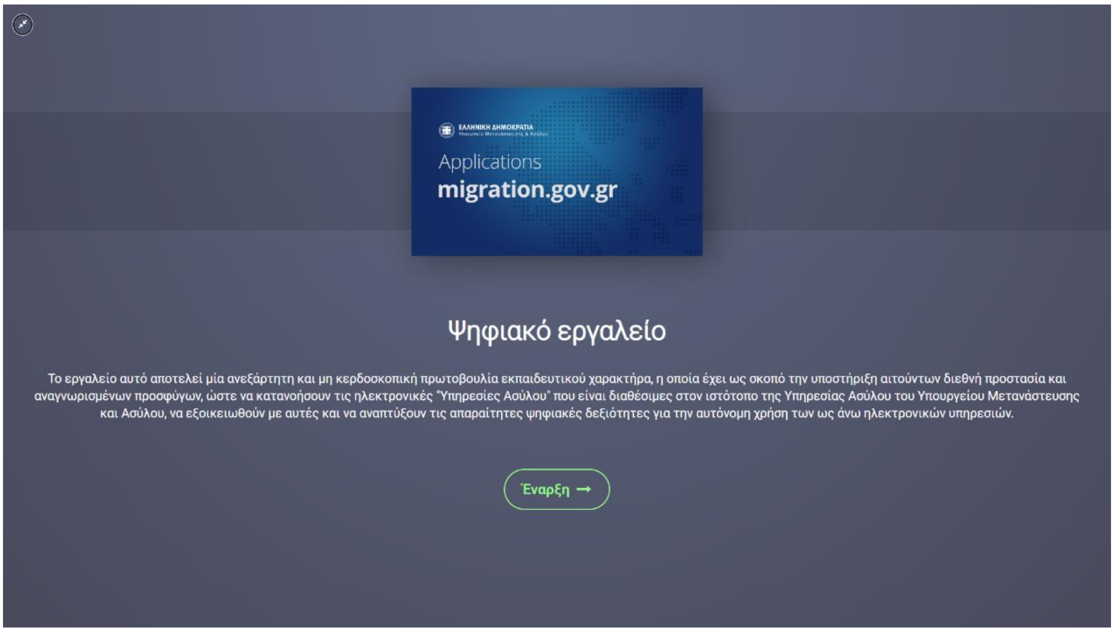
Αγορίτσα Γόγουλου, ΕΔΙΠ, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

1 ΕΙΣΑΓΩΓΗ

Η Ελλάδα ανέκαθεν αποτελούσε ένα από τα κύρια σημεία εισόδου προσφυγικών ροών λόγω της ευνοϊκής γεωπολιτικής της θέσης. Συνεπώς, έχει αναπτύξει υπηρεσίες και διαδικασίες που απευθύνονται σε αυτή την πληθυσμιακή ομάδα και σχετίζονται άμεσα με το νομικό καθεστώς υπό το οποίο παραμένουν στην ελληνική επικράτεια.

Οι κοινωνικοπολιτικές και υγειονομικές εξελίξεις των τελευταίων χρόνων, έχουν οδηγήσει στην αξιοποίηση των ραγδαίων τεχνολογικών εξελίξεων και των Τεχνολογιών Πληροφορίας και Επικοινωνίας στις κοινωνίες υποδοχής, με σκοπό τη διαμόρφωση συνθηκών ηλεκτρονικής διακυβέρνησης. Στο πλαίσιο αυτό, βασικές υπηρεσίες που απευθύνονται σε προσφυγικό πληθυσμό, προχώρησαν σε άμεση ψηφιοποίηση των διαδικασιών τους, προκειμένου να εναρμονιστούν με τις απαιτήσεις της εποχής. Ωστόσο, εκ του αποτελέσματος, είναι έκδηλο ότι οι ενέργειες αυτές γέννησαν και αρκετές προκλήσεις, καθώς η πληθυσμιακή ομάδα των προσφύγων αντιμετωπίζει εμπόδια και δυσκολίες στην προσπάθειά της να αποκτήσει πρόσβαση στις εν λόγω υπηρεσίες που σχετίζονται με γλωσσικούς παράγοντες, αδυναμία πρόσβασης σε έγκυρη και ολοκληρωμένη πληροφόρηση και έλλειψη βασικών ψηφιακών δεξιοτήτων. Η κατανόηση, η γνώση και η πρακτική αξιοποίηση όλων αυτών των στοιχείων δύνανται να προσφέρουν χρήσιμα εργαλεία ως στοχευμένες παρεμβάσεις για την αποτελεσματική διαχείριση των ανωτέρω προκλήσεων και την υποστήριξη του προσφυγικού πληθυσμού.

Ως εκ τούτου, η επικρατούσα κατάσταση στην Ελλάδα και οι προκλήσεις που ανακύπτουν στο πλαίσιο της αυτόνομης και αποτελεσματικής πρόσβασης του προσφυγικού πληθυσμού στις ψηφιοποιημένες υπηρεσίες, λειτούργησαν ως έναυσμα ώστε να εκπονηθεί η παρούσα πτυχιακή εργασία, η οποία εστιάζει στη μελέτη της εν λόγω πληθυσμιακής ομάδας και των διαδικασιών που την αφορούν, καθώς επίσης της αξιοποίησης των Τεχνολογιών Πληροφορίας και Επικοινωνίας για την υποστήριξη της ενταξιακής διαδικασίας των προσφύγων στην κοινωνία υποδοχής. Με βάση το θεωρητικό αυτό υπόβαθρο, δημιουργήθηκε ένα ψηφιακό μαθησιακό εργαλείο που απευθύνεται στην πληθυσμιακή ομάδα των προσφύγων και έχει εστιάσει το περιεχόμενό του στην πρόσβαση αυτής στις ηλεκτρονικές υπηρεσίες ασύλου (Εικόνα 1).



Εικόνα 1: Αρχική σελίδα ψηφιακού εργαλείου

Ο σκοπός της παρούσας πτυχιακής εργασίας συνίσταται στα κάτωθι:

- Η αιτιολογημένη επιλογή της ομάδας στόχου, μέσα από την ανάλυση των διαδικασιών ασύλου και των ιδιαίτερων χαρακτηριστικών και αναγκών τους σε συνάφεια με την ψηφιοποίηση της κοινωνίας υποδοχής
- Η θεωρητική παρουσίαση των Τεχνολογιών Πληροφορίας και Επικοινωνίας (ΤΠΕ) και της σύνδεσης αυτών με την ψηφιοποίηση υπηρεσιών και διαδικασιών σε εθνικό και ευρωπαϊκό επίπεδο

- Η δημιουργία ενός ψηφιακού μαθησιακού εργαλείου καθοδήγησης και ανάπτυξης ψηφιακών δεξιοτήτων για την αποτελεσματική και αυτόνομη πρόσβαση της ομάδας στόχου στις ψηφιοποιημένες υπηρεσίες ασύλου, μέσα από την κατάκτηση από τον χρήστη των ακόλουθων προσδοκώμενων αποτελεσμάτων:
    - Να πλοηγείται σε ένα περιβάλλον ψηφιακών υπηρεσιών και να εντοπίζει και να επιλέγει τις επιθυμητές υπηρεσίες
    - Να εντοπίζει τα υπομνήματα βοήθειας που παρέχονται σε μία ηλεκτρονική φόρμα
    - Να διαχειρίζεται και να συμπληρώνει μία ηλεκτρονική φόρμα
    - Να εφαρμόζει τα βήματα της προσομοίωσης στην πλατφόρμα της Υπηρεσίας Ασύλου (ΥΑ) και να υποβάλλει με επιτυχία την εκάστοτε αίτηση
    - Να πειραματίζεται στη χρήση της πλατφόρμας της ΥΑ μέσω των προσομοιώσεων
    - Να αναγνωρίζει τα πλεονεκτήματα της ψηφιοποίησης των γραφειοκρατικών διαδικασιών, να προσπαθεί αυτόνομα να χρησιμοποιεί τις αντίστοιχες ψηφιακές πλατφόρμες και να παροτρύνει άλλους στη χρήση αυτών
    - Να προσαρμόζεται στις ψηφιακές απαιτήσεις της κοινωνίας υποδοχής και να τις αντιμετωπίζει με αυτοπεποίθηση και όχι με φόβο

1.1 Η επικρατούσα κατάσταση στην Ελλάδα

Η πανδημία του COVID-19 έφερε αδιαμφισβήτητα ραγδαίες αλλαγές σε όλους τους τομείς της ανθρώπινης δραστηριότητας. Στο πλαίσιο αυτό, έφερε στο προσκήνιο την ανάγκη αξιοποίησης των ΤΠΕ ώστε να εξασφαλιστεί η ανεμπόδιστη πρόσβαση όλων των πολιτών σε υπηρεσίες για την ικανοποίηση των αναγκών τους και τη διεκπεραίωση θεμάτων της καθημερινότητάς τους [1]. Οι υπηρεσίες του δημοσίου τομέα στην Ελλάδα έχουν ενταχθεί σε μια διαδικασία ψηφιακού μετασχηματισμού προκειμένου να προσαρμόσουν λειτουργίες και παροχές στα σύγχρονα δεδομένα ενός ψηφιοποιημένου κράτους (Πρόγραμμα «Πολιτεία» [2], Ενιαία Ψηφιακή Πύλη της Δημόσιας Διοίκησης (gov.gr) [3] κ.ά.).

Υπό το πρίσμα αυτό, το Υπουργείο Ψηφιακής Διακυβέρνησης, αναγνωρίζοντας την αναγκαιότητα απόκτησης ψηφιακών δεξιοτήτων, ώστε όλοι οι πολίτες να ανταποκρίνονται με αποτελεσματικότητα στις απαιτήσεις της ψηφιοποιημένης παροχής υπηρεσιών, σύστησε την ομάδα «Εθνικό Κέντρο Ανάπτυξης Ψηφιακών Δεξιοτήτων/Ικανοτήτων», καθώς και τη Βίβλο του Ψηφιακού Μετασχηματισμού που περιέχει σχετικές στρατηγικές ως προς την ανάπτυξη δραστηριοτήτων για την εκπαίδευση πολιτών και επιχειρηματιών στις ΤΠΕ [4].

Παράλληλα, η Εθνική Στρατηγική του Υπουργείου Μετανάστευσης και Ασύλου για την ένταξη των προσφύγων [5] δίνει επίσης έμφαση στην αξιοποίηση των ΤΠΕ και στην ανάπτυξη ψηφιακών δεξιοτήτων για τη διευκόλυνση της ενταξιακής τους διαδικασίας μέσα από τη διασφάλιση της πρόσβασής τους σε δημόσιες υπηρεσίες και σε περιεκτικές και ακριβείς πληροφορίες σχετικά με τα δικαιώματα και τις υποχρεώσεις τους.

1.2 Η επικρατούσα κατάσταση στην Ευρωπαϊκή Ένωση

Η Ευρωπαϊκή Ένωση (ΕΕ) ορίζει την ψηφιακή ικανότητα ως μία από τις οκτώ βασικές ικανότητες τις οποίες χρειάζονται όλοι οι άνθρωποι για την προσωπική τους ολοκλήρωση και ανάπτυξη, την ενεργό ιδιότητα του πολίτη, την κοινωνική ένταξη και την απασχόληση. Το Ευρωπαϊκό Κοινοβούλιο, μέσα από σχετικό ψήφισμα, προβλέπει τις ψηφιακές δεξιότητες ως απαραίτητο στοιχείο για μια ψηφιακή κοινωνία χωρίς αποκλεισμούς, στην οποία όλοι οι πολίτες ανεξαιρέτως έχουν την υποχρέωση και λαμβάνουν τα κατάλληλα κίνητρα προκειμένου να τις αναπτύξουν [6]. Στο πλαίσιο αυτό, ο Ευρωπαϊκός Ιστότοπος για την Ένταξη (EWSI) διερεύνησε το εύρος της ψηφιοποίησης των υπηρεσιών ένταξης μεταναστών σε ολόκληρη την ΕΕ κατά τη διάρκεια της πανδημίας COVID-19, ορίζοντας ως θεματικές εστίασης την προσαρμογή των υπηρεσιών που απαιτούν φυσική παρουσία, τη δημιουργία νέων διαδικτυακών υπηρεσιών και την ψηφιακή προσβασιμότητα για τον μεταναστευτικό πληθυσμό.

Μέσα από τη δημοσιοποίηση των σχετικών ευρημάτων τον Φεβρουάριο 2022, διαπιστώθηκε ότι οι δημόσιες υποστηρικτικές υπηρεσίες τοπικού και εθνικού επιπέδου προσαρμόστηκαν, ώστε να λειτουργήσουν διαδικτυακά σε ολόκληρη την ΕΕ, για να αντιμετωπίσουν μεταξύ άλλων τη μοναξιά και την απομόνωση που προκλήθηκαν λόγω της εφαρμογής των περιοριστικών μέτρων για τον COVID-19. Πολλές κυβερνήσεις της ΕΕ βρίσκονταν ήδη στη διαδικασία ψηφιοποίησης του συνόλου των υπηρεσιών τους, ενώ για ορισμένους διεθνείς οργανισμούς η ηλεκτρονική παροχή υπηρεσιών αποτελούσε προϋπάρχον χαρακτηριστικό τους. Για άλλες κυβερνήσεις και μικρότερους οργανισμούς, η πανδημία του COVID-19 αποτέλεσε έναυσμα για τη μετάβαση από την παραδοσιακή στην ψηφιακή παροχή υπηρεσιών. Ωστόσο, υπήρχαν επιπλέον κυβερνήσεις ή/και οργανισμοί που εξακολουθούσαν να μην έχουν τους απαιτούμενους πόρους για ψηφιοποίηση των παρεχόμενων υπηρεσιών.

Ως επί το πλείστον, οι δυτικές, οι βόρειες και οι μεσογειακές χώρες, όπως η Γαλλία, η Πορτογαλία, η Σουηδία, η Δανία και η Μάλτα τείνουν να έχουν υψηλότερο επίπεδο παροχής διαδικτυακών υπηρεσιών για τον προσφυγικό πληθυσμό συγκριτικά με τις χώρες της Ανατολικής και Κεντρικής Ευρώπης, όπως η Σλοβακία, η Ρουμανία, η Ουγγαρία και η Βουλγαρία, συνθήκη η οποία ίσχυε και πριν από την πανδημία του COVID-19.

Ανεξάρτητα από το βαθμό ψηφιοποίησης των παρεχόμενων υπηρεσιών στις χώρες της ΕΕ, όπως προκύπτει μέσα από τη μελέτη ξενόγλωσσης βιβλιογραφίας σχετικά με τον σχεδιασμό και την ανάπτυξη ψηφιακών εργαλείων για την υποστήριξη της πρόσβασης του εν λόγω πληθυσμού σε ψηφιακές υπηρεσίες, εντοπίζονται μεμονωμένες πρωτοβουλίες που συνδέονται κυρίως με την κατάρτιση σχεδίων ένταξης, την αξιολόγηση αναγκών επαγγελματικής εκπαίδευσης και την παροχή πληροφοριών για την πρόσβαση σε υπηρεσίες ένταξης και διοικητικές διαδικασίες, οι οποίες σε ορισμένες περιπτώσεις αποτέλεσαν προσωρινή συνθήκη κατά τη διάρκεια των εθνικών απαγορεύσεων κυκλοφορίας λόγω της πανδημίας του COVID-19, ενώ σε άλλες διαπιστώθηκε έλλειψη πολυγλωσσικού χαρακτήρα [7].

1.3 Πρόσβαση του προσφυγικού πληθυσμού σε ηλεκτρονικές υπηρεσίες

Στην Ελλάδα, από το 2020, στο πλαίσιο προώθησης της ψηφιακής διακυβέρνησης με σκοπό την κάλυψη των αναγκών της κοινωνίας σε επίπεδο διαδικασιών και υπηρεσιών εν μέσω της πανδημίας του COVID-19, το Υπουργείο Μετανάστευσης και Ασύλου έχει ανακαλέσει την απαραίτητη αυτοπρόσωπη παρουσία των αιτούντων άσυλο στις κατά τόπους αρμόδιες Περιφερειακές Υπηρεσίες, οπότε οι διαδικασίες της Υπηρεσίας Ασύλου υποστηρίζονται αποκλειστικά ηλεκτρονικά. Συγκεκριμένα, έχει διαμορφωθεί ηλεκτρονική πλατφόρμα μέσω της οποίας οι καταγεγραμμένοι αλλοδαποί έχουν πρόσβαση σε ένα σύνολο από ηλεκτρονικές αιτήσεις (e-Services). Παράλληλα, για τις διαδικασίες που απαιτούν τη δέσμευση ραντεβού, είναι απαραίτητη η ηλεκτρονική κράτηση αυτού μέσα από τη σχετική δυνατότητα που παρέχεται στην πλατφόρμα.

Οι αιτούντες ή/και δικαιούχοι διεθνούς προστασίας μπορούν να υποβάλλουν ηλεκτρονικά αιτήσεις οι οποίες εντάσσονται στις κάτωθι κατηγορίες των Υπηρεσιών Ασύλου [8]:

- Προγραμματισμός ραντεβού
- Ανανέωση Δελτίων Αιτούντων Διεθνούς Προστασίας
- Αίτηση Αλλαγής Στοιχείων Επικοινωνίας
- Αίτηση Αλλαγής Στοιχείων Ταυτότητας
- Αίτηση Αναβολής/Επίσπευσης Συνέντευξης
- Αίτηση Βεβαίωσης Κατάστασης Αιτήματος
- Αίτηση Διαχωρισμού Φακέλων
- Αίτηση Κατάθεσης Εγγράφων
- Αίτηση Χορήγησης Αντιγράφων
- Αίτηση για Νομική Συνδρομή
- Αίτηση γνωστοποίησης ΠΑΑΥΠΑ
- Έλεγχος απόδοσης ΑΦΜ

Εντούτοις, η ψηφιακή πρόσβαση παραμένει ιδιαίτερα δύσκολη για τους νεοαφιχθέντες υπηκόους τρίτων χωρών και για ευάλωτες ομάδες προσφύγων [9]. Τα βασικά εμπόδια για τους υπηκόους τρίτων χωρών στην πρόσβαση σε υπηρεσίες ηλεκτρονικής υποστήριξης περιλαμβάνουν γλωσσικά εμπόδια και έλλειψη στα κάτωθι: σύνδεση στο διαδίκτυο, τεχνολογικά μέσα, ηλεκτρονική ταυτότητα (e-ID), ψηφιακές δεξιότητες (συμπεριλαμβανομένης της εξοικείωσης με συγκεκριμένους τύπους τεχνολογίας), χρόνο και χώρο στο σπίτι [7][10].

Παράλληλα, πρόσφατη αξιολόγηση αναγκών του προσφυγικού πληθυσμού που διεξήγαγε ο Ελληνικός Ερυθρός Σταυρός (ΕΕΣ) σε συνεργασία με τη Διεθνή Ομοσπονδία Ερυθρού Σταυρού Ερυθράς Ημισελήνου (ΔΟΕΣ/ΕΗ) [11] ανέδειξε ότι η περιορισμένη ή παντελής έλλειψη γνώσης της ελληνικής γλώσσας αποτελεί βασικό εμπόδιο για την απόκτηση πρόσβασης σε αγαθά και υπηρεσίες και την εξασφάλιση διαμονής και απασχόλησης. Επιπρόσθετα, μεγάλη πλειοψηφία του ως άνω πληθυσμού βιώνει θλίψη, απομόνωση και άγχος λόγω του μεγάλου χρόνου αναμονής και της μη κατανόησης των διαδικασιών που αφορούν στο νομικό τους καθεστώς.

Ειδικότερα σε σχέση με το τελευταίο, αναδείχθηκε ότι η πρόσβαση στις διαδικασίες ασύλου είναι εξαιρετικά δύσκολη λόγω ασάφειας και του γεγονότος ότι η ανωτέρω πλατφόρμα ηλεκτρονικών υπηρεσιών της Υπηρεσίας Ασύλου είναι κυρίως στα ελληνικά και στα αγγλικά, ενώ παρουσιάζει ελλείψεις σε πεδία που προσφέρονται σε περισσότερες γλώσσες. Το συγκεκριμένο θέμα κρίνεται ιδιαίτερα προβληματικό για μια πλατφόρμα που προορίζεται για τον προσφυγικό πληθυσμό, δημιουργώντας προκλήσεις στη χρήση της.

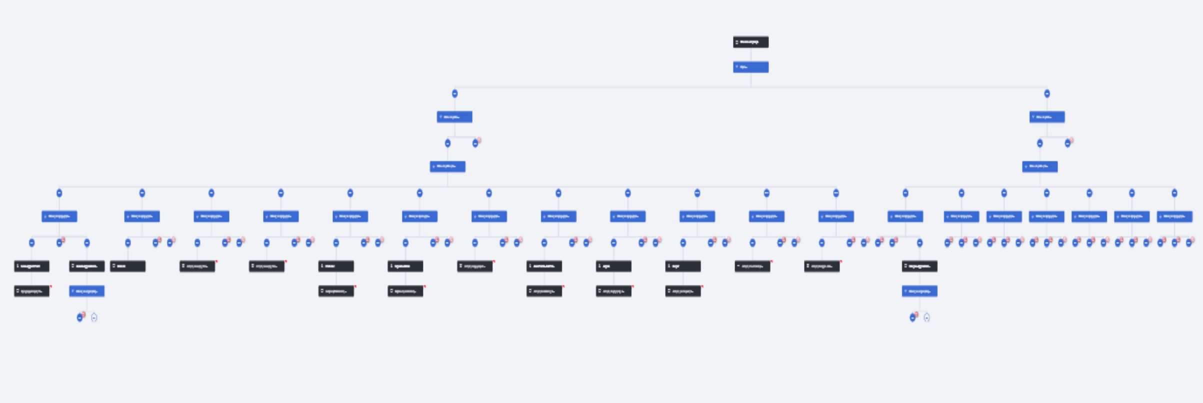
2 ΨΗΦΙΑΚΟ ΕΡΓΑΛΕΙΟ

Το ψηφιακό εργαλείο μάθησης που σχεδιάστηκε και υλοποιήθηκε στο πλαίσιο της παρούσας πτυχιακής εργασίας, συνδυάζοντας την καλλιέργεια συναφών δεξιοτήτων και την εξοικείωση με τις ηλεκτρονικές υπηρεσίες της Υπηρεσίας Ασύλου, επιδιώκει τον μετριασμό των προαναφερθέντων δυσλειτουργιών και αξιολογείται ως στοχευμένη και αποτελεσματική παρέμβαση από τον Ελληνικό Ερυθρό Σταυρό, ο οποίος αφενός το έχει ενσωματώσει στις παρεχόμενες υπηρεσίες του για την υποστήριξη και την κοινωνική ένταξη του προσφυγικού πληθυσμού, αφετέρου ενισχύει την περαιτέρω ανάπτυξή του.

Συγκεκριμένα, αποτελεί ένα εύχρηστο υπερμεσικό εκπαιδευτικό περιβάλλον, το οποίο παρέχει άμεση, διαδραστική και ευέλικτη πρόσβαση σε πολλαπλές μορφές πληροφορίας (κείμενο, εικόνα, ήχο, βίντεο, γραφικά, κινούμενα σχέδια) και σε ένα σύνολο διαφορετικών τύπων μαθησιακών αντικειμένων.

Το περιεχόμενό του επικεντρώνεται στις ηλεκτρονικές υπηρεσίες (e-services) που είναι διαθέσιμες στην κατηγορία «Υπηρεσίες Ασύλου» της ηλεκτρονικής πλατφόρμας της Υπηρεσίας Ασύλου (ΥΑ) του Υπουργείου Μετανάστευσης και Ασύλου, οι οποίες απευθύνονται σε αιτούντες και δικαιούχους διεθνούς προστασίας.

Η οργάνωσή του είναι δομημένη σε ένα δίκτυο συνδεδεμένων μεταξύ τους κόμβων, όπου ο κάθε κόμβος αντιστοιχεί σε συγκεκριμένη αίτηση της ψηφιακής πλατφόρμας και αποτελεί μία πολύμορφη πληροφοριακή ενότητα ή ένα μαθησιακό αντικείμενο. Οι κόμβοι διακρίνονται σε δύο κατηγορίες ως προς το περιεχόμενό τους και τη λειτουργία που επιτελούν: 1) κύριοι κόμβοι που αντιστοιχούν σε πληροφοριακές ενότητες και μαθησιακά αντικείμενα (μαύρο χρώμα) 2) κόμβοι διακλάδωσης με πληροφορίες που βοηθούν τον χρήστη να επιλέξει το κατάλληλο μονοπάτι βάσει της απάντησης που δίνει (μπλε χρώμα) (Εικόνα 2). Ως εκ τούτου, ο εκπαιδευόμενος έχει τον έλεγχο της πλοήγησης στο εκπαιδευτικό περιβάλλον και μπορεί να καθορίζει δυναμικά την πρόσβαση στις διάφορες ενότητες βάσει των αναγκών του και του επιπέδου γνώσεων και δεξιοτήτων που διαθέτει. Παράλληλα, η σχεδίαση του συγκεκριμένου εκπαιδευτικού περιβάλλοντος, λαμβάνοντας υπόψη ότι η εκάστοτε ηλεκτρονική αίτηση αποτελεί για την ομάδα στόχου μία σύνθετη κατάσταση την οποία καλείται να διαχειριστεί, κάνει χρήση επεξηγηματικών μηνυμάτων με πολλαπλές και συνάμα απλές αναπαραστάσεις, προκειμένου να την παρουσιάσει με εύληπτο και αποτελεσματικό τρόπο που δε θα απωθήσει τους εκπαιδευόμενους, αλλά θα αποτελέσει γι' αυτούς κίνητρο μάθησης και ενεργούς ενασχόλησης. Παρ' όλα αυτά, διατηρεί απλή τη διεπαφή, προκειμένου να ταιριάζει στα χαρακτηριστικά και τις ανάγκες της ομάδας στόχου.



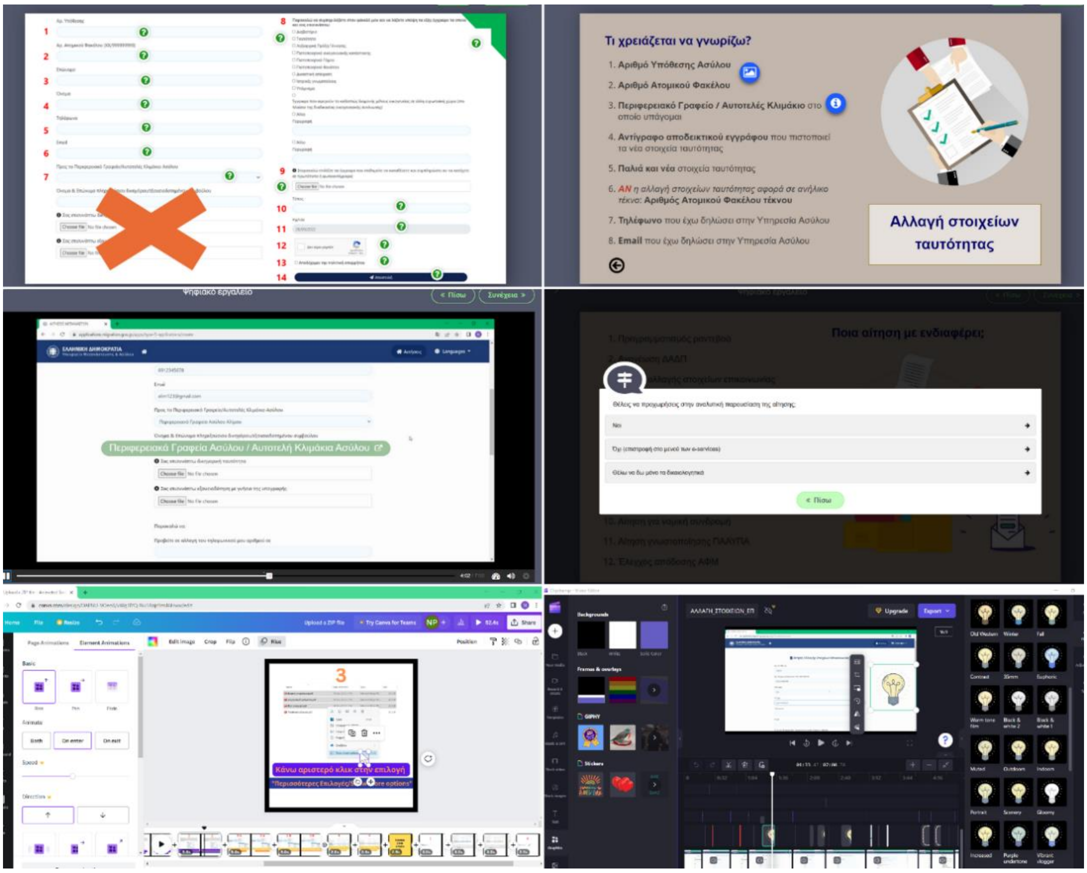
Εικόνα 2: Δομή ψηφιακού εργαλείου σε δίκτυο κόμβων (από το περιβάλλον σχεδίασης)

Όσον αφορά τη νοητική διάρθρωση του μαθησιακού εργαλείου, αυτή περιλαμβάνει οργανωμένα σύνολα πληροφοριών σχετικά με το είδος των αιτήσεων της ΥΑ, τα στοιχεία που είναι απαραίτητα για τη συμπλήρωση της κάθε αίτησης και τη διαδικασία υποβολής της κάθε αίτησης από το πρώτο μέχρι το τελευταίο στάδιο. Τα σύνολα πληροφοριών αποτυπώνονται με τρόπο που ακολουθεί τον υπερμεσικό χαρακτήρα του εργαλείου, ενώ οι διαδικασίες υποβολής των αιτήσεων που είναι διαθέσιμες στην ως άνω πλατφόρμα, παρουσιάζονται στους εκπαιδευόμενους κυρίως μέσα από δύο τύπους μαθησιακών αντικειμένων: λογισμικό καθοδηγούμενης διδασκαλίας και προσομοίωση.

Το λογισμικό καθοδηγούμενης διδασκαλίας (tutorial) σχεδιάστηκε με στόχο να προσομοιώσει σταδιακά τη διαδικασία και να παρουσιάσει στους εκπαιδευόμενους το περιεχόμενο αυτής βήμα προς βήμα, συνδυάζοντας εικόνα, ηχητικές περιγραφές, υπερσυνδέσμους και κινούμενα σχέδια σε ένα βίντεο. Αποτελεί αυτόνομη διδακτική μονάδα με τη μορφή βίντεο, η οποία παρέχει όλες τις πληροφορίες που θα τους υποστηρίξουν κατάλληλα, ώστε να κατανοήσουν την εκάστοτε αίτηση με το δικό τους ρυθμό, χωρίς τη χρήση άλλου εκπαιδευτικού υλικού. Το βίντεο-tutorial δίνει τη δυνατότητα στους εκπαιδευόμενους να ελέγξουν τη ροή (έναρξη/παύση), να προχωρήσουν σε επόμενο βήμα ή/και να γυρίσουν σε προηγούμενο, να επαναλάβουν βήματα της διαδικασίας και να τα συνδυάσουν με άλλες πηγές περιεχομένου (π.χ. επίσημο έγγραφο με πληροφορίες, επικουρικό βίντεο εστιασμένης θεματολογίας κ.ά.) προκειμένου να επιτύχουν τα προσδοκώμενα αποτελέσματα.

Η προσομοίωση (simulation) αποτελεί ένα ανοικτό μαθησιακό αντικείμενο, το οποίο αναπαριστά την πλατφόρμα της Υπηρεσίας Ασύλου για την εκάστοτε αίτηση. Οι εκπαιδευόμενοι έχουν τη δυνατότητα να εξασκηθούν πρακτικά και να πειραματιστούν κατ' επανάληψη με το περιβάλλον της πλατφόρμας, τη συμπλήρωση και την υποβολή της αίτησης, οικοδομώντας με αυτό τον τρόπο μία ολοκληρωμένη αναπαράσταση της διαδικασίας, οπότε καλλιεργούν την ικανότητα να την εφαρμόσουν με επιτυχία στην πραγματική αίτηση που είναι διαθέσιμη στην ιστοσελίδα της ΥΑ. Επιπλέον, την απλότητα και την ευχρηστία της προσομοίωσης συμπληρώνουν ένα σύμπλεγμα πληροφοριών που δίνονται στους εκπαιδευόμενους με διαφορετικές μορφές ανάλογα με τον βαθμό δυσκολίας και πολυπλοκότητας της επεξηγούμενης έννοιας/διαδικασίας, η επιλογή προβολής παραδείγματος συμπληρωμένης αίτησης, καθώς επίσης η δυνατότητα ανίχνευσης σφαλμάτων ως προς τη συμπλήρωση της αίτησης (π.χ. καταχώρηση λανθασμένης τιμής, παράλειψη υποχρεωτικού πεδίου) σύμφωνα με τις αντίστοιχες προδιαγραφές του πραγματικού περιβάλλοντος αιτήσεων της ΥΑ.

Τέλος, όσον αφορά τα τεχνολογικά μέσα που χρησιμοποιήθηκαν για την ανάπτυξη όλων των στοιχείων που συνθέτουν το ψηφιακό εργαλείο, αξιοποιήθηκε κυρίως το framework συνεργασίας περιεχομένου (content collaboration framework) HTML5 Package (H5P) ως ενσωματωμένο πρόσθετο εργαλείο (plugin) σε περιβάλλον WordPress για τη βασική δόμηση του εργαλείου μέσα από την αξιοποίηση των εξής τύπων περιεχομένων: εικόνα με hotspots, παρουσίαση, διαδραστικό βίντεο, διακλαδισμένο σενάριο. Παράλληλα, συμπληρωματικά μέσα διαμόρφωσης του ψηφιακού εργαλείου αποτέλεσαν η συγγραφή πηγαίου κώδικα με συνδυασμό στοιχείων HTML, CSS και JavaScript, η δημιουργία εικόνων με μορφότυπο Graphics Interchange Format (GIF) και επεξηγηματικών βίντεο μέσω του σχεδιαστικού εργαλείου Canva, καθώς και η ανάπτυξη βίντεο tutorials μέσα από την καταγραφή οθόνης, την επεξεργασία του παραγόμενου υλικού και την προσθήκη κινούμενων σχεδίων (animations) με τη χρήση του εργαλείου Climpchamp της Microsoft.



Εικόνα 3: Τεχνολογικά μέσα

3 ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Μεθοδολογία ανάπτυξης εργασίας

Αρχικά, πραγματοποιήθηκε έρευνα στην υπάρχουσα ελληνόφωνη και ξενόγλωσση βιβλιογραφία σχετικά με την προσφυγική ιδιότητα και τις συναφείς διαδικασίες στο ελληνικό πλαίσιο προς αιτιολόγηση της επιλογής της ομάδας στόχου. Η ίδια μέθοδος ακολουθήθηκε τόσο για την παρουσίαση των ΤΠΕ στο ελληνικό και ευρωπαϊκό πλαίσιο ψηφιοποίησης διαδικασιών όσο και για τη διαμόρφωση του γνωστικού περιεχομένου του εργαλείου που δημιουργήθηκε. Στο τελευταίο, πραγματοποιήθηκε επίσης συνεργασία με διεπιστημονική ομάδα επαγγελματιών για την αξιοποίηση της εμπειρίας της στον προσφυγικό τομέα. Τέλος, δημιουργήθηκε το ψηφιακό εργαλείο με αξιοποίηση πολλαπλών τεχνολογικών μέσων.

3.2 Μεθοδολογία δημιουργίας ψηφιακού εργαλείου

Για τη δημιουργία του ψηφιακού εργαλείου ακολουθήθηκαν τα κάτωθι βήματα:

1) Προσδιορισμός προσδοκώμενων αποτελεσμάτων.
2) Διερεύνηση και αποτύπωση εννοιών και διαδικασιών που επρόκειτο να υποστηρίξει το συγκεκριμένο εκπαιδευτικό εργαλείο, με παράλληλη εκτίμηση των γνωστικών δυσκολιών της ομάδας στόχου.
3) Σχεδιασμός της δομής του εργαλείου και προσδιορισμός των ψηφιακών στοιχείων και των τεχνικών χαρακτηριστικών του (διεπαφή και στοιχεία αλληλεπίδρασης, διασφάλιση διαλειτουργικότητας κ.ά.).
4) Ανατροφοδότηση και συνεργατική διαπραγμάτευση από τα μέλη της ομάδας σχεδιασμού για το γνωστικό περιεχόμενο και τα τεχνολογικά χαρακτηριστικά του εργαλείου.
5) Διαδοχικοί κύκλοι ελέγχου και διαμορφωτικής αξιολόγησης του εργαλείου.
6) Ολοκλήρωση του εκπαιδευτικού εργαλείου με πλήρη λειτουργικότητα και διασύνδεση των επιμέρους τμημάτων του.
7) Πιλοτική αξιολόγηση του πρωτότυπου μαθησιακού εργαλείου μέσα από αναλυτική παρουσίαση σε διευρυμένη διεπιστημονική ομάδα επαγγελματιών, η οποία περιέκλειε την αρχική ομάδα σχεδιασμού εμπλουτισμένη με επιπλέον διερμηνείς/πολιτισμικούς διαμεσολαβητές, κοινωνικούς λειτουργούς, κοινωνική ανθρωπολόγο και πρόσωπο αναφοράς για θέματα Συμμετοχικότητας, Επικοινωνίας και Δέσμευσης (ΣΕΔ), και την αλληλεπίδραση των μελών της ομάδας με το εργαλείο.
8) Δημοσίευση του μαθησιακού εργαλείου στην ελληνική γλώσσα, ώστε να είναι ευρέως διαθέσιμο και να δοκιμάζεται σταδιακά, ενόσω προετοιμάζονται οι εκδόσεις του εργαλείου σε επιπρόσθετες γλώσσες, οπότε θα διευρυνθεί ακόμα περισσότερο η αξιολόγησή του. Διευκρινίζεται ότι η διάθεση του εργαλείου σε άλλες γλώσσες και η τελική αξιολόγηση αυτού αποτελούν μεν σημαντικά στάδια του πλήρους κύκλου ανάπτυξης του εργαλείου, ωστόσο επεκτείνονται πέρα από τα όρια της παρούσας πτυχιακής εργασίας.

4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το ψηφιακό μαθησιακό εργαλείο που διαμορφώθηκε έρχεται να καλύψει το κενό που δημιουργείται από τη μη συμπερίληψη της πολυπολιτισμικής διάστασης της κοινωνίας στα βήματα ψηφιοποίησης που η τελευταία επιχειρεί, ενώ συγχρόνως αποσκοπεί να επηρεάσει θετικά την ενταξιακή διαδικασία των προσφύγων στην ελληνική κοινωνία, καθώς εστιάζει στην παροχή συγκεντρωτικής καθοδήγησης και στην καλλιέργεια αναγκαίων ψηφιακών δεξιοτήτων, επιδιώκοντας την αποτελεσματική και αυτόνομη πρόσβαση του προσφυγικού πληθυσμού στις ηλεκτρονικές υπηρεσίες ασύλου της Υπηρεσίας Ασύλου, έχοντας θέσει στο επίκεντρο της σχεδίασης και υλοποίησής του τις ανάγκες και τα ιδιαίτερα χαρακτηριστικά της ομάδας στόχου.

Η αξιολόγηση του εργαλείου που έχει πραγματοποιηθεί σε πρώτη φάση από επαγγελματίες που δραστηριοποιούνται στον προσφυγικό τομέα, έχει οδηγήσει στα εξής πρωτογενή συμπεράσματα:

- Οι περισσότερες υπηρεσίες που απευθύνονται στον προσφυγικό πληθυσμό και χρειάστηκε να ψηφιοποιηθούν τα τελευταία χρόνια λόγω των υγειονομικών και κοινωνικών συνθηκών, έχουν σχεδιαστεί με τα πρότυπα και τα δεδομένα της κοινωνίας υποδοχής και δε λαμβάνουν υπόψη τα ιδιαίτερα χαρακτηριστικά και τις ανάγκες των ανθρώπων που χρειάζεται να τις χρησιμοποιήσουν. Το παρόν εργαλείο, αποτελεί μία καινοτόμο προσπάθεια, η οποία επιδιώκει να γεφυρώσει αυτό το χάσμα, συγκεντρώνοντας σε ένα μέρος όλη την απαιτούμενη πληροφορία και γνώση τόσο σε επίπεδο βασικών πληροφοριών περί νομικού καθεστώτος και σχετικών δεδομένων όσο και σε επίπεδο ανάπτυξης των αναγκαίων ψηφιακών δεξιοτήτων.
- Η υιοθέτηση του συγκεκριμένου εργαλείου από μεγάλο μέρος της ομάδας στόχου αποτελεί μία επιδίωξη, η οποία χρειάζεται υπομονή και επιμονή, παρακολούθηση των μεταβολών που σημειώνονται στην πλατφόρμα της ΥΑ, συστηματική επαφή με τους ανθρώπους που το χρησιμοποιούν και μετουσίωση της ανατροφοδότησης που λαμβάνεται από αυτούς σε πράξη. Το συγκεκριμένο εργαλείο έχει υλοποιηθεί με τρόπο που επιτρέπει

τη δυναμική μεταβολή του περιεχομένου του και την προσθαφαίρεση εκπαιδευτικών λογισμικών, εφόσον κριθεί απαραίτητο.

- Η δομή του εργαλείου είναι σαφής και παρουσιάζει την ίδια πληροφορία με πολλαπλούς τρόπους και σε πολλά σημεία, με αποτέλεσμα να μην αφήνει τον χρήστη χωρίς καθοδήγηση και να επιτρέπει σε λιγότερο ή περισσότερο έμπειρους χρήστες να πλοηγηθούν σε αυτό.
- Η χρήση διαφορετικών τύπων εκπαιδευτικού λογισμικού για την παρουσίαση του εύρους των διαθέσιμων αιτήσεων της ΥΑ, έχει επιτευχθεί σε ικανοποιητική αναλογία, ώστε το εργαλείο να μην είναι βαρετό και μονοδιάστατο, αλλά ταυτόχρονα να μην μπερδεύει και αγχώνει τον χρήστη, αναγκάζοντάς τον να έρθει σε επαφή με πολλά άγνωστα εργαλεία και περιβάλλοντα, εντείνοντας το ψηφιακό άγχος που ήδη βιώνει.
- Δεδομένου ότι η κοινωνία υποδοχής σημειώνει μεγάλα και διαρκή βήματα μετάβασης στην ψηφιακή εποχή, η εξοικείωση με τις σύγχρονες τεχνολογίες και την ψηφιακή διακυβέρνηση, αποτελεί πλέον σημαντικό άξονα της κοινωνικής ένταξης του προσφυγικού πληθυσμού. Στο πλαίσιο αυτό, το παρόν μαθησιακό εργαλείο με τα προσδοκώμενα αποτελέσματα που υπηρετεί, μπορεί να συμβάλλει αποτελεσματικά στην πορεία που διαγράφει η ομάδα στόχου, επιδιώκοντας την πλήρη ένταξή της στην κοινωνία υποδοχής. Επομένως, πρόκειται για ένα εργαλείο με πολλές προοπτικές και δυνατότητες.
- Η διάθεση του εργαλείου στις βασικές ομιλούμενες γλώσσες του προσφυγικού πληθυσμού, ήτοι αγγλικά, γαλλικά, αραβικά, φαρσί και ουρντού, μετά την ολοκλήρωση της μετάφρασής του, ανοίγει τη βεντάλια πιλοτικής πλοήγησης από ένα σημαντικό εύρος χρηστών, γεγονός που πρόκειται να συμβάλλει στη συλλογή πολύτιμης ανατροφοδότησης με σκοπό τη βέλτιστη διαμόρφωση του ψηφιακού εργαλείου.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

[1] European Commission, ‘’2030 Digital Compass: The European Way for the Digital Decade, COM/2021/118/final’’, European Union, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118

[2] Πρόγραμμα Πολιτεία, Ν. 3345/2005 αρ.17, παρ.1, ΦΕΚ Α’138/16.06.2005.

[3] Ενιαία Ψηφιακή Πύλη της Δημόσιας Διοίκησης, Ν. 4635/2019 αρ. 52, ΦΕΚ Α’167/30.10.2019

[4] “Βίβλος Ψηφιακού Μετασχηματισμού 2020 - 2025,” Υπουργείο Ψηφιακής Διακυβέρνησης, Βίβλος Ψηφιακού Μετασχηματισμού 2020-2025. [Online]. Available: https://digitalstrategy.gov.gr/

[5] Ελληνική Δημοκρατία, Υπουργείο Μετανάστευσης και Ασύλου, "Εθνική Στρατηγική για την Ένταξη", Ελληνική Κυβέρνηση, Αθήνα, 2022. [Online]. Available: http://www.opengov.gr/immigration/?p=1599

[6] Ευρωπαϊκό Κοινοβούλιο, "Μια νέα ψηφιακή ατζέντα για την Ευρώπη : 2015.eu, Ψήφισμα του Ευρωπαϊκού Κοινοβουλίου της 5ης Μαΐου 2010 σχετικά με ένα νέο ψηφιακό θεματολόγιο για την Ευρώπη: 2015.eu (2009/2225(INI))", 2010. [Online]. Available: www.europarl.europa.eu/RegData/seance_pleniere/textes_adoptes/definitif/2010/05-05/0133/P7_TA%282010%290133_EL.pdf

[7] "Digitalising migrant integration services during the COVID-19 pandemic: adaptation, funding and accessibility", European Website on Integration, 2022. [Online]. Available: https://ec.europa.eu/migrant-integration/special-feature/digitalising-migrant-integration-services-during-covid-19-pandemic-adaptation_en.

[8] “Applications”, Ελληνική Δημοκρατία, Υπουργείο μετανάστευσης και ασύλου, 25-Aug-2022. [Online]. Available: https://applications.migration.gov.gr/

[9] United Nations High Commissioner for Refugees, “Greece must ensure safety net and integration opportunities for refugees,” UNHCR. [Online]. Available:

https://www.unhcr.org/news/briefing/2020/6/5ed603924/greece-must-ensure-safety-net-integration-opportunities-refugees-unhcr.html.

[10] Eurostat, EU-SILC (2018 data)

[11] Hellenic Red Cross and IFRC Joint Assessment, "Migration and Displacement Assessment Report, Greece", 2022. Available: https://www.ifrc.org/sites/default/files/2022-02/Greece_Country_Plan_2022.pdf

Early Forest Fire Detection using UAV and Computer Vision

Dimitrios A. Foteinos

ABSTRACT

In recent years, a major environmental problem with huge economic consequences that concern most European countries are forest fires. Classical identification and treatment techniques are found to be insufficient with large losses occurring each year due to them. There have been many proposals in the field of fire detection, with most of them being very costly and technologically advanced, requiring large technological infrastructure to maintain them. The solution we propose in this research concerns fire identification using UAVs combined with computer vision. More specifically, we have trained an object recognition model (Yolov5) through a custom fire dataset (images). The data collected by the drone are sent through the computer to smart devices and through an application that the fire authorities will have installed on their mobile phones, they can immediately see the place, the time, the date and also the photo of the fire, in order to intervene immediately and control it effectively.

Subject Area: Computer Vision, Drone Programming, Mobile App Development

Keywords: Unmanned Aerial Vehicles, Object Recognition, Fire Identification, Smart Devices, Forest Fires

ADVISOR

Stathes Hadjiefthymiades, Professor NKUA

1 INTRODUCTION

1.1 The cost of forest fires

Forest fires contribute to global greenhouse gas emissions and negatively affect public health, the economy and the provision of ecosystem services. In the humid tropics, fires are largely human-caused and lead to forest degradation. Studies have shown changing fire dynamics around the world due to changing climate and land use [1].

With the expansion of the European Union (EU from now on) appropriate defence against forest fires becomes more and more important. The number of annual forest fires within the EU is between 50–70 thousand, reaching 3–5 thousand square kilometres and causing a lot of financial problems, losing millions of Euros. It was recognized by environmental experts of the EU that forest fires cause significant ecological, economic and social problems in many European countries with possible long term consequences to the natural environment and the economy.

This year (2022) was a disastrous year for the countries of the European Union. Specifically, Forest fires have burned a record 700,000 hectares in the EU so far this year - the biggest amount since records began from 2006. Figures from the European Forest Fire Information System (EFFIS) [2] show that Spain has been the most heavily impacted so far with more than 297,000 hectares burned. It is followed by Romania (149,278), Portugal (103,462), France (61,911), and Italy (51,056). The number of fires has also shot up in recent years with more than 2,300 fires recorded across the bloc by mid-August, well above the average over the 2006-2021 period of 1,349 fires.

Table 1: Annual Fire Reports based on EFFIS

Country	**Country Area (ha)**	**Year 2022 (ha)**	**Annual Avg. 2006 - 2021 (ha)**
ESP - Spain	50,604,375	297,801	66,965.13
ROU - Romania	23,833,860	149,278	14,313.38
PRT - Portugal	9,187,803	103,462	96,625.44

FRA - France	54,951,621	61,911	9,825.69
ITA - Italy	30,075,506	51,056	53,961.38

1.2 Proposed IoT System

Many ways to identify forest fires have been proposed in recent years. Some of them are, for example, a network of thermal cameras, where identification can be made through them. Also, in recent years, satellite coverage has helped to identify and predict the course of the fire. Unfortunately, these systems are very expensive and require advanced technological infrastructures to maintain and be in continuous operation. The idea we propose in this thesis is an IoT system, which with the help of machine learning and a Drone, can help in the early identification of forest fires. More specifically, we did a spatial analysis of the area which is going to be surveillanced by the Drone. Then we trained a machine learning model through a custom fire dataset, and loaded it into our Drone. Thus, in mid-flight, we are able to locate a forest fire with great precision and send constant data from our Drone to a computer. Then, the computer in turn sends this data through an Android application to smart devices, which can be installed by the fire authorities, and thus intervene directly at the point where the fire has started. Through this low-cost concept, we achieve automation in the identification of forest fires, and significantly reduce a large part of the workforce, which can be used for other purposes.

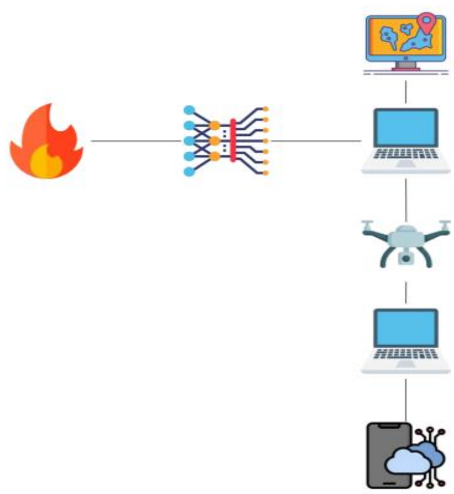


Figure 1: IoT System Workflow

1.3 Related Work

The last years, many proposals have been made in the field of Fire Detection with the aid of UAVs. Mostly, Machine Learning and Computer Vision algorithms are used in combination with Drones, making possible the fire detection from an aerial point of view, without the presence of humans (i.e from Helicopters or Aerial Manned Vehicles).

In this paper [3], a method using 2 UAVs is used. The first one consists of a (VTOL) fixed-wing UAV, which flies at a height of 350m - 5500m. If this particular drone observes anything suspicious, in order to reduce the chance of a false positive error, a second drone is sent to the area where the suspicious activity was observed, flying at a height of 10m - 350m.

Figure 2: Fire Detection with the use of 2 UAVs

The specific system consists of a control system on the ground which controls the mobility of these 2 UAVs. The model used for recognition is the Tensorflow Object - Detection API. The input data consists of images that contain smoke as well as fire. The disadvantage of this particular system is initially that it is quite a costly undertaking but also that the monitoring is controlled by a human presence.

2 TECHNICAL EQUIPMENT

2.1 System of Usage

The type of drone we choose to use for our approach is a Multi-Rotor drone, because it can offer great flexibility during the flight. After an extend research, we've decided to use the DJI Ryze Tello. This drone offers us the possibility to program it through a special SDK with the Python programming language. Thus, we have the possibility in real time, to run the fire detection program during the flight, without having to transfer the video streaming to another screen and run our algorithms from there.

Table 2: DJI Ryze Tello Specs

Aircraft	Flight Perfomance	Camera
Weight: 80g	Max Flight Distance: 100m	Photo: 5MP (2592 x 1936)
Dimensions: 98 x 92.5 x 41mm	Max Speed: 8m/s	FOV: 82.6°
Propeller: 3 inches	Max Flight Time: 13min	Video: HD720P30
Batteries: 3	Max Flight Height: 30m	Format: JPG(Photo), MP4 (Video)

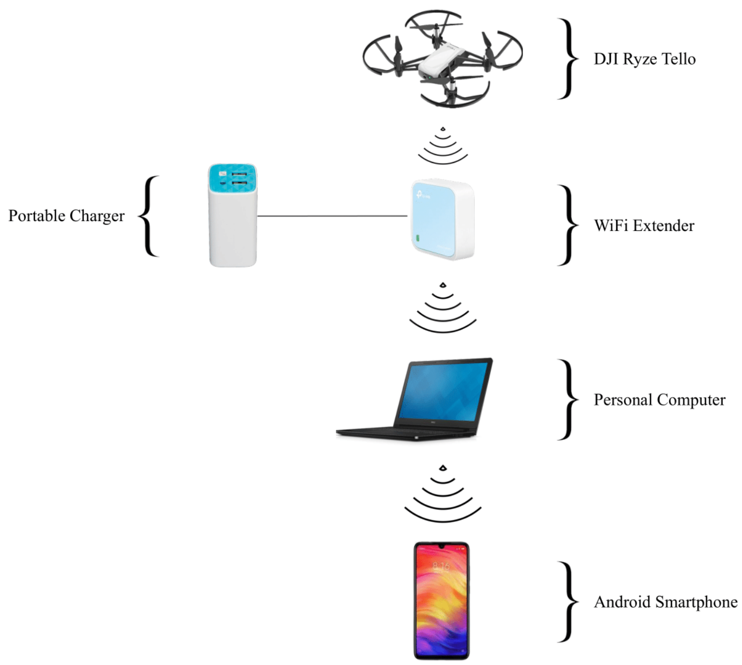


Figure 3: System Workflow

3 SPATIAL ANALYSIS

3.1 Study of the area of interest

In order to achieve the best possible identification of the fire, we first made a geographical study of the area. More specifically, we took the weather data at the specific time we make the flight, with the ones we are mainly interested in being the wind direction and speed. Then, we calculate based on the range radius of the Drone, the starting point where we'll start the surveillance. We used the mathematical tool GeoGebra to illustrate: 1) The area, 2) The wind lines, 3) The circle that our drone will make its flight. The total area our drone can cover is calculated as follows: $V = \left(\frac{2}{3}\right)\pi r^3$.

Where r = Drone's max range.

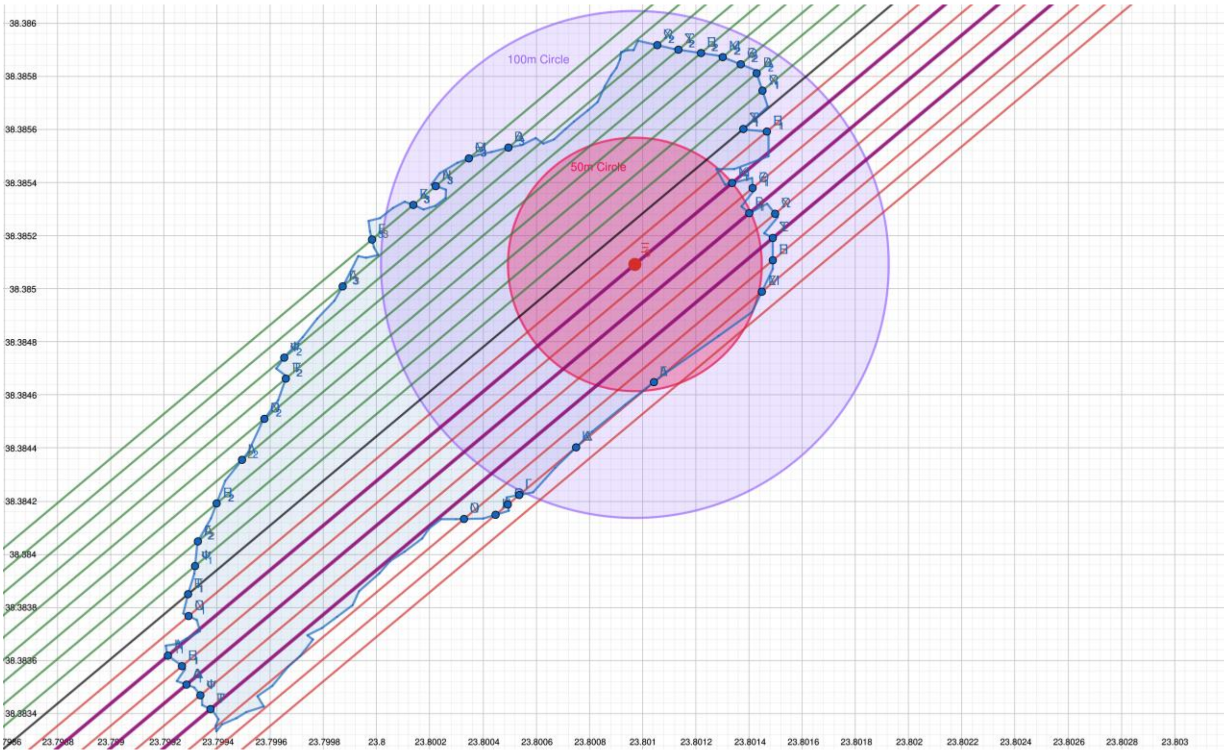


Figure 4: GeoGebra Analysis

The red circle is a hypothetical scenario where our drone can reach up to 50m range, and the purple is 100m. In this scenario, the wind direction is 50°. The green lines are the wind lines that are 'right' from the central and the red lines are these ones which are ‘left’ from the central. The purple lines are the lines that cover up the most part of the island.

3.2 Cost (Heuristic) Function

We also created a cost function that will evaluate in real-time if our decisions are right or not. We also must considerate our limits of the polygons. Our drone must fly within the limits and to not fly far away from there. The cost function is as follows:

$$C(x,y) = \big(w \cdot d(x,y) + t(x,y)\big) + a(x,y)$$

1. **Weight parameter:** $w = \mathbb{R}^*$

   This parameter will determine how ”important” is the d(x, y) factor. We’ve decided that this factor won’t be fixed but will be affected by the speed of wind. A first approach is to define this factor equal to the speed of wind, or a linear combination of it.

2. **Distance parameter:** $d\,(x,y) = d\left(p(x,y),\ set\big(inc_p\big)\right)$

   Where $p(x,y)$ is the current position of our drone and $set\big(inc_p\big)$ is the first ’set’ of the intersection points of the lines that are in the same direction of the wind, and the intersect our polygon. The reason that this distance interests us is because the cost of catching a fire close to intersection points is bigger than to be presented inside the polygon.

3. **Time parameter:** $t\big(d(x,y) > dis(`danger`)\big)$

   In particular, the time parameter counts the seconds, when the value of $t\big(d(x,y)\big)$ is bigger than the ’distance of danger’, which it means, the maximum distance where the drone is too far away from the intersection points. This parameter help us ’contain’ our drone close by the dangerous areas, with an option of freedom to explore different areas of interests near by.

4. **Activation Function:** $a(x,y) = \vdash^{1\ Polygon.contains(x,y)}_{\propto\ !Polygon.contains(x,y)}$

This function ensures that the drone will always move within the bounds of the polygon and will never escape from it. To make our UAV navigation completely autonomous, we need to find a way to approach that moves inside the polygon. Some initial thoughts are that the drone will follow a meandric approach for its tour.

4 YOLOV5 CUSTOM MODEL

4.1 Training Process

We trained the YOLOv5 Model with a custom dataset. More specifically, our dataset contains 502 annotated images, which have been split into 412 images for train and 90 images for validation. The images contain both small scale fires (lighters, candles) and large-scale fires (in cars, buildings and also forests). The training parameters are *416* as **input size**. The **batch size** we chose is *16*. We've trained our model for total *150* **epochs**. We also used the COCO dataset weights in order to achieve a **transfer learning** approach.

The matrix we're seeing in the figure 5 can be explained as follows:

- P_c : The probability of a class. If there's a fire or a smoke in the fire, this number will be 1. If there's no smoke or fire the probability will be 0.
- B_x : The coordinate in x-axis that determines the center of our bounding box.
- B_y : The coordinate in y-axis that determines the center of our bounding box.
- B_w: The width of the bounding box.
- B_h: The height of the bounding box.
- C_1: Fire class.
- C_2: Smoke class

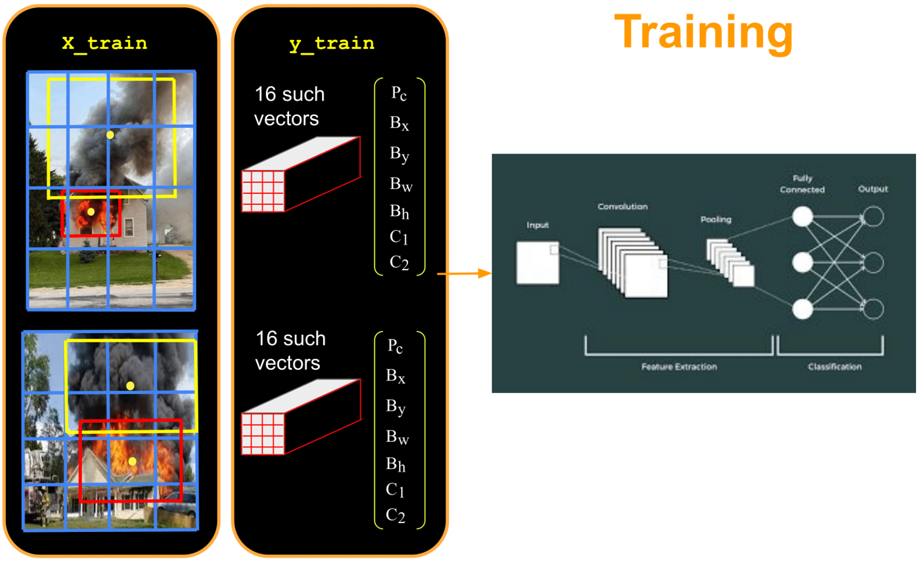


Figure 5: Training of the YOLOv5

4.2 Evaluation of the model

How can we measure if our model training was 'good' ? How can we measure this? We understand that it is not possible every time we want to check if our model is good, to check ourselves if the annotation and the confident score are satisfactory on each image. So, in the object detection field, there are certain metrics that tell us if our model is good, and thus, the work of evaluation is done automatically. This metric is **mAP** (mean average precision), which we will analyze further. Let's first look at the general results of our model in the evaluation that was done.

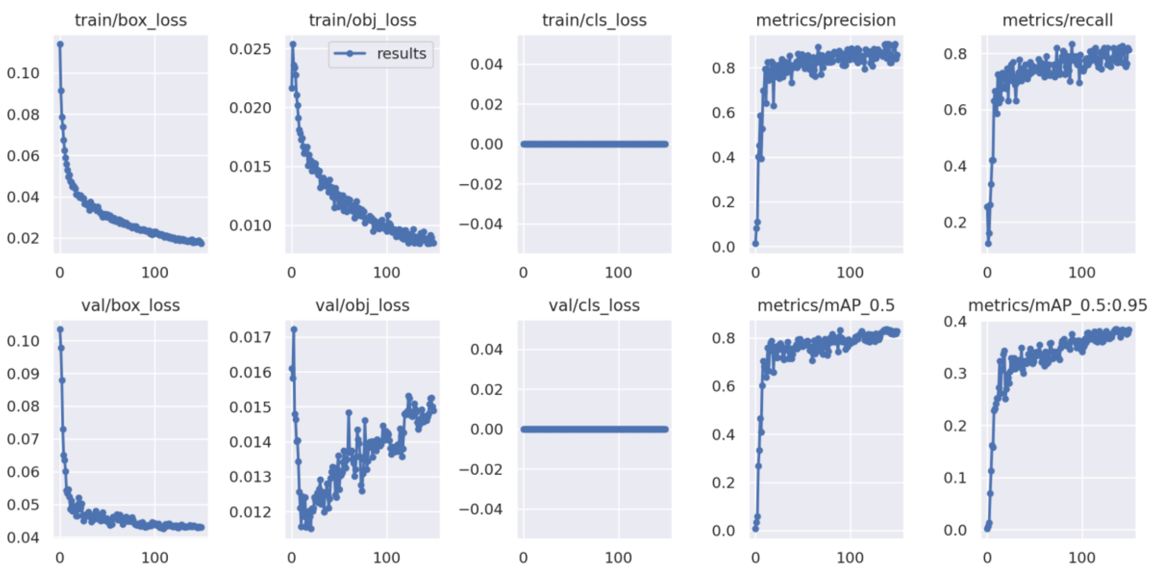


Figure 6: Evaluation Results

Let's now do some explain of the most important metric in Object Detection Field, the mAP. Models that involve an element of confidence can tradeoff precision for recall by adjusting the level of confidence they need to make a prediction. In other words, if the model is in a situation where avoiding false positives (stating a RBC is present when the cell was a WBC) is more important than avoiding false negatives, it can set its confidence threshold higher to encourage the model to only produce high precision predictions at the expense of lowering its amount of coverage (recall). The process of plotting the model's precision and recall as a function of the model's confidence threshold is the precision recall curve. It is downward sloping because as confidence is decreased, more predictions are made (helping recall) and less precise predictions are made (hurting precision).

5 RESULTS

5.1 Real Fire Experiment

In order to check our model ability to identify the fire, we tested our drone in multiple altitudes. So, we're going to see in which altitude had the best detection and if our model could possibly be useful at different conditions. The problems we've faced are two in this case: The first is the wind speed, which was good in our case, but because the fire creates self winds conditions and can easily change the wind's speed and direction, when we were back from the fire (in the first case, we are looking the fire forward), the drone had a hard time staying in the same position. The second problem, is the battery life. Each test lasted only about 13min, and we had to change the batteries in mid of the flight in order to continue our experiment.

Figure 7: Fire Cases

As we can see, we achieved the score the best results when facing the fire from a close distance and height. The first test case, was also the best one as we can see, since the confidence score is **0.7265**. The other test cases were not so good, but our system still managed to detected the fire, but with low confidence scores.

5.2 Mobile App Push Notification

The final stage of this IoT System, contains a mobile application which receives the data that the drone sent into our computer. In the figure below, we can see how the push notification is sent through the computer into our phone's screen as a push notification.

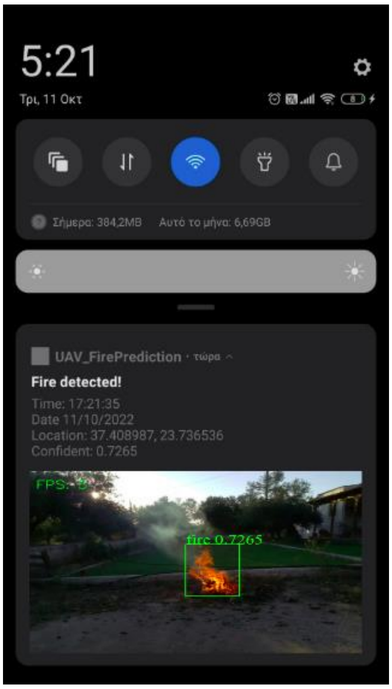


Figure 8: Push Notification

REFERENCES

[1] Alexandra Tyukavina et al. Forest fires concequenses. In Global Trends of Forest Loss Due to Fire From 2001 to 2019, page 1, 2022

[2] European forest fire information system: Annual reports. https://tinyurl.com/2p8pj783

[3] Diyana Kyuchukova, et al. Early forest fire detection using drones and artificial intelligence. In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019.

Διπλωματικές Εργασίες

Indexing N-Gram Graphs with Metric Access Methods

Nikolaos Gialitsis

ABSTRACT

This MSc thesis investigates the use of n-gram graph similarities in metric access methods for nearest neighbour search. While n-gram graphs offer efficient similarity functions that encapsulate semantic information, little research has been done on their metric properties and indexing capabilities. This thesis fills this gap by examining the satisfaction of metric properties by n-gram graph similarities and developing methods to convert them into indexable forms. A reference implementation is provided and evaluated on retrieval experiments of increasing complexity, demonstrating the potential of n-gram graph similarity search.

Subject Area: Information Retrieval

Keywords: indexing, n-gram graphs, similarity search, metric distances, natural-language-processing

ADVISOR

Dr George Giannakopoulos, Researcher (N.C.S.R. “Demokritos”)

Dr Stavros Perantonis, Researcher (N.C.S.R “Demokritos”)

Dr Haris Papageorgiou, Researcher (ATHENA-RC)

1 INTRODUCTION

Searching is an essential component of many applications including, among others, classification [1], multimedia retrieval [2], [3], clustering [4], abnormality detection [5] and recommendation systems [6]. A prominent problem in this domain is that of nearest neighbour search, where the goal is to find the most similar dataset object to a reference object called the query [7]. At the heart of search methods is the construction of a supporting data structure called an index, so that similarity searches can be carried out on that structure. Indexing exploits assumptions on the inner structure of datasets to make the nearest neighbour queries cheaper to resolve. Indexing methods adopt the query-by-example paradigm, where the database objects are ranked according to similarity to the query [7]. Several data structures are used for similarity search indexes, including trees, filter files, hash tables, and proximity graphs [8]–[12].

The field of similarity search, though originally aimed at an euclidean-space (vector) representation of indexed entities, has expanded beyond euclidean spaces, with the ability to index graphs and complex objects. In cases where the distances satisfy specific mathematical properties, called the metric axioms, Metric Access Methods (MAMs) can be employed to speed up similarity search [13]–[15]. On the flip side, searching via non-metric distances is generally a more challenging problem as many traditional solutions become suboptimal or require modifications to render them applicable [7], [16]. Nevertheless, the choice of searching with a metric or non-metric distance depends on the application domain and on the desired balance between accuracy and retrieval speed.

Efficient search methods for graph-structured data are in high demand, given the rise of graph-based ecosystems including NoSQL Graph databases, the semantic web, and gene-protein molecule interaction networks. Unfortunately, methods for graph comparison (or graph matching) are often computationally intensive as they attempt to detect local structures, like paths, trees, and subgraphs, as features for indexing schemes. These methods are primarily based on graph and subgraph isomorphism [17], [18] or graph edit-distance [19]. To this end, frameworks for graph representation exist that mitigate some of the costs associated with comparing graphs. The n-gram graph framework [20] integrates generalisation mechanics, which can be helpful in settings such as language [20]–[23] and biomedical informatics [24], [25], where the similarity between observations should be robust in the presence of noise or variation. N-gram graphs are equipped with similarity functions that are considered more lightweight than their traditional counterparts. Nevertheless, searching over collections of n-gram graphs has remained an open problem.

This thesis describes the design and development of a system for performing efficient nearest neighbour searches over n-gram graph collections. To this end, we (1) study whether the proposed similarity functions for n-gram graphs are usable by an index (2) study whether transformations to the graph distance functions can improve indexing performance (3) develop a reference software implementation for indexing n-gram graphs (4) evaluate and report the performance of each search configuration in the presence of noise.

2 BACKGROUND KNOWLEDGE

To develop a system for performing nearest neighbour search, one must carefully examine how closeness can be quantified and measured between the objects of interest. This chapter presents preliminary notions vital to understanding how indexing algorithms operate. We formulate nearest neighbour search and explain the shortcomings of searching in datasets of high dimensionality. We distinguish distances based on particular properties into metrics and non-metrics and describe their differences. Lastly, we present the n-gram graph framework, which is the application domain of this thesis.

2.1 Nearest Neighbour Search

Given a dataset $P = \{p_i \in D\}$, a distance $d : D \times D \rightarrow \mathfrak{R}$ and a query $q \in D$, the objective of *nearest neighbour search* is to retrieve an object $p^{\star}$such that $p^{\star} = arg\, min_{p_i \in P}\, d(p_i, q)$. From an algorithmic point of view, nearest neighbour searches often occur in distance space. A distance space is a mathematical construct composed of a set of objects along with a set of neighbourhoods for each object (a topological space) and a distance function (distance). More formally, a distance space is denoted as $D = (X, d)$ where X is a topological space and d is a distance function with signature $X^2 \rightarrow \Re$ that assigns higher values to more dissimilar objects and vice versa. Each space uses a number of dimensions in order to represent objects. A d-dimensional space is a space in which one uses d pieces of information (i.e., characteristics), called dimensions, to describe each object in that space. Equivalently, we can say that the dimensionality of the space is d.

The nature of the distance space ultimately dictates the final performance of a search system [26]. When the dimensionality of the space is high, the performance of nearest neighbour search methods is hindered by the manifestation of the so-called *curse of dimensionality* phenomena. The higher the dimensionality of the dataset, the more the objects become sparse (*emptiness of space* [27]) and equally dissimilar (*concentration of measure* [11]). The latter means that as dimensionality increases, the distance from the query to its nearest neighbour approaches the distance to its farthest neighbour, which also increases. In practice, the true dimensionality of the search space is unknown. Knowing the underlying generation mechanism of all possible objects is required, which is typically unavailable. Nevertheless, we can estimate the true dimensionality through indexability indicators. Indexability indicators measure the ability to efficiently search a distance space regardless of the indexing method used. A renowned indexability indicator is intrinsic dimensionality [11]. The intrinsic dimensionality is computed as $\rho = \mu^2 / (2\sigma^2)$ where μ and σ^2 are the mean and the variance of the distance distribution within a dataset. The intrinsic dimensionality indicates whether clusters of objects exist and how tight they might be. The intrinsic dimensionality can function as an early estimate of the performance of the search system.

2.2 Metric and Non-Metric Distances

Nearest neighbour searches can be significantly accelerated by using metric distances. A distance d is metric when it satisfies the following properties $\forall\ (x_{,i}, x_j, x_k) \in X$:

1. $d(x_i, x_j) \geq 0$ non-negativity
2. $d(x_i, x_j) = 0 \Leftrightarrow i = j$ identity of indiscernibles
3. $d(x_i, x_j) = d(x_j, x_i)$ symmetry
4. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$ triangle inequality

A space equipped with a metric distance d is called a *metric space*. Metric spaces are associated with a number of beneficial properties to indexing. For instance, all metric spaces are Hausdorff spaces[1], which means that objects can be grouped into distinct neighbourhoods. The Encyclopedia of Distances [28] maintains an extensive list of metrics applied throughout the different scientific disciplines. In addition to choosing one out-of-the-box metric distances, it is possible to derive new ones through metric-learning algorithms [14], [29], [30] . Metric learning algorithms often learn parametrized forms of existing metrics such as the linear Mahalanobis distance [31] or the non-linear χ^2 distance.

Although metric axioms are considered necessary for distances, scenarios have been presented in which non-metric distances naturally emerge, such as in bioinformatics and cognitive experiments [2], [15], [32], [33]. Distances that are robust to outliers or extremely noisy data, such as partial Hausdorff distance and non-metric L_p distances with $p \leq 1$ are typically non-metric. Other popular non-metric distances include the cosine distance, Dynamic Time Warping (DTW) distance, and the Kullback-Leibler (KL) divergence. In some cases, the distance measure may be the output of a complex algorithm or a black-box device, and an analytical form of the distance function may not be available, making it potentially non-metric [7].

2.3 N-Gram Graphs

Here we discuss n-gram graphs [20] which are graphs that represent sequences by retaining a portion of information from the original sequences. N-gram graphs disregard the exact order of sequences but maintain information on the co-occurrences of their sub-sequences. This condensing ability makes them robust in the presence of noise or variation. N-gram graphs can encode local and

[1] https://www.britannica.com/science/Hausdorff-space

global information of sequences, are agnostic to the position of the co-occurrences, and importantly, they assign unique labels to nodes, which is a key to efficient indexing.

To construct an n-gram graph, a sequence is split into overlapping subsequences of n elements, called n-grams, which are considered to co-occur when found within a window of a maximum distance threshold of each other. These neighbours are represented as connected vertices in an n-gram graph. For each selection of n in a user-specified range, overlapping n-grams are extracted by sliding a window over the length of the whole sequence. Smaller windows lead to more, but more compact neighbourhoods. By incorporating different-sized n-grams (i.e., different ranks), the n-gram graph captures information at multiple levels of detail, accounting for both high-level and low-level interactions.

Multiple similarity functions have been proposed in the literature to compare n-gram graphs [20]. Where $|G|$ refers to the number of edges in the n-gram graph G and e refers to a particular edge with weight w_e, popular similarities include the:

- Co-occurrence Similarity $CS(G_i, G_j) = \frac{|G_i \cap G_j|}{max\ (|G_i|, |G_j|)}$
- Containment similarity $CoS(G_i, G_j) = \frac{|G_i \cap G_j|}{min\ (|G_i|, |G_j|)}$
- Value Similarity $VS(G_i, G_j) = \frac{\sum_{e \in G_i} \quad VR(e)}{max\ (|G_i|, |G_j|)}\ where\ V(e)_{e \in (G_i \cap G_j)} = \frac{min(w^i{}_e, w^j{}_e)}{max(w^i{}_e, w^j{}_e)}$

All presented similarities are symmetric and in the range $[0,1]$ with a value of 1 for equivalent graphs and of 0 for absolutely distinct graphs. The similarities between n-gram graphs may be optionally enriched with information beyond the word or character level. For example, n-gram graph similarities can be leveraged together with named entity information on the word level in order to produce richer data representations [21]. Moreover, in classification settings, similarities may be used between n-gram graphs representing entire classes and n-gram graphs representing individual queries, to project sequences into a C-dimensional space of similarities where C is the number of classes.

3 RELATED WORK

In this chapter, we explore how indexing algorithms can potentially exploit the properties of metric distances to speed up search. We also present three approaches for transforming initial distances or similarities into more convenient forms for indexing.

3.1 Metric Access Methods

Metric Access Methods (MAMs) use the triangle inequality property to efficiently retrieve objects relevant to similarity queries. [11], [34], [35]. They are categorised into pivoting [35], [36] and compact partitioning techniques[10], [37]–[39], both of which filter candidate objects to improve search efficiency.

Pivoting techniques select objects from the dataset as pivots and pre-compute the distances between each pivot and other objects. Lower bounds for distances between the query and candidate objects are derived through triangle inequality, and objects are filtered based on these bounds. Compact partitioning techniques use bounding regions to divide objects into compact regions and filter objects during search. VP trees [10], [13], [40], [41] are binary trees with internal nodes representing partitions based on a pivot and are applicable to any metric space. The performance of a VP-tree deteriorates with increasing dimensionality, but it can speed up sequential search in low-dimensional spaces. MVP trees [42] are an extension of VP trees that use fewer pivots and store precomputed pivot-object distances in the leaves to improve filtering during querying.

3.2 Metric Conversion Methods

Converting non-metric distance functions to metric ones can be achieved using concave distance transformations called T-bases, controlled by a single parameter w that adjusts concavity or convexity. T-bases become triangle-generating modifiers (TG-modifier) for $w > 0$ and triangle-violating modifiers (TV-modifier) for $w \leq 0$. The TriGen algorithm [7] discovers the transformation that achieves the lowest intrinsic dimensionality without exceeding a user-specified triangle violation threshold.

Kernel functions can be transformed into a metric function, the kernel distance [43], [44], when the similarity function is a positive semi-definite kernel. This transformation allows for projections of objects into higher dimensions, making non-linearly separable data linearly separable without actually projecting objects. This is known as the "kernel trick" [45] and has been useful in machine learning, particularly for Support Vector Machines (SVMs) [46], [47].

Random projections, such as Locality-Sensitive Hashing (LSH) [48]–[52], are used to reduce high-dimensional data to a lower-dimensional space for approximate nearest neighbour queries. LSH uses hash functions that have a higher probability of collision for closer objects. By finding a hash family for a similarity function that satisfies certain conditions, a metric distance can be derived, making it useful for indexing objects using metric access methods based on a similarity function.

4 PROPOSED METHOD

Our objective is to efficiently handle near neighbour queries on n-gram graph collections, ensuring high accuracy and efficiency. To accomplish this, we recommend the combination of Metric Access Methods (MAMs) with similarity functions that operate on n-gram graphs. This approach allows us to create a unified framework for n-gram graph indexing without the need for separate solutions for each distance function. We propose transformations to n-gram graph similarities that always generate metric distances under certain conditions, which aids in developing indexing algorithms that can handle non-metric distances. Our approach comprises four distinct parts: Distance Modelling, Distance Characterization, Metric Access Method Selection, and Software Library Implementation. We discuss these steps in detail in the following sections.

To transform n-gram graph similarities into distances for efficient and precise nearest-neighbour searches, we use Value Similarity (VS) and a variation of Containment Similarity (CoS) called MinVS. VS is versatile and widely used, while

MinVS is more appropriate for smaller query objects. We explore three n-gram graph similarity transformations, resulting in six different n-gram graph distances.

Table 4.1: Examined transformations of n-gram graph similarities

Transformation	Formula
Standard d_S	$d_S(x,y) = 1 - s(x,y)$
Kernel d_K	$d_K(x,y) = \sqrt{s(x,x) - 2s(x,y) + s(y,y)}$
TriGen d_T	$d_T(x,y,w) = \begin{cases} (1-s(x,y))^{\frac{1}{1+w}} & \text{for } w > 0 \\ (1-s(x,y))^{1-w} & \text{for } w \leq 0, \end{cases}$

To determine if a derived distance is metric, we use different proving methods based on the distance transformation. We check if the similarities are positive-definite kernels for kernel-based distances. For d_S, we use various proof methods such as direct proofs, proofs by cases, proofs by contradiction, proofs by counter-example, and proofs by equivalence. For d_T distances, we only need to prove symmetry, non-negativity, and the identity of indiscernibles since the TriGen algorithm controls the triangle inequality. If we cannot determine if a distance is metric, we experimentally evaluate its metric satisfaction by randomly sampling a large number of distance triplets from a dataset and checking for violations of the triangle inequality.

We use MVP trees to index n-gram graphs since they have demonstrated good performance with non-metric similarities. We prefer the more advanced MVP tree over the traditional VP tree because of its better use of pivots and potentially faster search times. To address the issue of exploring partitions that cannot be pruned, we implement an early-stopping strategy by setting a limit on the number of visited leaves. Prior research on approximate VP-tree search has also discussed early-stopping strategies.

To create a usable library for n-gram graph search, we establish requirements for minimal RAM during index construction, optimised disk storage for precomputed distances, and ease of use for reproducible experiments. We avoid loading large collections of complex data types into memory by designing an algorithm for computing the distance matrix in partitions.

Data: N sequences, distance d, max graphs M
Result: distance matrix D
split dataset into $P = \lfloor N/(M/2) \rfloor$ partitions;
for *each partition* $X \in P$ **do**
  compute $X \times X$ submatrix;
  for *each partition* $Y \in [X+1,\ P]$ **do**
    compute $X \times Y$ submatrix;
  end
end

Our library is implemented in C++, integrated with the Metric Spaces Library, and publicly available with source code and documentation

5 INDEXING EXPERIMENTS

To develop an index based on the described distances, we aim to determine the speed gain, noise tolerance, accuracy, and retrieval time of the search system. In this chapter, we present our indexing experiments to address these questions, including the dataset, parameters, and evaluation process used. We used the 20 newsgroup text collection as our dataset, which contains forum posts covering 20 categories. We constructed n-gram graphs from the training set by keeping only the first 100 characters of each text and removing metadata. This resulted in 10,294 texts for evaluation.

We used the TriGen algorithm to determine the Trigen distance parameter w for n-gram graph similarity. We set w to 1 if the proportion of triplets violating the triangle inequality was greater than $\theta = 0.02$ and w to -1 otherwise. We identified the optimal value of w using an iterative process that doubled its value until T_{error} exceeded θ, then performed an exhaustive search in the range $[min(w/2, w), max(w/2, w)]$. We represented each of the 10,294 evaluation set sequences with a character-level n-gram graph and evaluated the search system's robustness against noisy input by performing a fuzzy search with variations of 1%, 2%, and 4%, 8%, and 16% noise. We measured recall, total distance computations, and differences between found neighbours and true neighbours, and included a baseline query with 0% noise.

6 EXPERIMENTAL RESULTS AND DISCUSSION

We created 6 indexes, one for each distance configuration, by computing the distance matrix, serialising it, and building an MVP tree. Index construction time ranged from 1.5 to 1.7 hours. Serialisation and deserialization times were between 3.1 and 4.5 seconds. The index required 46.6 KB of disk storage, while the distance matrix required 847.7 MB.

Table 7.1: Time and space requirements for index construction

Index Construction Time	1.5 - 1.7 hours
Distance Matrix Serialization Time	3.1 - 3.5 seconds
Distance Matrix Deserialization Time	3.3 - 4.5 seconds
Index Disk Storage	46.6 KB
Distance Matrix Storage	847.7 MB

The d_S and d_K spaces exhibit high intrinsic dimensionality and poor performance. d_T has a low intrinsic dimensionality due to the TriGen algorithm application, resulting in improved performance. d_K becomes non-metric, which can harm MVP-tree search.

Table 7.2: Statistics on the distance space induced by each configuration

transformation	similarity	mean	variance	intrinsic dimensionality
d_S	VS	0..998	0,0003	1989.21
	MinVS	0.998	0.0003	1948.45
d_K	VS	1.413	0.0004	2351.39
	MinVS	1.412	0.0004	2330.45
d_T	VS	0.931	0.0287	15.11
	MinVS	0.930	0.0294	14.69

6.1 Results of the Experimental Evaluation

We compared the speed and accuracy of our index against sequential search. Sequential search needs 10294 distance computations per query, while our index needs, on average, 546.47 and 608.99 computations for MinVS and VS distances, respectively, achieving a 95% speed improvement. Our index always identifies the true nearest neighbour for non-noisy inputs. Retrieval accuracy declines as noise increases, with less than half of true nearest neighbours identified at the highest noise percentage. Under the $d_S MinVS$ distance, we can achieve approximately 85% accuracy at 4% noise by comparing the query to just 10% of the entire database. $d_S VS$ and $d_K VS$ distances retrieve neighbours that are relatively close to the true answers, with average retrieved distances being less than 13% higher than the original distances. While d_T distances may return distant neighbours, they require the least amount of distance computations in all noise experiments. This renders them suitable for noisy query applications where retrieval speed is crucial. Distance computations peak at 8% noise and decrease sharply at 16% noise.

Table 7.3: Performance statistics on each distance configuration and noise percentage

			noise percentage%					
evaluation statistic	**transformation**	**similarity**	***0%***	***1%***	***2%***	***4%***	***8%***	***16%***
recall	d_S	VS	1.000	0.953	0.886	0.788	0.679	0.424
		MinVS	1.000	0.954	0.916	0.835	0.676	0.428
	d_K	VS	1.000	0.953	0.886	0.788	0.679	0.424
		MinVS	1.000	0.953	0.888	0.792	0.676	0.423
	d_T	VS	1.000	0.949	0.907	0.819	0.659	0.403
		MinVS	1.000	0.949	0.908	0.821	0.664	0.415
average distance computations per query	d_S	VS	608.98	1131.44	1112.52	1068.49	973.45	757.46
		MinVS	546.47	1142.61	1127.61	1084.15	986.49	780.28
	d_K	VS	608.98	1130.22	1111.85	1067.39	972.89	757.46
		MinVS	546.47	1143.04	1127.31	1083.26	985.78	780.28
	d_T	VS	608.98	620.56	628.45	642.50	656.06	600.36
		MinVS	546.47	564.81	577.66	611.41	650.60	641.67
avg. retrieved neighbor distance / avg. true neighbor distance	d_S	VS	-	1.108	1.120	1.131	1.116	1.073
		MinVS	-	1.494	1.409	1.360	1.280	1.153
	d_K	VS	-	1.108	1.120	1.131	1.116	1.073
		MinVS	-	1.112	1.120	1.129	1.118	1.073
	d_T	VS	-	> 4000	> 8000	> 19900	1180	192.035
		MinVS	-	> 4000	> 7000	> 16600	50000	412.664

7 CONCLUSIONS AND FUTURE WORK

The thesis aimed to develop an effective indexing method for n-gram graphs by formulating it as a nearest-neighbour search and extensively reviewing indexing methods. We demonstrated that proposed similarity functions can lead to metric distances and used the TriGen algorithm to improve indexing performance. Our approach combined MVP trees with n-gram graph distances and applied an early termination strategy to improve search efficiency. We evaluated our approach in multiple retrieval experiments and identified the most suitable configuration for efficiency and accuracy.

Future work will focus on evaluating more similarities, exploring distance transformations, experimenting with additional access methods, and testing our method on more datasets. We will also investigate using metric learning methods to learn custom distances from annotated examples. These efforts will unlock the potential of n-gram graph indexing and pave the way for further improvements.

REFERENCES

[1] J. Lokoăź, J. Kohout, P. Čech, T. Skopal, and T. Pevnỳ, “k-NN Classification of Malware in HTTPS Traffic Using the Metric Space Approach,” in Proceedings of the 11th Pacific Asia Workshop on Intelligence and Security Informatics-Volume 9650, 2016, pp. 131–145.

[2] D. Jacobs, D. Weinshall, and Y. Gdalyahu, “Class representation and image retrieval with non-metric distances,” IEEE Trans Pattern Anal Mach Intell, vol. 22, no. 6, pp. 583–600, 2000.

[3] J. Lokoč, T. Grošup, and T. Skopal, “SIR: the smart image retrieval engine,” in Proceedings of the 5th international conference on Similarity Search and Applications, 2012, pp. 240–241.

[4] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” Pattern Recognit., vol. 36, no. 2, pp. 451–461, 2003.

[5] P. Boniol, J. Paparrizos, T. Palpanas, and M. J. Franklin, “SAND: streaming subsequence anomaly detection,” Proc. VLDB Endow., vol. 14, no. 10, pp. 1717–1729, 2021.

[6] J. Suchal and P. Návrat, “Full text search engine as scalable k-nearest neighbor recommendation system,” in IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, 2010, pp. 165–173.

[7] T. Skopal, “Unified framework for fast exact and approximate search in dissimilarity spaces,” ACM Trans. Database Syst. TODS, vol. 32, no. 4, pp. 29-es, 2007.

[8] K. Echihabi, K. Zoumpatianos, and T. Palpanas, “Scalable machine learning on high-dimensional vectors: From data series to deep network embeddings,” in Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics, 2020, pp. 1–6.

[9] K. Echihabi, K. Zoumpatianos, T. Palpanas, and H. Benbrahim, “Return of the Lernaean Hydra: Experimental Evaluation of Data Series Approximate Similarity Search,” Proc VLDB Endow, vol. 13, no. 3, pp. 403–420, Nov. 2019, doi: 10.14778/3368289.3368303.

[10] P. N. Yianilos, “Data Structures and Algorithms for Nearest Neighbor,” in Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, SIAM, 1993, p. 311.

[11] E. Chávez, G. Navarro, R. Baeza-Yates, and J. L. Marroquín, “Searching in metric spaces,” ACM Comput. Surv. CSUR, vol. 33, no. 3, pp. 273–321, 2001.

[12] T. Skopal and J. Lokoč, “NM-tree: Flexible approximate similarity search in metric and non-metric spaces,” in International Conference on Database and Expert Systems Applications, Springer, 2008, pp. 312–325.

[13] P. Ciaccia, M. Patella, and P. Zezula, “M-tree: An efficient access method for similarity search in metric spaces,” in Vldb, 1997, pp. 426–435.

[14] A. Bellet, A. Habrard, and M. Sebban, “Metric learning,” Synth. Lect. Artif. Intell. Mach. Learn., vol. 9, no. 1, pp. 1–151, 2015.

[15] M. J. Cullinane, “Metric axioms and distance,” Math. Gaz., vol. 95, no. 534, pp. 414–419, 2011.

[16] Y. Mu and S. Yan, “Non-metric locality-sensitive hashing,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2010, pp. 539–544.

[17] R. C. Read and D. G. Corneil, “The graph isomorphism disease,” J. Graph Theory, vol. 1, no. 4, pp. 339–363, 1977.

[18] J. R. Ullmann, “An algorithm for subgraph isomorphism,” J. ACM JACM, vol. 23, no. 1, pp. 31–42, 1976.

[19] X. Gao, B. Xiao, D. Tao, and X. Li, “A survey of graph edit distance,” Pattern Anal. Appl., vol. 13, no. 1, pp. 113–129, 2010.

[20] G. Giannakopoulos, V. Karkaletsis, G. Vouros, and P. Stamatopoulos, “Summarization System Evaluation Revisited: N-gram Graphs,” ACM Trans Speech Lang Process, vol. 5, no. 3, p. 5:1-5:39, Oct. 2008, doi: 10.1145/1410358.1410359.

[21] L. Tsekouras, I. Varlamis, and G. Giannakopoulos, “A graph-based text similarity measure that employs named entity information,” in Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 765–771.

[22] G. Giannakopoulos, V. Karkaletsis, and G. A. Vouros, “Testing the Use of N-gram Graphs in Summarization Sub-tasks.,” in TAC, 2008.

[23] J. Violos, K. Tserpes, I. Varlamis, and T. Varvarigou, “Text Classification Using the N-Gram Graph Representation Model Over High Frequency Data Streams,” Front. Appl. Math. Stat., vol. 4, p. 41, 2018, doi: 10.3389/fams.2018.00041.

[24] D. Polychronopoulos, D. Sellis, and Y. Almirantis, “Conserved Noncoding Elements Follow Power-Law-Like Distributions in Several Genomes as a Result of Genome Dynamics,” PLOS ONE, vol. 9, no. 5, pp. 1–12, 2014, doi: 10.1371/journal.pone.0095437.

[25] S. Kazantzidis, A. Krithara, and G. Paliouras, “Splice Site Prediction: Transferring Knowledge Across Organisms.,” in BIOINFORMATICS, 2017, pp. 160–167.

[26] A. Hoyos, U. Ruiz, S. Marchand-Maillet, and E. Chávez, “Indexability-based dataset partitioning,” in International Conference on Similarity Search and Applications, Springer, 2019, pp. 143–150.

[27] D. W. Scott and J. R. Thompson, “Probability density estimation in higher dimensions,” in Computer Science and Statistics: Proceedings of the fifteenth symposium on the interface, North-Holland, Amsterdam, 1983, pp. 173–179.

[28] M. M. Deza and E. Deza, “Encyclopedia of distances,” in Encyclopedia of distances, Springer, 2009, pp. 1–583.

[29] K. Q. Weinberger and L. K. Saul, “Unsupervised learning of image manifolds by semidefinite programming,” Int. J. Comput. Vis., vol. 70, no. 1, pp. 77–90, 2006.

[30] H. Oh Song, Y. Xiang, S. Jegelka, and S. Savarese, “Deep metric learning via lifted structured feature embedding,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4004–4012.

[31] E. Xing, M. Jordan, S. J. Russell, and A. Ng, “Distance metric learning with application to clustering with side-information,” Adv. Neural Inf. Process. Syst., vol. 15, 2002.

[32] F.-M. Schleif and P. Tino, “Indefinite proximity learning: A review,” Neural Comput., vol. 27, no. 10, pp. 2039–2096, 2015.

[33] A. Tversky, “Features of similarity.,” Psychol. Rev., vol. 84, no. 4, p. 327, 1977.

[34] L. Boytsov, “Efficient and accurate non-metric k-NN search with applications to text matching,” PhD Thesis, Ph. D. Dissertation. Carnegie Mellon University, 2018.

[35] M. L. Hetland, “The basic principles of metric indexing,” in Swarm intelligence for multi-objective problems in data mining, Springer, 2009, pp. 199–232.

[36] M. L. Micó, J. Oncina, and E. Vidal, “A new version of the nearest-neighbour approximating and eliminating search algorithm (AESA) with linear preprocessing time and memory requirements,” Pattern Recognit. Lett., vol. 15, no. 1, pp. 9–17, 1994.

[37] J. Beniley, “Multidimensional binary search trees used for associative searching,” ACM Commun., vol. 18, no. 9, pp. 509–517, 1975.

[38] S. Brin, “Near neighbor search in large metric spaces,” 1995.

[39] A. Guttman, “R-trees: A dynamic index for geometric data,” in Proceedings of the ACM SIGMOD International Conference on Management of Data, 1984, pp. 47–57.

[40] J. K. Uhlmann, “Satisfying general proximity/similarity queries with metric trees,” Inf. Process. Lett., vol. 40, no. 4, pp. 175–179, 1991.

[41] T. Liu, A. Moore, K. Yang, and A. Gray, “An investigation of practical approximate nearest neighbor algorithms,” Adv. Neural Inf. Process. Syst., vol. 17, 2004.

[42] T. Bozkaya and M. Ozsoyoglu, “Distance-based indexing for high-dimensional metric spaces,” in Proceedings of the 1997 ACM SIGMOD international conference on Management of data, 1997, pp. 357–368.

[43] J. M. Phillips and S. Venkatasubramanian, “A gentle introduction to the kernel distance,” ArXiv Prepr. ArXiv11031625, 2011.

[44] S. Joshi, R. V. Kommaraji, J. M. Phillips, and S. Venkatasubramanian, “Comparing distributions and shapes using the kernel distance,” in Proceedings of the twenty-seventh annual symposium on Computational geometry, 2011, pp. 47–56.

[45] S. Theodoridis and K. Koutroumbas, “Pattern recognition. 2003,” Google Sch. Google Sch. Digit. Libr. Digit. Libr., 2009.

[46] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[47] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” IEEE Intell. Syst. Their Appl., vol. 13, no. 4, pp. 18–28, 1998.

[48] P. Indyk and R. Motwani, “Approximate nearest neighbors: towards removing the curse of dimensionality,” in Proceedings of the thirtieth annual ACM symposium on Theory of computing, 1998, pp. 604–613.

[49] E. Kushilevitz, R. Ostrovsky, and Y. Rabani, “Efficient search for approximate nearest neighbor in high dimensional spaces,” SIAM J. Comput., vol. 30, no. 2, pp. 457–474, 2000.

[50] A. Gionis, P. Indyk, R. Motwani, and others, “Similarity search in high dimensions via hashing,” in Vldb, 1999, pp. 518–529.

[51] Q. Huang, J. Feng, Y. Zhang, Q. Fang, and W. Ng, “Query-aware locality-sensitive hashing for approximate nearest neighbor search,” Proc. VLDB Endow., vol. 9, no. 1, pp. 1–12, 2015.

[52] J. Gao, H. V. Jagadish, W. Lu, and B. C. Ooi, “DSH: data sensitive hashing for high-dimensional k-nnsearch,” in Proceedings of the 2014 ACM SIGMOD international conference on Management of data, 2014, pp. 1127–1138.

Low Earth Orbit Microsatellite Constellation Utilizing Satellite Hellas Sat 5 as a Relay

Alexandros K. Lais

ABSTRACT

Driven by investments from private and venture capital, technological advancements are fueling a rekindled interest in Satellite Communications (SatComs). These systems have seen widespread use in various fields such as media broadcasting, news gathering, and backhaul. Presently, SatComs are shifting from system design to data services as the Internet applications continue to progress. Companies that were previously reserved for governments and large international industries now offer various manufacturing and listing options aided by the "New Space" project. Innovative earth observation missions and broadband efforts necessitate the refinement of satellite communication systems.

The concept of New Space is not centered around a particular technology, but rather around a fresh mindset toward space. The origins of this idea can be traced to three main factors: 1) the privatization of space, 2) the miniaturization of satellites, and 3) new services that make use of space data. Privatization involves the making and launching of satellites by non-governmental companies. Meanwhile, advancements in satellite and component miniaturization have made it possible to get multiple cube/micro/nano-satellites into space using just one launcher. By achieving quick and affordable entry into space, the integration of the initial two elements has resulted in the latter. Additionally, 5G terrestrial network technologies have been integrated and are currently concentrated on leveraging satellite networks in order to enhance their efficiency and accessibility. In the near future, 5G is set to surpass its predecessors by accommodating a vast array of new applications, catering primarily to pivotal

market segments such as transportation and automotive, entertainment and media, Industry 4.0, and e-Health.

In recent decades, thanks to technological advances in airborne and miniaturized satellite platforms, an intermediate layer of communication systems has emerged between the terrestrial and traditional satellite segments. Regardless of application, these new platforms can be categorized by their operation altitude. All in all, the rapid development of the space sector coupled with the cost-reducing approach taken by the private sector suggests that the upcoming future space technologies seem very promising.

Having said all the above, this thesis presents a comprehensive analysis of all the required steps that a system engineer / designer must consider in order to build and deploy a fully functional and reliable satellite communication system. The methodology entails a fully description of the basic laws of the space environment as well as an extensive analysis of the orbital mechanics and parameters. The objective is to demonstrate how theory can be utilized in an actual satellite communication system and evaluate its performance through computer simulations. The implementation scenario was to utilize a GEO satellite acting as a relay station communicating through ISLs with a four-satellite constellation in low Earth orbital position. The GEO satellite is going to be facing the ground station receiver at all times while the constellation satellites are going to connect to the GEO satellite with the use of ISLs. Towards this end, the last step involves the design and implementation of the system on specialized software and the presentation / discussion of the results.

The main conclusion of the above implementation is the fact that using a low Earth orbit (LEO) satellite constellation combined with a geostationary (GEO) satellite as a relay, it’s possible to achieve a resilient and reliable communication link with exceptional high data rates and an almost worldwide coverage.

Subject area: Satellite communications

Keywords: Link budget, satellite orbit, atmospheric losses, inter-satellite links, orbital design

ADVISOR

Demosthenes Vouyioukas, Professor University of the Aegean

1 PREFACE

The methodology followed in authoring this thesis was a comprehensive presentation of all the considered parameters in satellite communications, covering all stages starting from the basic laws of physics to the modern technologies used for a successful data satellite link.

Almost all of the relative theory was implemented by creating a scenario that included a constellation of microsatellites using the Hellas-Sat geostationary satellite as a relay station. The software used was the Systems Tool-Kit (STK) of the AGI company, for the use of which a special permission was requested via e-mail for the construction and study of this thesis. The request was granted with the provision of a four-month free license providing access to the entire capabilities of the software.

The design and implementation were conducted regarding the Mediterranean area which was designated as the area of study / coverage. Findings such as hours per coverage rates, access times per satellite, azimuth / angle / distance changes as well as complete link budget analysis, were calculated for the designated area.

The main result is the fact that by using a constellation of four microsatellites with specific orbital parameters in combination with a geostationary satellite that operates as a relay, the provided coverage for the considered area is 94.6% for 18.34 hours per day per satellite. This means that by applying a constellation with four microsatellites, complete coverage is possible for a 24-hour period.

2 INTRODUCTION

2.1 Satellite Communications

Satellite communication networks are considered to be one of the most significant parts of modern telecommunications systems mainly because of its capabilities regarding development, modernization, data broadcasting techniques and point-to-point links. Given the fact that modern applications require high stability, interference resilience, low latency and increased throughput, all the above technologies would be impossible to implement without the use of satellite systems.

Communications with the use of satellites are the result of research in the field of communications, space, and modern technologies in shipping and trade, aiming to increase coverage and manage spectrum and quality of services (QoS) with the lowest possible cost of implementation and maintenance. Apart from the research, satellite communications have also been part of every day's life, as many applications that contribute to people's lives, would be impossible to operate without them (e.g., positioning services that contribute to facilitating transport and navigation, or weather forecasts for emergency measures in adverse weather conditions). Satellite communications has also great contribution in telecommunication applications by providing global coverage and communication to anywhere on Earth in which there is no terrestrial network support.

In the past, the performance of satellites was severely restricted. To overcome this limitation, it was necessary to construct exceptionally large ground stations with dish antennas measuring over twenty meters in diameter, in order to establish communication links with them. The use of satellites was limited to long distance telephone and television signal transportation between studios. By 1990, two out of every three intercontinental telephone calls were transmitted by telecommunication satellites. Satellites proved particularly useful for communicating with countries in which the terrestrial infrastructure was insufficient or even non-existed.

Recent technology changes and various kinds of demands have changed the way communications satellites are used. More powerful satellites and the use of higher frequencies have made it possible for many people to receive direct signals from the sky. At the beginning of the 21st century, more than one hundred million European houses were able to watch television programs transmitted by satellites, either by direct reception or through cable distribution systems. All of these were made possible through the rapid development in innovative technologies and techniques such as satellite and component miniaturization. These new technologies has allowed easy and relatively inexpensive access to space by multiplexing multiple cube/micro/nano-satellites into a single launcher.

3 LEO MICROSATELLITE CONSTELLATION UTILIZING SATELLITE HELLAS SAT-5 AS RELAY

3.1 Introduction

This thesis analyzes and presents the capabilities of a satellite ground station communicating with a low-earth-orbit (LEO) satellite constellation via the geostationary satellite, Hellas-Sat 5, which acts as a relay. Toward this end, the design and implementation steps concerning the LEO constellation as well as the overall satellite communication system which are given in detail, i.e., the LEO-GEO-Earth station communication path. Moreover, the software platform that has been used to design, implement, and assess the performance of the considered space-to-ground communication system is presented. The simulations were carried out using the AGI Systems Tool Kit (STK v.12.2[1]) software with the use of a requested 4-month prolonged free license. Additional on-line tools will be used for calculating the uplink and downlink budgets as well as the structure of the satellite constellation.

[1] https://www.agi.com/new-stk/stk-12-2 - AGI Systems Tool-Kit version 12.2

3.2 AGI Systems Tool-Kit

Systems Tool-Kit (formerly Satellite Tool Kit), often referred to by its initials STK, is a multi-physics software application from Analytical Graphics Inc, which enables engineers and scientists to perform complex analyses of ground, sea, air, and space platforms, and to share results in one integrated environment. The STK interface is a standard GUI display with customizable toolbars and dockable maps and 3D graphic windows. All analysis can be done through mouse and keyboard interaction.

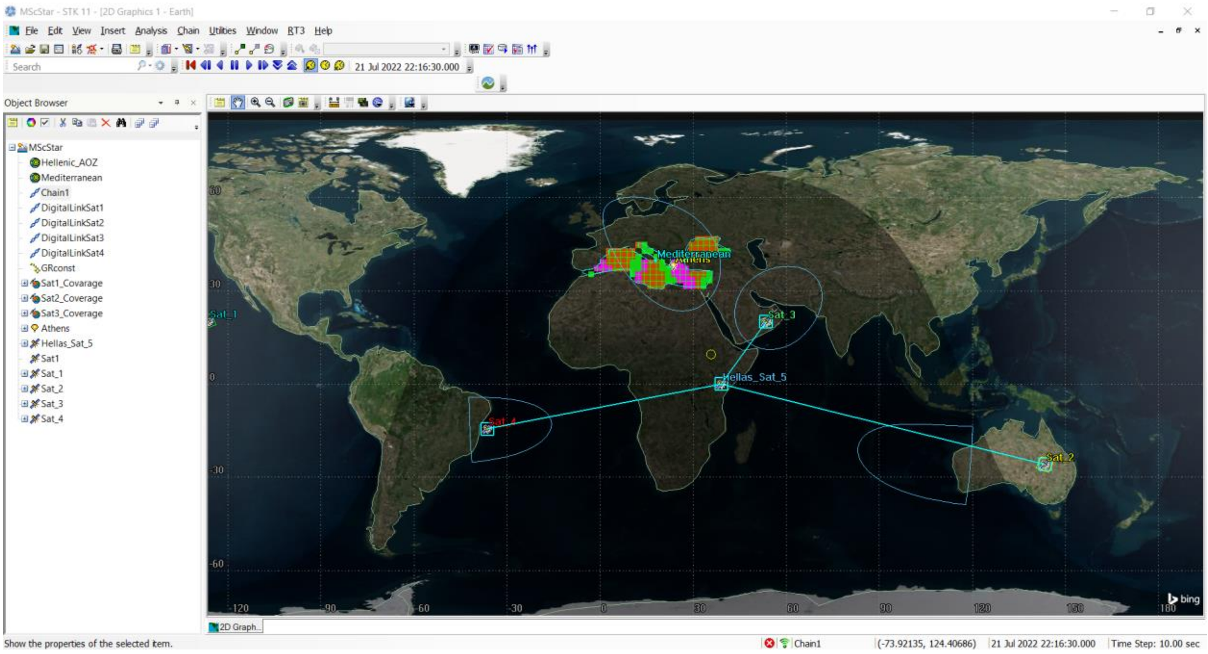


Figure 1: STK v12.2 Platform

3.3 Methodology

Regarding the implementation of this thesis, the following steps are considered in order to formulate a complete scenario of a satellite communication system:

Acknowledgements

Walker LEO Satellite Constellation / GEO Satellite Design and Implementation (STK Analysis)

LEO Satellite Communicational System Configuration (STK)

GEO Satellite Communicational System Configuration (STK)

LEO Constellation Satellite Link Budget Analysis (STK & Link Budget Calculator)

GEO Satellite Link Budget Analysis (STK & Link Budget Calculator)

Overall System Presentation (STK)

3.4 Acknowledgements

Low Earth Orbit satellite constellations are very popular nowadays since they may offer a significant number of services and technologies rather than using a single satellite in a stationary or single orbital position. Earth observation, fire detection, next generation communication systems as well as many other satellite technologies uses LEO constellations in order to facilitate these services. The LEO position is ideal for a broad spectrum of services which takes advantage of the low altitude, the cost efficiency, and the simplicity regarding placing into orbit. That is the reason several world-wide companies design and implement LEO constellations for their operations.

On the other hand, LEO position also has several disadvantages such as increased velocity (high Doppler effect), short passes, designing limitations due to spacecraft size and overpopulated orbits. In order to overcome these disadvantages, certain technologies are used as well as alternative methods. Specifically talking about designing limitations, this thesis will analyze the possibility of implementing a LEO constellation with the use of microsatellites which inherently are small sized spacecrafts and with limitations regarding the communication capabilities (e.g., limited power budget).

Microsatellites face a significant challenge when it comes to their communication equipment's ability to overcome atmospheric / propagation losses and communicate with Earth's stations as well as the limited communication window time offered. However, using a relay geostationary satellite offers an appealing solution since it provides the necessary high data-rate telecommunication equipment and offers a constant view of the same area on Earth's surface. Furthermore, with an altitude of approximately 36,000 km, the geostationary spacecraft offers significant access times between the LEO constellation and itself (i.e., providing full time coverage).

A geostationary data relay receives the data from LEO satellites via inter-satellite links and forwards it to the ground. The main advantage, compared to direct LEO-Ground links, is the large Field of View (FOV). A GEO satellite can cover up to 69% of the LEO satellite orbits. A single geostationary data relay is enough to provide uninterrupted real-time access to LEO satellite constellation. Only one single ground station is required for transferring the data from each LEO satellite to the ground. Due to the extended contact times, data rates of 30 - 100 Mbit/s on the inter-satellite links (ISLs) are sufficient to serve download volumes of several Terabytes per day [1]. This rate exceeds current downlink requirements with margins. With the use of the STK software, a summarize of the access times will be presented in order to better understand the use of ISLs between LEO constellation and GEO satellite.

In this particular thesis, only the necessary equipment for the ISL between the LEO constellation and the GEO relay will be studied. The actual payload of the LEO microsatellites will be given in examples of possible use cases.

3.5 Walker LEO Constellation / GEO Satellite - Design and Implementation

In the frame of this thesis, the design objective of the LEO constellation is to maximize the coverage area of the Mediterranean Sea, which is considered the area of interest, within a 24-hour time period. The coverage would also be possible to be computed for global coverage but for this thesis, is out of the scope.

AGI STK software offers a powerful tool for analyzing and implementing trade studies regarding the ideal selection of orbital parameters. This tool is called «Analyzer Plugin» for which a special request was made for a 4-month prolonged free license in addition to the other STK components.

In order to commence the implementation of this thesis, a custom area was designed called “Mediterranean”. A basic circular orbit satellite was inserted with a J2Perturbation propagator, 60° inclination and 6.678 km semi-major axis. This represents the seed-satellite concerning the trade study as well as the design of the walker constellation.

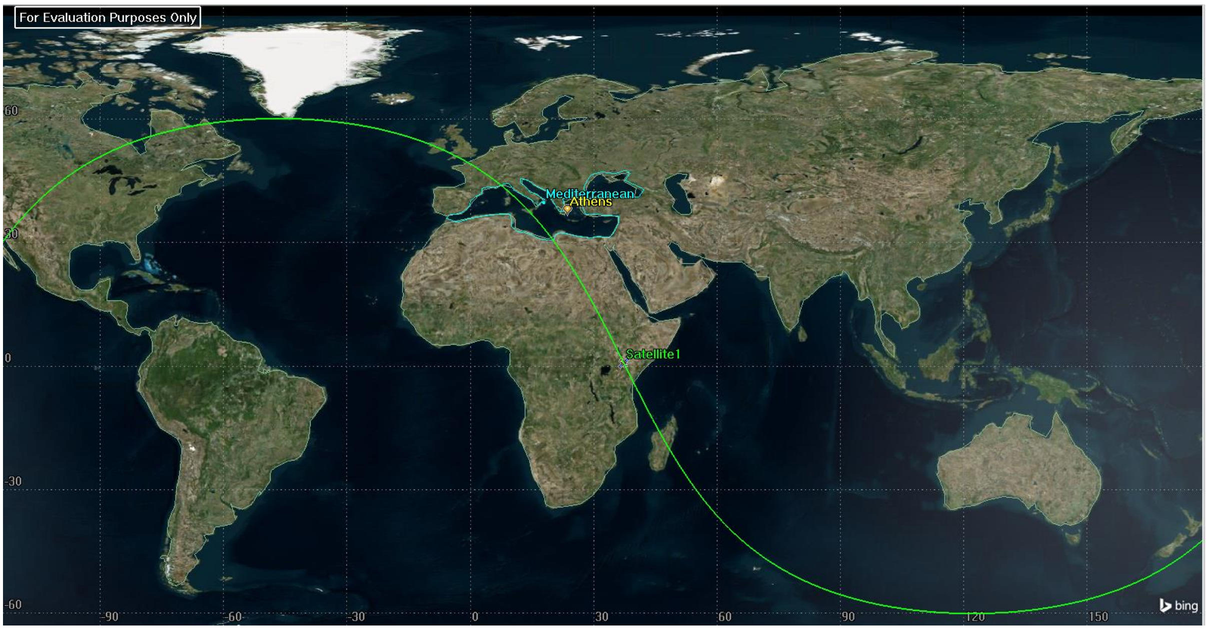


Figure 2: Seed-Satellite and custom area

The first step is to compute Satellite 1 (i.e., the seed satellite) access times to the Mediterranean area (seconds). This procedure will produce the total duration (sec) of the access time of Satellite 1 at the area of interest.

Following the completion of Satellite 1 access times for 60^o inclination and 6.678 km semi-major axis, a first stage trade study was implemented. STK Analyzer plugin is a tool which performs a trade study between one or several variable inputs and exports a specific output. For this specific thesis, a trade study performs a comparison between different inclinations and semi-major axis and their effect in total access times. The goal is to find the optimum combination of inclination and altitude in order to maximize the access time duration.

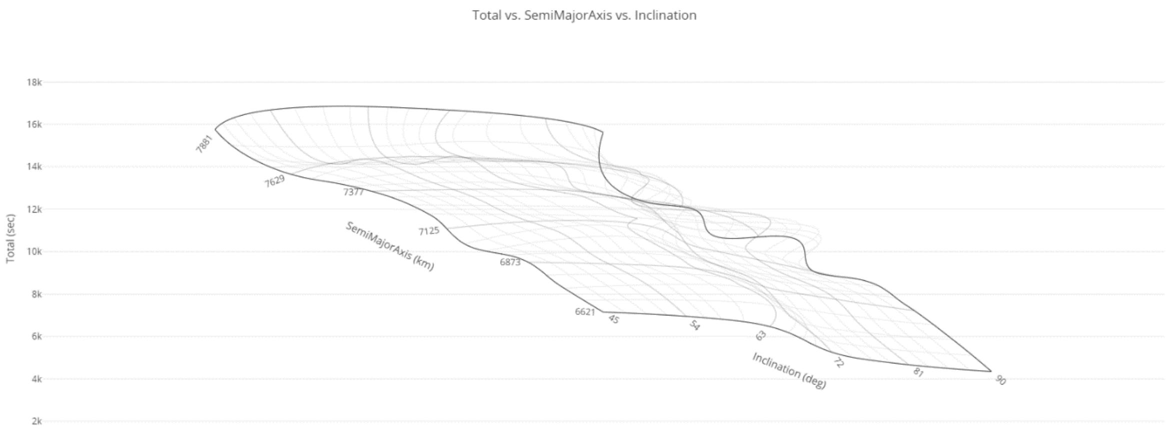


Figure 3: Carpet plot Inclination vs semimajor axis and access times

The trade study calculated the ideal combination between inclination and semi-major axis which achieves the highest access time for Satellite 1. Based on the above, inclination of 60^o deg and semi-major axis equal to 7.881 km (1500 km) will produce a total access time of 16.861 seconds (4.68 hours) in a 24-hour period.

After the fine-tunning of Satellite 1 orbital parameters based on the above trade study, a coverage definition should be carried out in order to determine the coverage of the Mediterranean Sea. For this reason, the coverage definition tool of the STK will is used to determine the percentage coverage of the area of interest for a 24-hour period.

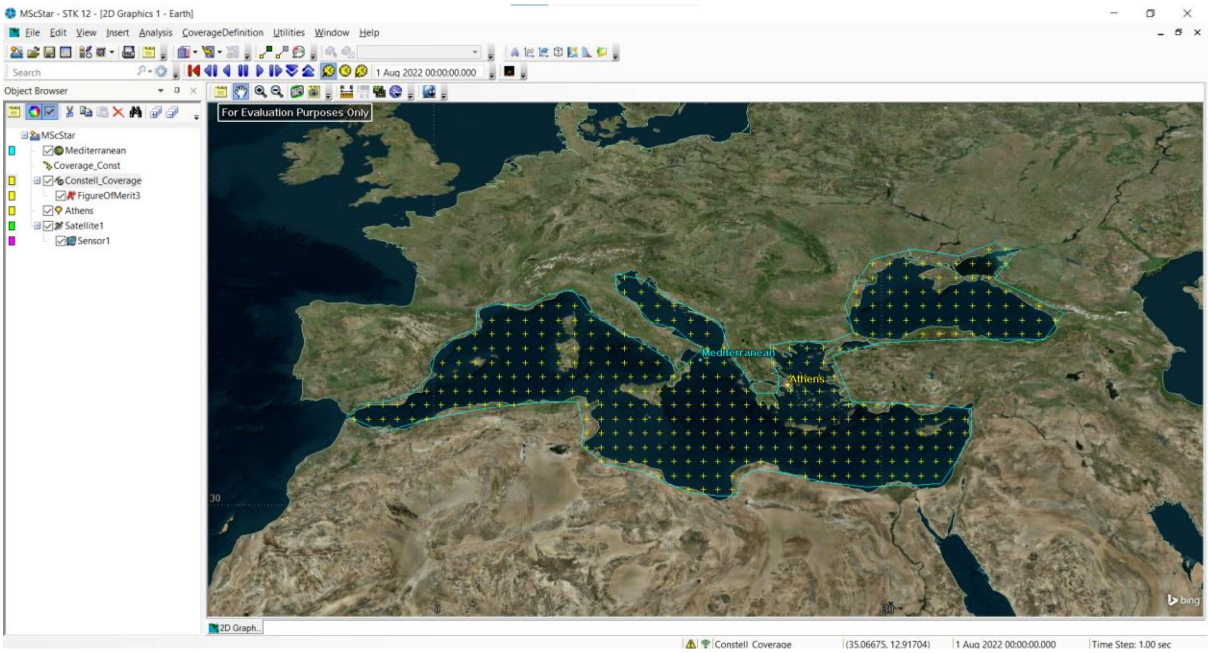


Figure 4: Mediterranean area coverage definition

As the optimum seed-satellite orbital parameters have been defined, the walker constellation should be calculated following the same approach as before. This time a two-step trade study will be performed in which the first will determine the walker constellation based on the maximum coverage percentage of the designated area and the second based on the maximum access time regarding the GEO satellite in order to design the ISL (inter-satellite link).

3.5.1 Walker Constellation for Maximum Coverage and Access times

Using the STK Analyzer module, the software is able to determine the designated area coverage percentage. For this specific thesis, a standard walker constellation is used at 60° inclination (according to the previous trade study), with four planes each of one containing one satellite (60° 4/4/0 walker constellation), RAAN spread between planes $\frac{360}{4 \text{ planes}} = 80^{\circ}$ and 20° true anomaly differences between satellites. Having simulated several combinations with different true anomaly angles, the 20° angle offers a 94.6% percent coverage of the designated area.

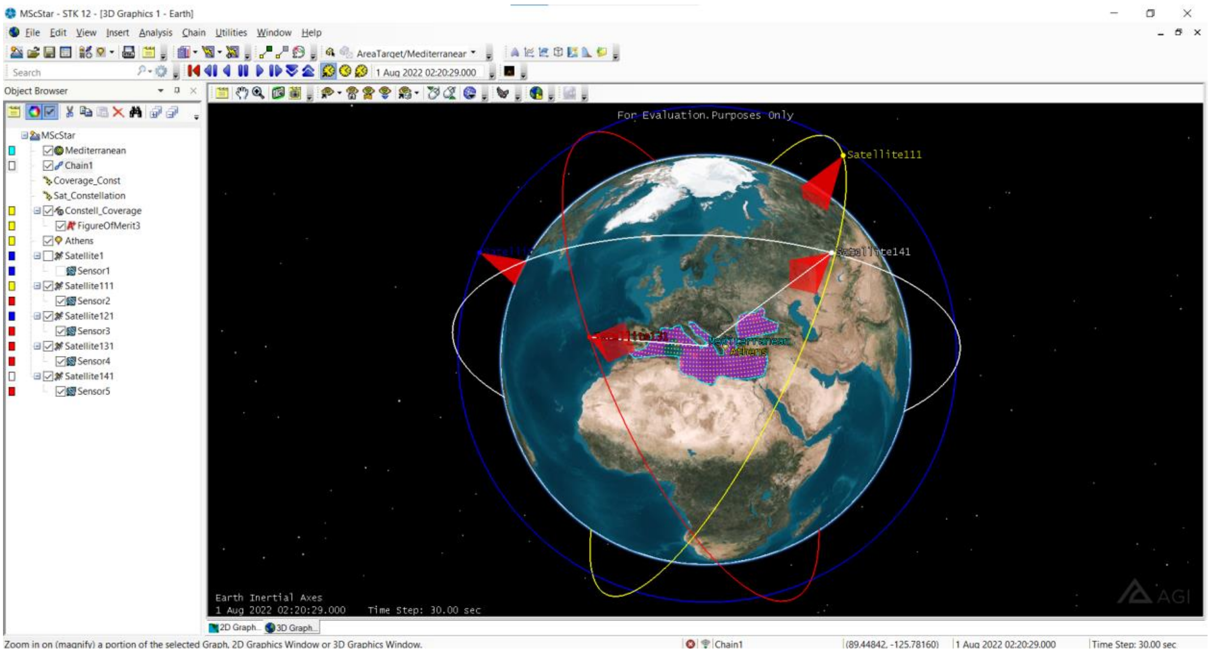


Figure 5: Satellite constellation 3D map

3.5.2 Geostationary Satellite

As mentioned in this thesis title, the upcoming Hellas Sat 5 geostationary satellite will function as a relay station between the LEO constellation and the Earth station. A standard geostationary satellite will be loaded on STK using the same orbital parameters as the Hellas Sat 4 (Altitude 35.794.2 km, inclination: 0^o, location: 39^o East).

The next step is to compute the access times between the constellation satellites and the GEO satellite in order to determine the number of hours that the satellites are visible by the GEO satellite. Using the built-in tool of the STK software, it is possible to calculate the access times. The first step is to add the constellation satellites in a constellation group.

Next, the chain element is going to be used in order to calculate the access times. In the chain element it is required to assign the constellation group as well as the Hellas Sat 5. Applying the above changes and performing the compute access command in chain element, the software is determining the times in which one or more LEO satellites are visible from and to the GEO satellite.

As mentioned before, only the configuration concerning the communication aspects for the LEO constellation Inter-Satellite-Link with the GEO satellite, is analyzed in the frame of this thesis. This means that a link budget analysis is performed in order to calculate the required telecommunication equipment for the ISL links. The payload part of the LEO constellation will be presented with potential use-cases.

For the implementation of the satellite constellation ISL, it is necessary for the microsatellite to be equipped with an antenna which is going to be targeted (when visible) at the receiver of the GEO satellite. This means that the antenna should be able to perform tracking procedures in order to be aligned with the GEO receiver. After that, the communications equipment must be powerful enough to be able to achieve sufficient data rates through the ISL. An important observation showing why ISLs are so widely adopted in modern satellite networks, is the fact that an ISL is occurring only in an environment with no atmosphere (space). This means that the necessary equipment, which is going to be installed on the microsatellite, is not required to be able to overcome the physical losses of the Earth's atmosphere. This is exactly why the use of a relay station is much easier to implement, because a GEO satellite is large enough to accommodate much more comprehensive communicational equipment, something that is not an option with microsatellites.

3.6 GEO Satellite Communication System Configuration (STK)

For this thesis implementation, the GEO satellite communications schema entails four separate servomotors and receivers, each of one pointing at a specific LEO microsatellite. In actual configurations, this method is unfeasible because a GEO satellite cannot contain separate antennas allocated to a single satellite. A specific configuration must be set to accommodate the ISL's between LEO and GEO satellites such as specific time slots, switching equipment as well as decision making algorithms base on range, azimuth angle and elevation angle. These special configurations are out of this thesis's scope.

As the ISL's with the LEO constellation are set, another connection should be established between the GEO satellite and the Earth station. For this purpose, as before, a sensor will be inserted on the ground facility which will function as a servomotor tracking the GEO satellite, a receiving antenna and lastly, a transmitting antenna on GEO satellite which will be facing the Earth station.

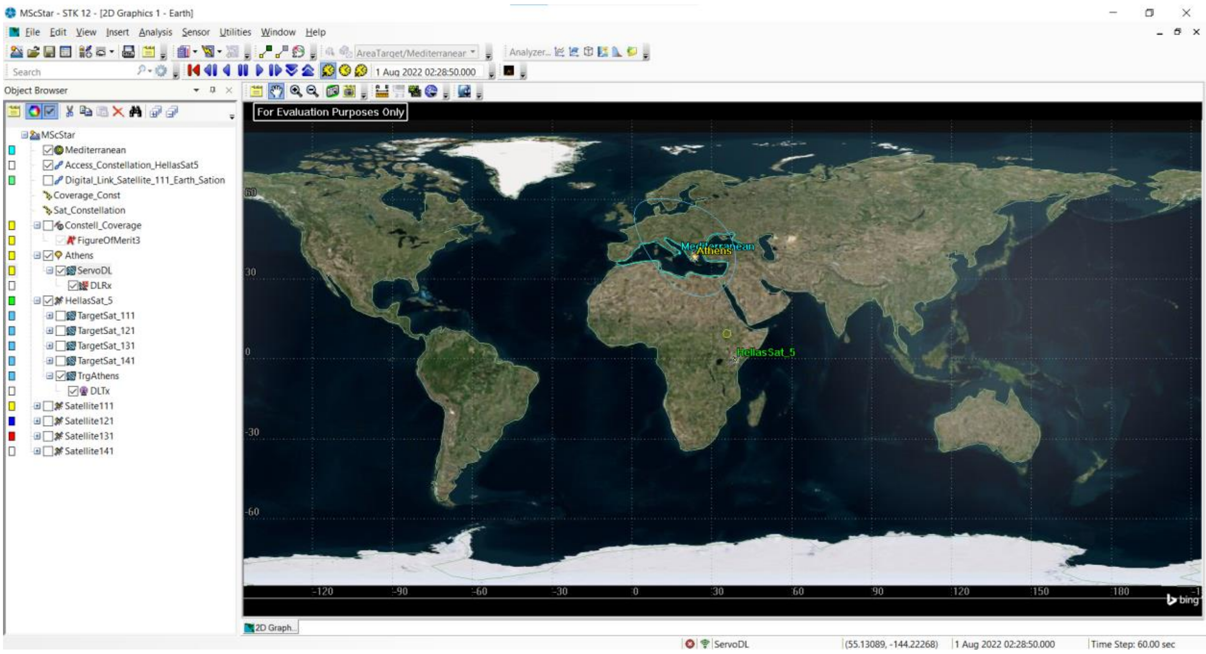


Figure 6: GEO to Earth station connection

These elements conclude the implementation of the ISL's between the constellation and the GEO satellite as well as the connection between Hellas Sat 5 and the ground station. In the following chapter, a complete link budget analysis is presented in detail regarding the ISL and the GEO downlink.

3.7 LEO Constellation Satellites Link Budget Analysis (STK & Link Budget Calculator)

Having analyzed all the basic communication equipment of the LEO satellite constellation, this subsection introduces the characteristics of its communication equipment used in detail. Also, a comprehensive link budget analysis between the microsatellite and geostationary satellite is provided. Furthermore, the potential use cases for the microsatellites and their payloads are presented and discussed.

The first step is to determine the ISL's characteristics regarding the communication between the microsatellites and the geostationary satellite. For this purpose and in order to cover all the potential use cases, the Digital Video Broadcasting version 2 (DVB-S2) protocol is utilized for the communication. Table 1 below illustrates all the specific values that will be inserted in STK software at the ULTxSat_111 element which is the transmitter of Satellite 1 in the constellation as well as the corresponding receiver on the geostationary satellite (RxSat_111).

Table 1: LEO microsatellite to GEO satellite parameters [5]

Microsatellite Tx Equipment	
Frequency	11.2 GHz X-Band
Transmitted Power	15 dBW
EIRP	39 dBW
Bandwidth	36 MHz
Antenna Diameter	0.2 m
Antenna Efficiency	65%
Polarization	Horizontal
Modulation & Coding	QPSK 2/3
Filter	Root Raised Cosine (RRC)
Roll Of Factor (ROF)	0.25
Transmit Symbol Rate	29.7 Mbaud
Target Data Rate	20 Mbit/s (Data transmission)
Geostationary Rx Satellite Equipment	
Antenna Diameter	1 m
Antenna to LNA line loss	2 dB
LNA Gain	10 dB
LNA to receiver line loss	2 dB
Modulation & Coding	QPSK 2/3
Link Requirements	
Minimum Eb/No Required	5 dB[2]
SNR Threshold (C/N)	6.6 dB[3]
Target BER	10^{-7} or less[4]

2 https://www.etsi.org/deliver/etsi_i_ets/300400_300499/300421/01_60/ets_300421e01p.pdf - ETS 300 421
3 http://www.satbroadcasts.com/news,81,Minimum_carrier_to_noise_ratio_values_CNR,_CN_for_DVB_S2_system.html - Minimum carrier to noise ratio values (CNR, C/N) for DVB-S2 system
4 https://tech.ebu.ch/docs/techreview/trev_300-morello.pdf - DVB-S2 - ready for lift-off

All the above parameters are used in standard DVB-S2 configuration in satellite communications. Furthermore, due to the fact that STK platform does not offer the QPSK 2/3 modulation as an option, this particular modulation type it is implemented and inserted in STK manually (Figure 8).

The next is to insert all the above parameters in the STK platform in order to calculate the link budget analysis. In order for STK software to be able to produce the link budget analysis, a chain element must be inserted and select the LEO satellite transmitter and the GEO satellite receiver as the main study elements.

The next and last step is to right click the chain element, report and graph manager and select the Digital Repeater Comm Link. The following Figure 7 illustrates the result:

Time (UTCG)	Xmtr Power1 (dBW)	Xmtr Gain1 (dB)	EIRP1 (dBW)	Prop Loss1 (dB)	Rcvd. Frequency1 (GHz)	Rcvd. Iso. Power1 (dBW)
1 Aug 2022 00:00:00.000	15.000	25.5407	39.983	204.4159	11.199872	-164.435

Flux Density1 (dBW/m^2)	g/T1 (dB/K)	C/No1 (dB*Hz)	Bandwidth1 (kHz)	C/N1 (dB)	Eb/No1 (dB)	BER1
-121.995286	14.895994	85.059801	36000.000	9.4968	12.0495	6.139707e-13

Figure 7: Link budget analysis ISL microsatellite to geostationary satellite

Considering Table 1 target parameters, it is shown that the ISL between the LEO satellite and the GEO relay, achieves those required parameters. Specifically, the achieved BER is 6.13×10^{-13} which is less than the targeted one as well as both Eb/No and C/N values are greater than the designated thresholds. Another important note is that the above performance is calculated on the optimum azimuth and elevation position between LEO satellite and GEO satellite. At the lowest position of the LEO satellite the performance drops at 1.77×10^{-7} which is also sufficient enough for this specific link. The above note illustrates the importance of the designing process of a link budget in order to cover all the possible positions of the satellite.

The above procedure concludes the study for the ISL between the LEO microsatellite and the GEO satellite. In the following subsection, a link budget analysis will be studied from the GEO satellite to the ground station.

3.8 GEO Satellite Link Budget Analysis (STK & Link Budget Calculator)

For the geostationary downlink to the ground station, a real world link budget analysis is considered and inserted in STK in order to simulate an actual satellite link connection. The link budget was obtained from Hellas Sat corporation and simulates a connection for data transmission.

Due to the fact that STK platform does not offer QPSK modulation with 2/3 coding rate and with the use of the same technique mentioned before (MATLAB Bit Error Rate Analysis – Figure 8), the BER vs Eb/No relation was derived in order to insert it manually on STK.

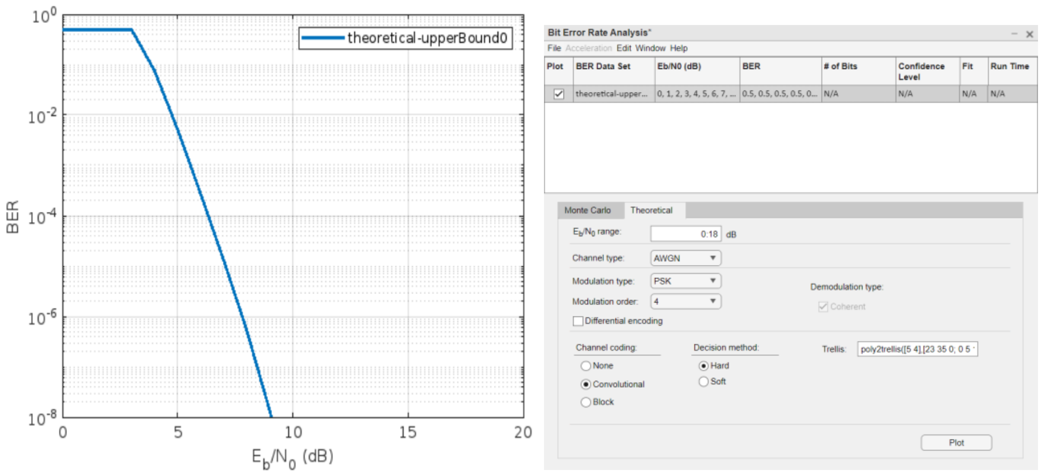


Figure 8: QPSK modulation with 2/3 coding rate bit error rate analysis tool

The following Table 2 illustrates all the parameters that was inserted in STK platform:

Table 2: GEO Satellite to Earth station parameters [5]

Geostationary Satellite Tx Equipment	
Frequency	11.6 GHz X-Band
Transmitted Power	8 dBW
EIRP	42 dBW
Bandwidth	36 MHz
Antenna Diameter	0.8 m
Antenna Efficiency	65%
Polarization	Horizontal

Modulation & Coding	QPSK 2/3
Filter	Root Raised Cosine (RRC)
Roll of Factor (ROF)	0.05
Transmit Symbol Rate	16 Mbaud
Target Data Rate	20 Mbit/s (Data transmission)
Earth Station Rx Equipment	
Antenna Diameter	1.2 m
Antenna to LNA line loss	2 dB
LNA Gain	5 dB
LNA to receiver line loss	2 dB
Modulation & Coding	QPSK 2/3
Link Requirements	
Minimum Eb/No Required	3.6 dB
SNR Threshold (C/N)	5.5 dB
Target BER	10^{-7} or less[5]

The last step is to insert all the above values in STK platform. As mentioned before, another chain element is required to be created and select the GEO satellite Tx system and the Earth station Rx system in order to produce the link budget.

Figure 9 below illustrates the link budget analysis of the STK platform regarding the link between GEO satellite and the ground station:

Time (UTCG)	Xmtr Power1 (dBW)	Xmtr Gain1 (dB)	EIRP1 (dBW)	Prop Loss1 (dB)	Rcvd. Frequency1 (GHz)	Rcvd. Iso. Power1 (dBW)
1 Aug 2022 00:00:00.000	8.000	37.8867	42.424	205.2330	11.600000	-162.809

Rcvd. Iso. Power1 (dBW)	Flux Density1 (dBW/m^2)	g/T1 (dB/K)	C/No1 (dB*Hz)	Bandwidth1 (kHz)	C/N1 (dB)	Eb/No1 (dB)	BER1
-162.809	-120.064512	16.784506	83.574288	36000.000	8.0113	10.5640	6.344040e-09

Figure 9: Link budget analysis STK platform

As Figure 9 shows, the achieved bit error rate of the link is 6.34×10^{-9} which is

[5] https://tech.ebu.ch/docs/techreview/trev_300-morello.pdf - DVB-S2 - ready for lift-off

more than enough than the targeted BER of 10^{-7}. Furthermore, both C/N and Eb/No values are greater than the desired one's presented in Table 2.

3.9 Overall System Presentation (STK)

Once the required communication equipment is determined and the necessary requirements are met to achieve the desired performance for each link, the overall system's performance should be evaluated. Hence, an overall link budget should be calculated to assess the performance of the complete communication link from the microsatellite to the Earth station via the GEO satellite Hellas-Sat 5 which acts as a relay. In order to perform an overall link budget, a chain element is created and named as «Digital_link_LEO_RelayGEO_Earth_Station». The difference now is that both ISL and downlink to Earth station will be inserted into the chain element in order to calculate the complete link budget. For the individual links and with the use of the chain element, the overall system link budget is calculated as shown in Figure 10:

Time (UTCG)	Xmtr Power1 (dBW) Xmtr Power2 (dBW)	Xmtr Gain1 (dB) Xmtr Gain2 (dB)	EIRP1 (dBW) EIRP2 (dBW)	Prop Loss1 (dB) Prop Loss2 (dB)	Rcvd. Frequency1 (GHz) Rcvd. Frequency2 (GHz)	Rcvd. Iso. Power1 (dBW) Rcvd. Iso. Power2 (dBW)
1 Aug 2022 00:00:00.000	15.000 8.000	25.5407 37.8867	39.983 42.424	204.4159 205.2330	11.199872 11.600000	-164.435 -162.809

Flux Density1 (dBW/m^2) Flux Density2 (dBW/m^2)	g/T1 (dB/K) g/T2 (dB/K)	C/No1 (dB*Hz) C/No2 (dB*Hz)	Bandwidth1 (kHz) Bandwidth2 (kHz)	C/N1 (dB) C/N2 (dB)	Eb/No1 (dB) Eb/No2 (dB)	BER1 BER2	BER Tot.2
-121.995286 -120.064512	14.895994 16.784506	85.059801 83.574288	36000.000 36000.000	9.4968 8.0113	12.0495 10.5640	6.139707e-13 6.344040e-09	6.344654e-09

Figure 10: Overall link budget analysis STK

Figure 10 illustrates the complete link budget of the system. The achieved end-to-end BER is 6.34×10^{-9} while the worst one is 1.84×10^{-7} (Figure 11). This means that the system will perform normally even in the lowest azimuth and elevation positions.

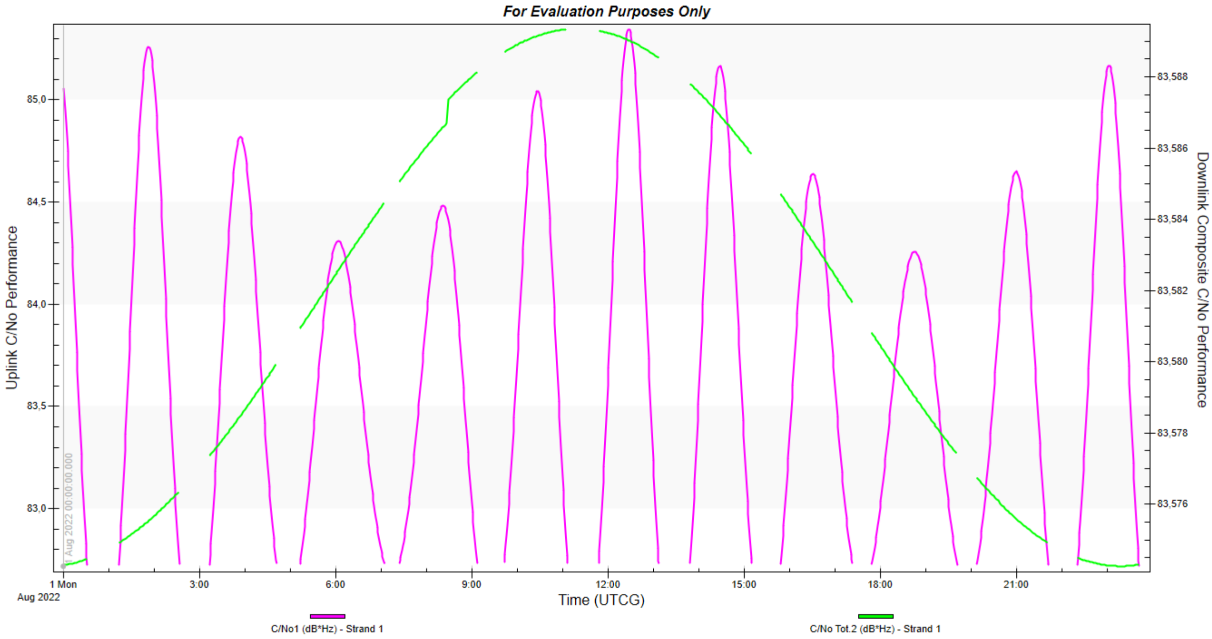


Figure 11: Total C/N value changes over time

The last step of the thesis is to configure all four LEO satellites that forms the constellation with the parameters specified in the previous sections in order to illustrate the complete system as shown in Figures 12 & 13:

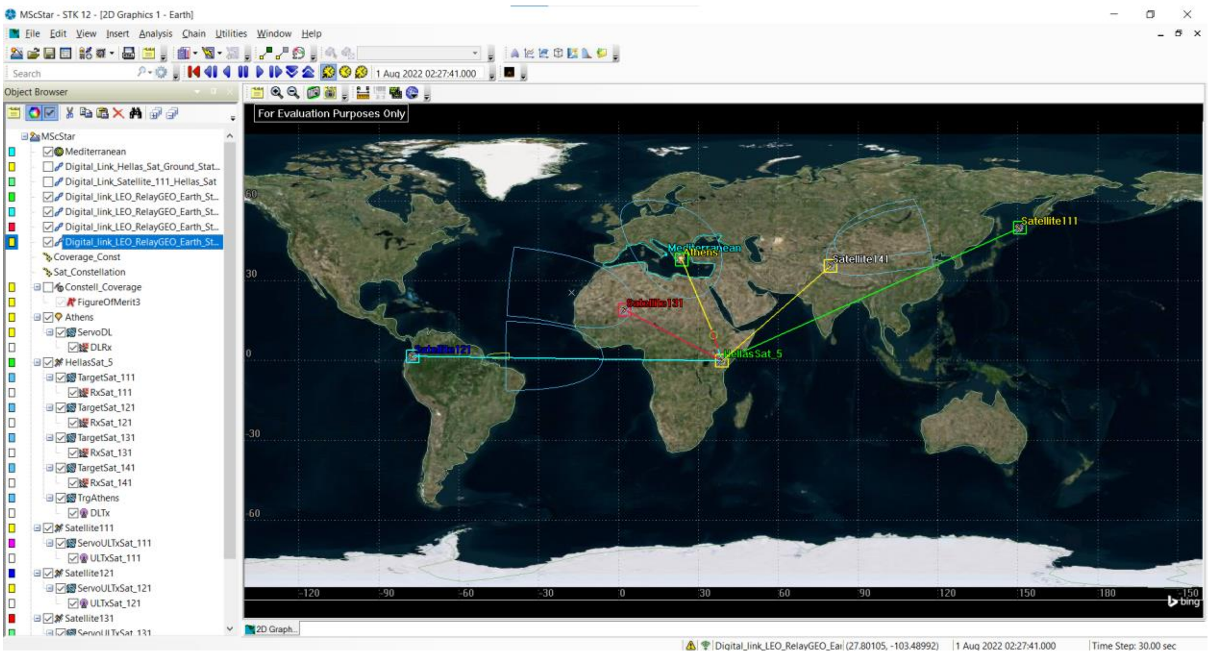


Figure 12: Complete LEO constellation with relay station STK 2D

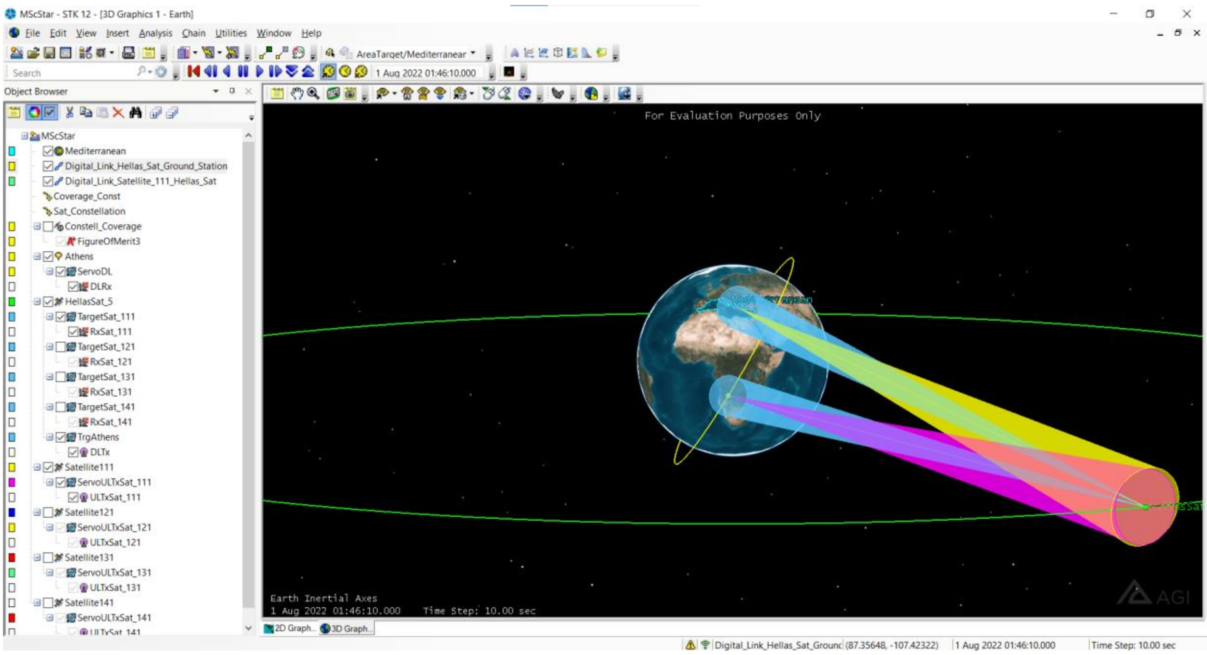


Figure 13: Complete LEO constellation with relay station STK 3D

4 CONCLUSIONS

4.1 General

Typically orbiting at low Earth orbit, satellite constellations provide the required data with low latency (downlink and uplink), particularly valuable when immediate responses are critical. Compared to geostationary satellites, swarms of small units (up to 500 kg) are cheaper and faster to deploy.

This particular thesis analyzes the theoretical approach to design and implement a LEO constellation with the use of a geostationary satellite as a relay node. By combining a LEO constellation with a GEO satellite as a relay, we can leverage the advantages of both systems. The LEO constellation offers global coverage and cost-effectiveness, while the GEO satellite provides high data transmission rates and continuous coverage on a specific region of interest.

After the completion of all the individual configuration steps for the implementation of this thesis, the following conclusions are drawn:

- The LEO orbital parameters are a critical component in the design of a complete LEO satellite communication network. The engineers should consider the necessary equipment in terms of implementation costs, predicted lifespan, and the purpose / application of the constellation system. Given the fact that due to the increased number of satellites in each orbital plane, the orbital position should be selected carefully in order to avoid possible hazards.
- Each individual LEO satellite, having in mind that it is a microsatellite, should be designed in such way that meets all the necessary communicational requirements in order to provide reliable communication and be able to be fitted in small compartments.
- The geostationary satellite requires comprehensive communication equipment and must be designed to accommodate current and future technological advancements. Additionally, it should be adaptable for future satellite links. All the communicational links should be designed and evaluated for the worst-case scenario. This means that in a comprehensive link budget analysis, all of the atmospheric losses should be considered in order to be able to predict the link's performance under the worst possible conditions. Having also the nominal link operation, it is useful to derive the link margin and the achieved availability.
- ISLs are a relatively recent technology and is complicated to implement, especially in microsatellites mainly due to increased Doppler effect.
- The design of a microsatellite with ISL must consider the fact that it should contain different equipment for the ISL connection and different for the payload, something that has certain restrictions due to limited space.

4.2 Advantages of using a geostationary satellite as relay

As mentioned before, this thesis's purpose is to demonstrate the benefits of using ISL's as well as relay stations. In this case, a combination of both occurred resulting in an extremely resilient and powerful satellite link for transmitting data at very high data rates.

Utilizing a GEO satellite as a relay offers several advantages such as:

As it is at greater height, it covers larger geographical area.

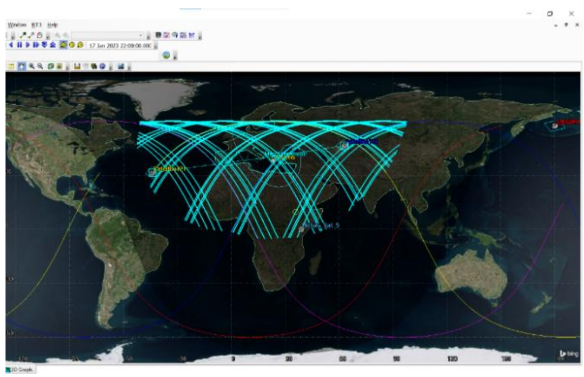
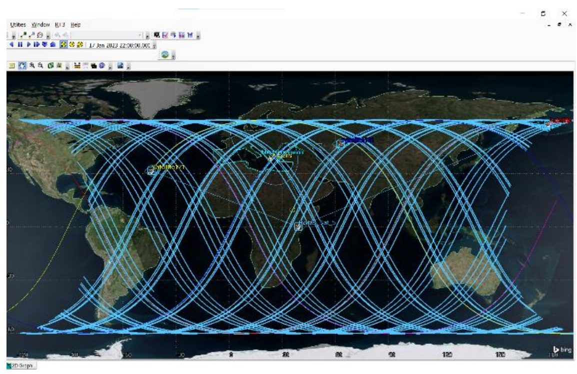
- Satellites are visible for 24 hours continuously from single fixed location on the Earth.
- It is ideal for broadcasting and multi-point distribution applications.
- Ground station tracking is not required as it is continuously visible from earth from fixed location.
- Almost no doppler shift and hence less complex receivers can be used for the satellite communication.

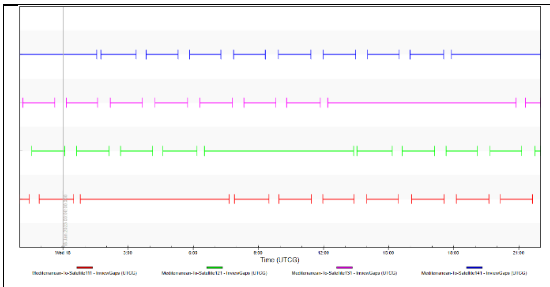
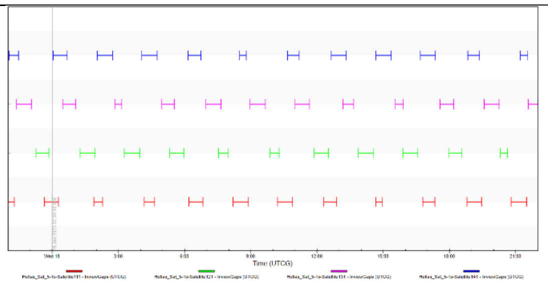
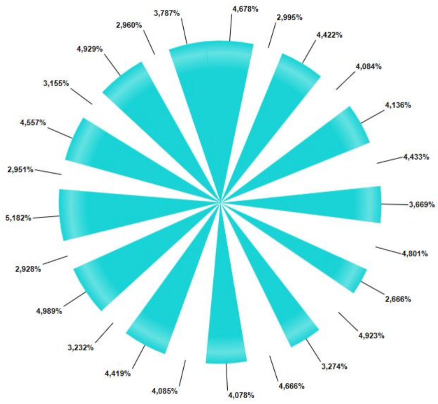
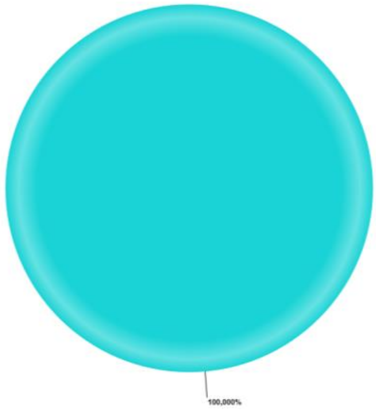
Through the use of STK's tools, several reports and graphs were produced in order to show the differences between a microsatellite constellation without a relay station comparing to using one.

4.3 Final comparison

Table 3 below presents the final comparison between the two cases regarding the access times as well as the link budget performance, one without the use of the geostationary relay and one with the use of Hellas Sat 5 as the data relay node.

Table 3: Final parameters comparison

Without relay station	With relay station
Access Time	
	

<table>
<tr><td></td><td></td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>4.5 hours per day per satellite</td><td>17.3 hours per day per satellite</td></tr>
<tr><td colspan="2">Link Budget</td></tr>
<tr><td>
Xmtr Power1 (dBW) | Xmtr Gain1 (dB) | EIRP1 (dBW)
15.000 | 25.5407 | 39.983

C/N1 (dB) | Eb/No1 (dB) | BER1
4.9466 | 7.4993 | 1.505172e-04
</td><td>
Xmtr Power1 (dBW) | Xmtr Gain1 (dB) | EIRP1 (dBW)
Xmtr Power2 (dBW) | Xmtr Gain2 (dB) | EIRP2 (dBW)
15.000 | 25.5407 | 39.983
8.000 | 37.8867 | 42.424

C/N1 (dB) | Eb/No1 (dB) | BER1 |
C/N2 (dB) | Eb/No2 (dB) | BER2 | BER Tot.2
9.4968 | 12.0495 | 6.139707e-13 |
8.0113 | 10.5640 | 6.344040e-09 | 6.344654e-0
</td></tr>
</table>

As a conclusion, a combination of:

- the increased access times between the constellation and the geostationary relay station,
- the limited atmospheric losses due to inter-satellite links,
- the benefits of a geostationary satellite regarding the communicational capabilities,
- the extended coverage times that a microsatellite constellation is offering,
- the link budget performance,

is able to establish a resilient, reliable and exceptionally high throughput communication link with the use of a geostationary relay station and a constellation consisting of four low-Earth orbit microsatellites.

REFERENCES

[1] D. Vouyioukas, Satellite Communications: Technologies, Systems and Applications, Kallipos, Open Academic Editions, 2015.

[2] E. Dr. Lutz , M. Werner and A. Dr. Jabn, Satellite Systems for Personal and Broadband Communications, Springer, 2000.

[3] J. J. Sellers, Understanding Space - An Introduction To Astronautics, 2nd ed., McGraw-Hill Primis Custom Publishing, 2003.

[4] G. Maral and M. Bousquet, Satellite Communications Systems: Systems, Techniques and Technology, 5th ed., John Wiley and Sons, 2010.

[5] T. Pratt, C. W. Bostian, and J. E. Allnutt, Satellite Communications, John Wiley and Sons, 2nd ed., 2009.

Evaluation of the Effect of Different Tracker-driven Direction Sources on Continuous Artificial Locomotion in VR

Theodoros E. Mandilaras

ABSTRACT

This research evaluated different direction sources with Continuous Artificial Locomotion in Virtual Reality (VR). In CAL, the user moves continuously in a virtual environment (VE) by pressing direction buttons on the controllers. However, the direction of the movement, in most cases, is defined either by the headset or by one of the controllers, which may feel unnatural. In this study, we propose two additional direction sources, one based on the direction of the user's hips and another based on the direction of their feet. We utilized trackers placed on these body parts to implement these direction methods. We evaluated the four methods in terms of performance, preference, motion sickness, and presence. We designed and implemented a VE and conducted a user study with 24 participants to evaluate. The users had to fulfill three tasks with each method: to navigate in a zigzag environment, then to cross a corridor while counting some dots positioned on the walls of that corridor, and finally to complete a mini-game by placing colored cubes in the desired area. Data were collected both quantitatively and qualitatively. The results of our study indicate that hip and feet methods are the optimal selection in terms of performance for executing different virtual tasks, that motion sickness is not affected by the different direction sources, and that the users preferred the feet and headset methods.

Subject area: Human Computer Interaction

Keywords: Virtual Reality, Locomotion, Interaction devices

ADVISOR

Maria Roussou, Assistant Professor NKUA

1 PREFACE

Navigation in Virtual Reality (VR) is one of the most widely used interaction components in VR since some navigation is required in most cases. It can be defined as a way for users to move through virtual worlds.

Providing the ability to comfortably navigate in a VE and maintain a high sense of presence is considered paramount for VR. A poorly designed locomotion technique (LT) may distract the user, reduce immersion and introduce motion sickness [1]. Consequently, in the past decades, researchers have been experimenting with numerous locomotion techniques, exploring their strengths and weaknesses and improving them constantly [2]. Physical locomotion techniques [5], like Real Walking (RW), despite being the most natural [3][4], cannot be scaled due to the limited physical space. Thus, researchers need to switch to Alternative Locomotion Techniques (ALTs) to navigate beyond those confines. One of the most commonly used is Continuous Artificial Locomotion (CAL). In this locomotion technique, the user continuously moves through the space by pressing directional buttons, sticks, or pads to move the viewpoint in the chosen direction. However, this technique is highly related to causing VR sickness, induced by the mismatch in sensory information between the visual and other senses. It can lead to headaches, blurred vision, dizziness, disorientation, or nausea [1].

Furthermore, one other issue with CAL techniques is the source of the direction that the user moves to. Most of the existing literature about CAL techniques mainly focuses on how to move the user and not on the direction of movement, which is defined as the user's "forward." Most of the time, the moving direction is driven by the orientation of the Head Mounted Display (HMD), which in a virtual world represents the user's viewpoint. Thus, the users will move where their head is looking at. That technique is usually called "gaze-steering" [4]. Another approach is "hand-steering," in which the direction is driven from the orientation of one of the hand-held controllers. In this case, the user moves toward where the controller points. Both techniques do not require extra hardware since the tracking mechanism of the VR system already tracks these devices. However, those approaches do not directly relate to how humans move in real life. Movement in the physical world is decoupled from the direction your head and hands are pointing, so you can look around and use your hands as you move. As a result, those approaches may contribute to the created motion sickness that the user may feel. Moreover, overloading the headset and the controllers with navigation functionalities [6] is considered detrimental to performance [7], given that in most VR experiences, those devices are already used for other interactions and tasks.

In this study, we explore the effects different sources of direction can have on the user experience when CAL is used in VR navigation. Apart from the two common sources of direction, head and hands, we also explore two additional sources for getting the direction of movement, one based on the hip and one other based on feet. In the latter case, we propose a novel way to infer the direction of the user movement by the mean angle from both feet. Besides the existing hardware, additional tracking devices are utilized to get orientation data. Then we evaluate the effects of those direction sources in CAL regarding users' performance, preference, motion sickness, and presence. Therefore, the examined direction sources are:

- **Headset-based**: The direction of movement is driven by the headset's orientation. The headset in a VE functions as our "virtual" head, so when it is used as the direction source, the user moves toward where they are looking.
- **Controller-based**: The orientation of the controller directs the movement. So, in this approach, the users' point of view moves towards the direction that the used controller points to.
- **Hip-based**: Locomotion is directed using a body-centric technique [8], based on the Hip. To achieve that, the hip is tracked by a tracker placed on the hip region of the user's body. Thus, the direction of movement is driven

by the orientation of this tracker. In this way, the user moves toward the direction that the hip is facing. That allows the users to move independently of where they look and decouples the controllers from the locomotion direction control.

- **Feet-based**: The direction of movement is calculated by the orientation of the feet. In this method, the user's feet are tracked by a tracker attached to each foot. The direction of movement is calculated as the mean rotation in the YAW axis of both trackers. Therefore, for the users to turn, they must adjust their feet to look in the desired direction. This direction source is our proposed technique.

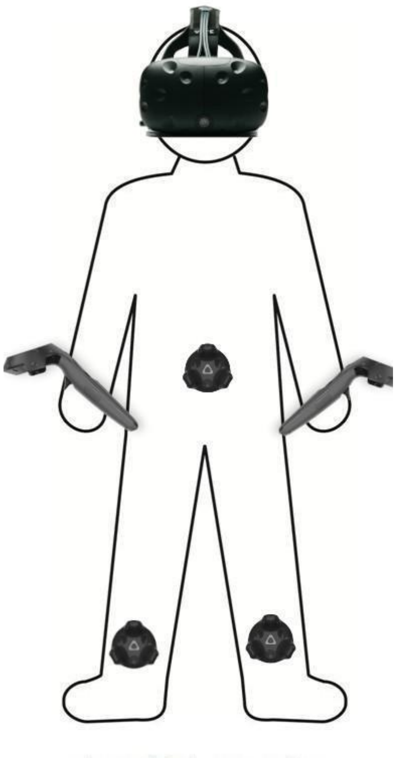


Figure1: Equipment over the body of the user.

2 EXPERIMENT DESIGN

This thesis aims to evaluate the effects of different direction sources in CAL. Those effects are examined regarding performance, motion sickness, preference, and presence. We designed and created an immersive and interactive VE to conduct a user study to achieve that. That VE served as the experimental ”canvas” for the evaluation. In this VE, the users can navigate freely by pressing directional buttons on the controllers and interacting with some created assets.

2.1 User Virtual Representation

In most VR applications, the user’s avatar consists mainly of their hands only since one cannot see their head. In this way, we avoided providing a whole body to the avatar, even though we leverage full-body tracking to not saturate our experimental goals and the focus of our study with other aspects of VR interaction, such as the virtual embodiment and the sense of ownership. Therefore, the hands of the user are visualized using the SteamVR Hand Slim3, which contains the animation for the fingers. The HMD has no visualization since the users cannot see their heads. About the rest of the hardware being used, meaning the three trackers, we decided to visualize them with relevant and suitable objects that were consistent with the rest of the avatar representation. For the tracker in the area of the hip, we used a Buckle, and for the trackers in the feet, we used a pair of boots. Those objects followed the movement of the trackers in real life, giving a sense to the users that they were wearing them.

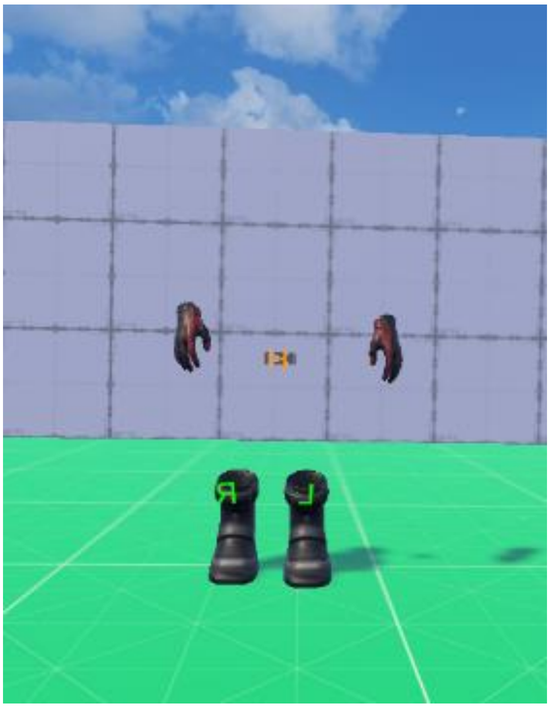


Figure 2: User’s virtual representation.

2.2 Virtual Environment

The VE was designed as simple as possible yet realistic for immersion purposes. Our goal was to create a VE that the users would enjoy and would make them feel comfortable and relaxed. Also, simple enough to avoid distracting their attention. We used simple materials, mainly gray, green, brown, and white colors, to be minimal. The map was decorated with plants and palm trees to create a feeling of an exotic place, and the sky was set to blue sky with few clouds and setting sun. We avoided adding a roof to the map because we did not want to create any feeling of claustrophobia.

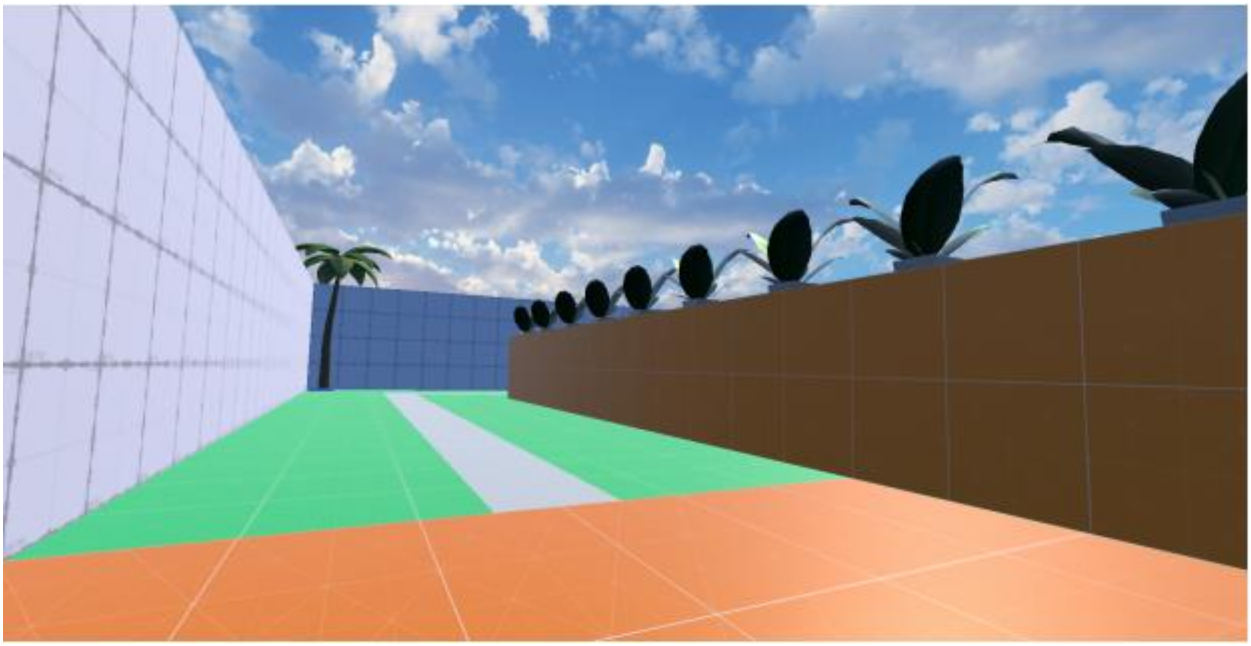


Figure 3: An example of the designed VE that was created for the experiment.

For the implementation of the experiment, three scenes were developed:

- **Training Scene:** in which the users were trained about the mechanics of the virtual application,
- **Experiment Scene:** in which the experiment was taken place,
- **Final Scene:** It was for conclusion purposes and was used to inform the users that the experiment was completed and thank them for participating.

For the correct functionality of the VE, some assets were created and repeatedly used in the map. Some of them can be seen in the figure 4.

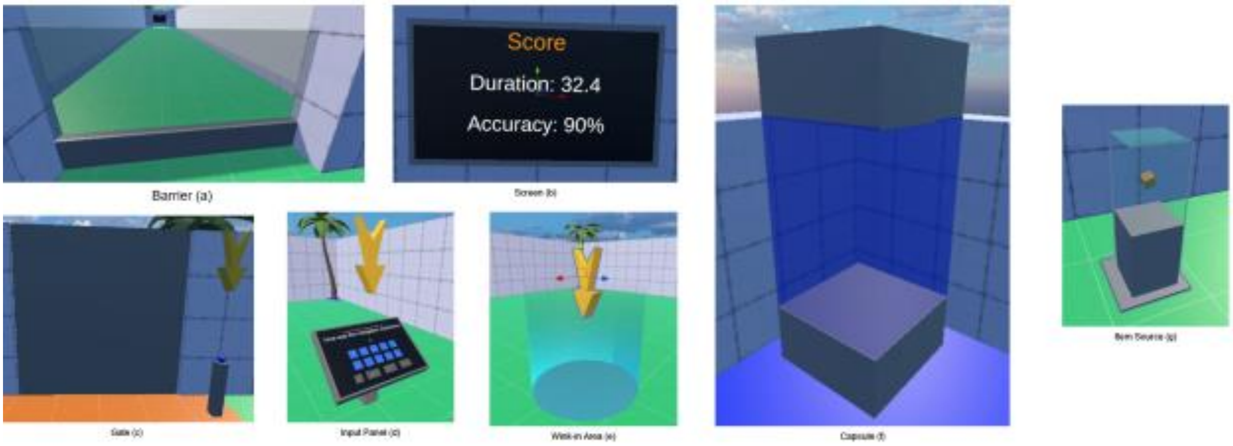


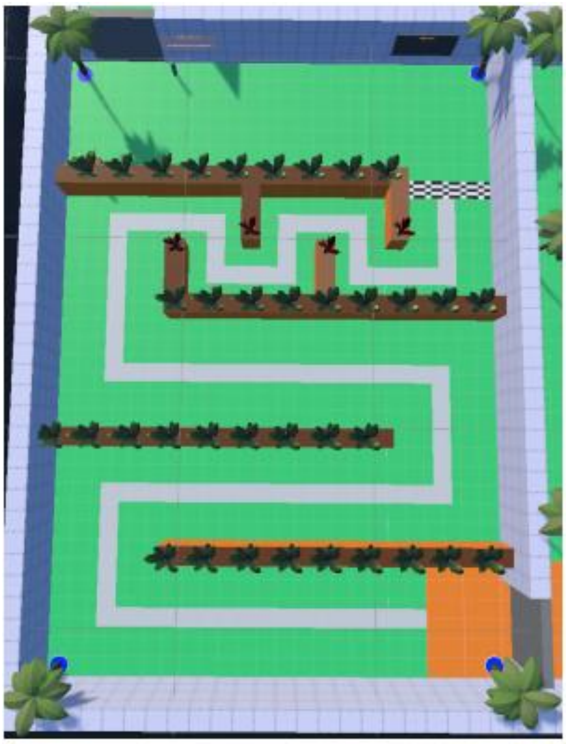
Figure 4: Some of the created assets.

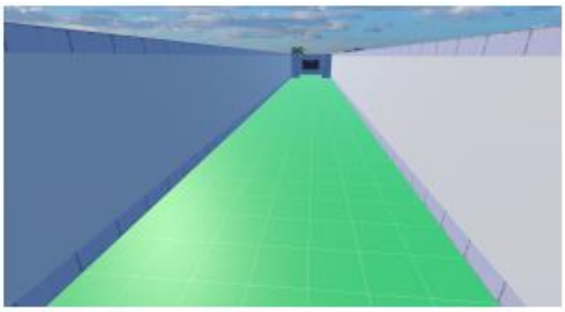
2.3 Experiment Tasks and Parts

We designed and developed three tasks to evaluate our methods, each in a different part of the VE.

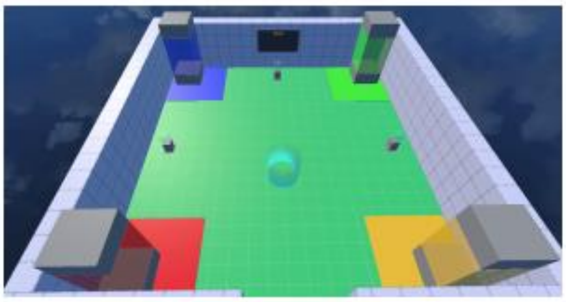
The **Tutorial** is the first part the users are introduced to when the application starts. This part aims to teach the user the basic mechanics of the VE so they are prepared for the experiment. Those mechanics include how to navigate in the VE and how to interact with the created assets. As for the utilized direction source, we selected the HMD so the users move in the direction they are looking. Besides, this selection may not be fair against the other methods, we had to choose one of them, and the head-based is used as the default method in most VR applications with CAL. This part takes place in the Training Scene.

Then the user is moved into the Experiment Scene, where the **Familiarization Part** starts. In this part, the user meets and practices the current direction source each time before moving to the experimental tasks. It is the first part of the experiment scene and has no evaluation significance. Users continue to the **ZigZag Part** when familiar with the method. This part is the first task of the experiment. Through this task, we aim to evaluate each direction method in a challenging navigation path with multiple turns. The users are asked to follow a white path on the floor and stay inside as much as possible. We measure the users' performance and accuracy with each direction source in this task. After that, in the next task, the users have to walk through a straight **corridor**. On both sides of that corridor is a wall with dots appearing as the user approaches them. The user has to walk through that corridor while counting the dots on the walls. At the end of the corridor, they are asked to submit how many they counted. This part was designed to evaluate the spatial awareness and the effect on the visual observation that each different direction source causes. Also, it is very common to look around while navigating, so we believe this represents a real scenario of VR applications. For the last part, the user has to complete an **interactive mini-game**. In this game, there are four colored capsules, each one placed on each corner of the room, and three item sources on the sides of this room. In this part, the users must grab the cube from the highlighted item source and place it on the capsule with the same color. Then, another cube appears in one of the sources. This action is repeated four times, once for each color, to complete the game. We designed this task to evaluate each method on a highly interactive game that requires hands busy for interactions and complex navigation with many turns.





(b) Corridor Part



(a) Zig-Zag Part

(c) Color-Game Part

Figure 5: Tasks in Virtual Environment

3 IMPLEMENTATION

For a VR experiment, we need a VR system capable of supporting the experiment's requirements and a high-end PC. Moreover, for our experiment, we need three trackers, one placed on the hip of the user and the other two placed on each leg. In our experiment, we used the **HTC Vive Pro** headset. Also, to make it work tetherless, we utilized the **Vive Wireless Adapter** to provide the users with a comfortable experience because the experiment requires a lot of mobility and many physical rotations. For trackers, we chose to use the **Vive Trackers**, which were ideal because they can be easily worn with specific straps for the legs and the hip. About the controllers, we chose to replace the classic controllers of the HTC Vive Pro with the **Valve Index Controllers**. These controllers contain a joystick for controlling movement which is more comfortable than the touchpad of the Vive. In addition, they are more stable over the user's hand due to the straps on the grip.

Figure 6: The used equipment for the experiment.

The Unity 3D Game Engine was used for the experiment's software development.

4 STUDY

We conducted a study to evaluate the effects of different direction sources in CAL in terms of motion sickness, performance, preference, and presence.

Performance was measured as the time and the accuracy of each user in each task by the logged data of the VE. The others were assessed using a mixed-method approach of data collection through questionnaires. More specifically, before we started the experiment, we asked the users questions to understand their profiles better. Then, when the users completed all the tasks with each direction source, they had to fill out an in-VR questionnaire. These questionnaires were related to the preference of the users with the method. After that, we prompted the users to take a break by removing their headsets and verbally administered the Motion Sickness Questionnaire. The break was intended to also provide comfort from any fatigue or dizziness the method could have induced before moving to the next one. At the end of the experiment, when the users completed all the tasks for all the methods and answered all the questionnaires, the virtual experience was over. Then, they responded to questions about all the methods comparatively. In this phase, we asked questions about the users' preferences and which method generated their highest sense of presence.

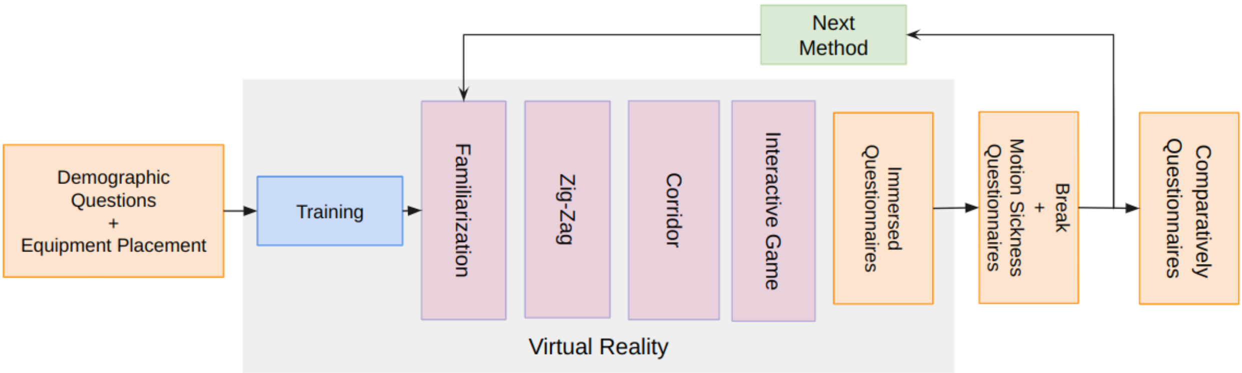


Figure 7: Experiment's flow.

In the experiment, **24 users participated**, and two could not complete it due to intense discomfort. The collected data from these two cases were ignored in the following analysis.

5 ANALYSIS

5.1 Performance

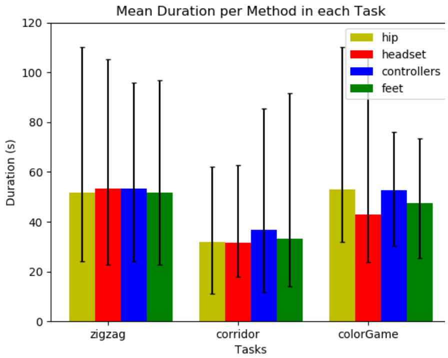


Figure 8: Mean duration per method in each task

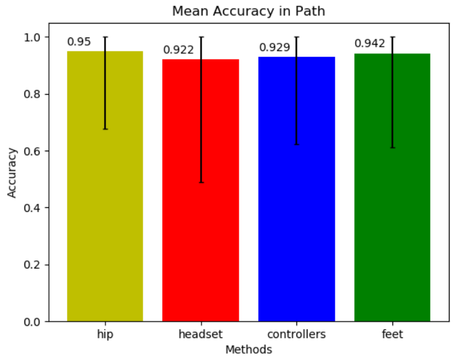


Figure 9: Mean accuracy in path

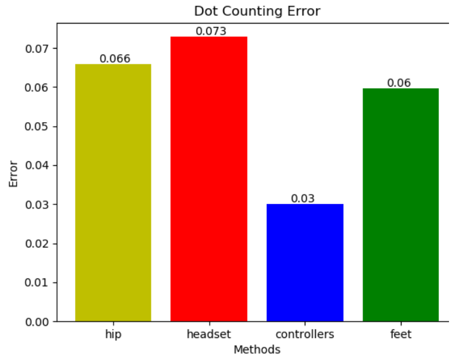


Figure 10: Dot Counting Error

In the first diagram, we can see the duration of the users to fulfill each task with each method. As we can see, there is no significant difference in performance for a method to stand out. In the second one, we can see the average accuracy of the users in the zig-zag task. In this task, we can see that most users performed very well, having a small error in most cases. However, the best accuracy occurred with the hip and feet methods. And the last diagram shows the error in the corridor task with the dot counting. In this task, the users performed worst with the headset and best with the controllers. The other two methods scored in the middle.

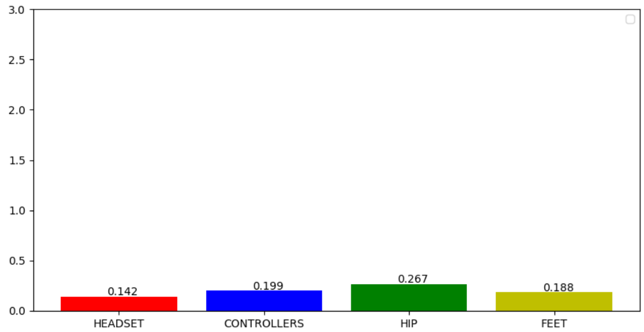
For the analysis of the **immersed (in-VR) questionnaire**, Friedman's nonparametric tests were used first to identify whether there was any significant difference between the scores for the four methods, which revealed significant differences between the method scores in two questions:

1. **How natural was the mechanism which controlled movement through the environment?** This question had a significant difference between the HMD, Controllers, and Feet. Therefore, according to the results, we can assume that the controllers were chosen as the least natural method and the feet as the most natural.

2. **How much did your movement in the virtual environment seem consistent with the real-world movement?** In this question, we had similar results. The controller was selected as the least related mechanism to the real world, and the feet as the most.

Symptoms	Headset	Controllers	Hip	Feet
General discomfort	0.31	0.4	0.54	0.31
Fatigue	0.0	0.0	0.18	0.18
Headache	0.13	0.4	0.09	0.0
Eye strain	0.04	0.4	0.22	0.0
Salivation increasing	0.0	0.0	0.0	0.0
Sweating	0.18	0.3	0.4	0.36
Nausea	0.18	0.3	0.31	0.31
Dizziness	0.27	0.3	0.36	0.31

Table 1: Motion sickness symptoms on users from each method



Total Discomfort that was induced with each method

Regarding the motion sickness questionnaires, as we can see from the results above, none of the questioned symptoms appeared significantly. Moreover, in the statistical analysis, we found no significant difference between the methods.

Regarding the comparative questionnaires, the users' favorite method was the HMD, followed by the Hip as a close second. The easiest-to-use method was also the HMD and, with equal votes, the Hip, and the Feet. As for the most natural method, the participants chose the HMD method, and in second place followed the Feet method. Last, even though the users selected the Feet-based method as the method that made them pay the most attention to the real world, it was chosen as their favorite method. It is also worth mentioning that the Controllers method was selected the least in most of the questions.

6 CONCLUSION

This thesis presents the design, analysis, and results of an experiment focused on the effect of different direction sources for CAL in VR. Literature about LTs mainly focuses on "how to move" the user and not the direction of movement. In most cases in CAL, the direction of movement is driven either by the headset or the controllers, i.e., devices that are already tracked by the tracking mechanism of the VR system. However, this approach may not feel natural to the users and may overload the headset and the controllers with navigation functionalities. In this study, we utilize extra-tracked devices to unlock different direction sources. One is positioned on the user's hip by placing a tracker in that area, and another is on the user's feet. We evaluate these four direction sources in terms of the user's performance, preference, motion sickness, and presence. To evaluate, we designed and implemented a VE and conducted a user study with 24 participants. The users had to fulfill three tasks with each method. Data were collected both quantitatively and qualitatively. The results of our research indicated that the hip and feet methods are the optimal selection for performance for executing different virtual tasks, that motion sickness is not affected by the different direction sources, and that the users preferred the feet and headset methods.

REFERENCES

[1] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc., USA, 2004.

[2] Massimiliano Di Luca, Hasti Seifi, Simon Egan, and Mar Gonzalez-Franco. Locomotion Vault: The Extra Mile in Analyzing VR Locomotion Techniques. Association for Computing Machinery, New York, NY, USA, 2021.

[3] Mahdi Nabiyouni, Ayshwarya Saktheeswaran, Doug A. Bowman, and Ambika Karanth. Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality. In 2015 IEEE Symposium on 3D User Interfaces (3DUI), pages 3–10, 2015.

[4] Evan A. Suma, Sabarish Babu, and Larry F. Hodges. Comparison of travel techniques in a complex, multi-level 3d environment. In 2007 IEEE Symposium on 3D User Interfaces, 2007.

[5] Costas Boletsis. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. Multimodal Technologies and Interaction, 1(4), 2017.

[6] Julius von Willich, Martin Schmitz, Florian Müller, Daniel Schmitt, and Max Mühlhäuser. Podoportation: Foot-based locomotion in virtual reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pages 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[7] Nathan Navarro Griffin, James Liu, and Eelke Folmer. Evaluation of handsbusy vs handsfree virtual locomotion. In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '18, pages 211–219, New York, NY, USA, 2018. Association for Computing Machinery.

[8] BoYu Gao, Ziiun Mai, Huawei Tu, and Henry Been-Lirn Duh. Evaluation of Body-centric Locomotion with Different Transfer Functions in Virtual Reality. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pages 493–500. IEEE, Mar 2021.