Computer Science and Information Systems 17(1):253–269 https://doi.org/10.2298/CSIS190427039E

A Mobile Crowd Sensing Framework for Suspect Investigation: An Objectivity Analysis and De-Identification Approach*

ElAlaoui ElAbdallaoui Hasna[1,2], ElFazziki Abdelaziz[1,3], Ennaji Fatima Zohra[1,4], and Sadgal Mohamed[1,5]

[1] Computing Systems Engineering Laboratory (LISI)
[1] Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco
[2] h.elalaoui@edu.uca.ac.ma
[3] elfazziki@uca.ma
[4] f.ennaji@edu.uca.ma
[5] sadgal@uca.ma

Abstract. The ubiquity of mobile devices and their advanced features have increased the use of crowdsourcing in many areas, such as the mobility in the smart cities. With the advent of high-quality sensors on smartphones, online communities can easily collect and share information. These information are of great importance for the institutions, which must analyze the facts by facilitating the data collecting on crimes and criminals, for example. This paper proposes an approach to develop a crowdsensing framework allowing a wider collaboration between the citizens and the authorities. In addition, this framework takes advantage of an objectivity analysis to ensure the participants' credibility and the information reliability, as law enforcement is often affected by unreliable and poor quality data. In addition, the proposed framework ensures the protection of users' private data through a de-identification process. Experimental results show that the proposed framework is an interesting tool to improve the quality of crowdsensing information in a government context.

Keywords: crowdsourcing, crowdsensing, law enforcement, objectivity analysis, de-identification.

* It is an extended version of the (8th International Conference on Model and Data Engineering MEDI 2018).

1. Introduction

For a long time, everyone was asked to perform organizational tasks, to generate content or simply to collect useful information. Today, the access to the Internet has enabled human crowds to be online 24/7, which has spurred the use of human intelligence in different types of problem solving [1]. Over the past decade, crowdsourcing has emerged as a powerful paradigm for engaging the crowd in complex tasks [2] that even powerful computing machines cannot perform. While participants could perform their tasks on computers, they can now do some of them on their mobile devices. The latter, being equipped with several integrated sensors (GPS, for example), make it possible to collect more relevant data in a given context. This technique of data collecting allows the emergence of the crowdsensing concept, especially mobile crowdsensing [3]. In addition to playing a leading role in different applications, crowdsensing has also attracted considerable attention in organizational practice and many institutions and government organizations have incorporated it into their decision-making policies [4]. By exploiting the new opportunities offered by the new information and communication technologies, these agencies are making good efforts to strengthen citizen participation. Several examples testify this and prove that crowdsensing played a key role in solving many problems: the best-known example is the Boston Marathon attack on April 15th, 2013 [5]. Crowdsensing has rapidly become a form of collaboration that allows governments (local, state or federal) to use citizens' skills in order to gather or evaluate information about a situation or a context related to a given event (natural disasters, crimes, wars, etc.).

Despite the breadth of the various crowdsensing applications mentioned above, a challenge of how to manage all the information provided by the crowd and ensure or verify their reliability emerges and requires effective support. Recurring questions about the concept of the participants' credibility and the information reliability can be raised. The success of a participatory activity depends on its quality, especially in critical contexts such as witnessing crimes or identifying suspects/criminals.

To address this problem, the purpose of this paper is to encourage the public to participate in activities that fall within the remit of the public authorities; especially e-participation. Recently, several platforms such as CrimeReports, WikiCrimes or CrimeMapping [6] have emerged in several countries. These platforms report, on digital maps, information about the location of crimes or suspects. Such platforms will be of undeniable added value in developing countries such as Morocco. To do this, we propose an approach to develop a framework for managing citizen collaboration in government activities. This framework will be a good support for the authorities allowing them a rational data management on the crimes and the criminals and an acceleration in the process of suspects' identification and localization. So, this framework will support:

- The collecting of information on crimes and suspects through an e-participatory infrastructure.
- The information about crimes and suspects can be visualized thanks to online user interfaces.
- The verification of the participants' credibility and the reliability of the information they provide will be done using an objectivity analysis.

– The anonymity of the participants will be set up using a de-identification process.

The implementation will be carried out using a set of tools. First, the process of de-identification is based on a k-anonymization algorithm effectively masking the identifiers of a participant. Subsequently, the objectivity analysis will be based on a K-Means clustering that brings together location information of a suspect to be processed by a reliability validation algorithm of each cluster.

The rest of the paper is structured as follows. After a literature review gathering a set of works developed for suspect identification and localization, we detail the proposed framework by presenting its structure and its implementation. Experiments and results are presented in Section 6 before concluding the paper with a discussion and a future work section.

2. Related Research

In this section, we review the work previously done in the context of suspect investigations. First, we discuss law enforcement and the integration of the information technologies. Then, we present some work where the crowdsensing concept was adopted and we discuss how the latter has been able to strengthen the quality of suspect investigations. Finally, we present the main methods proposed for verifying the users' credibility and the reliability of the information that the e-participants provide.

2.1. Law enforcement and information technologies

The identification and the prosecution of criminals have changed considerably and many tools have been developed to help the authorities find the suspects. By analyzing the literature, we found a considerable number of works made for similar purposes. For example, Mali P et al. in [7] propose a system based on the analysis of the images captured by the CCTV cameras in order to find a correspondence between suspects and the criminals mugshots registered in the authorities' databases.

With the aim of promoting generic reports while automating the process of detecting, and synthesizing information about a crime, Asquith [8] has implemented new techniques for sharing criminal information through sophisticated sensors. This is supported by Artificial intelligence algorithms and Natural Language Processing (NLP).

In a similar view, M. I. Pramanik et al. [9] see the Big Data Analytics and its applications as a potential for effective resolution of complex issues, including criminal analysis. Most law enforcement and government agencies need to think about adopting Big Data analytics techniques like Data Mining to cope with the large volume of data from a variety of data sources. This will help to develop effective strategies to prevent crime and the formation of criminal networks.

We note that these works neglect or rarely involve citizens who remain important sources of information in such scenarios. Human potential is a pillar in the suspect identification process. In other words, the application of crowdsourcing/crowdsensing has become an increasingly common practice among government authorities who see in the collaboration with citizens, an important step to take into consideration in any criminal case. In this part of the paper, we present some previous work that used the

crowdsensing concept for suspect identification and/or localization. We also summarize the main contributions and techniques proposed for verifying the information provided by the crowd as part of a crowdsensing-based activity.

2.2. Law enforcement and crowdsourcing

Transafe [10] is a system that allows citizens of the city of Melbourne, Australia to share location-based, time-stamped crime data, as well as their perceptions of security in a given city location. This data can be used by government agencies such as the public transit companies. The system also has a user tracking and emergency calls features. However, this platform, not yet evaluated, does not give details about the crimes and/or the criminals.

In a similar perspective, CrowdSafe [11] is a crowdsensing-based system for storing and displaying spatio-temporal data of criminal incidents. CrowdSafe includes other features such as a Safety Router that guides users through the least dangerous routes. Thanks to a dashboard, this system will enable better data analysis for smarter and safer public decision-making. The authors used real data from the Washington DC metropolitan area, but they did not propose any solutions to verify the reliability of the information shared on the platform.

On the other hand, Furtado et al. describe the WikiCrimes Web application [12], a multi-agent system that allows anyone to record and/or search for criminal information directly on maps. In WikiCrimes, the information is more credible if the user supports his statement with a document (link to a video, a newspaper article, a police report, etc.). In addition, the more people confirm a fact, the more it is reliable.

Following the same vision, Hairihan Tong proposed Bian Yi [6], a system dedicated to crime mapping allowing spatio-temporal visualization. It allows a user to mark on a Google Map, the location of a crime and all the underlying information (date, description, etc.).

Despite the importance of this platform (this has been proved through an online survey), this system has some limitations. For example, when a user submits a crime report, his contact information or credentials are not requested and remain optional. The accuracy and the authenticity of the data is still a problem. In addition, the scoring system (based on increasing or decreasing the reliability score of a report) is a feature granted only to witnesses who may not perform this task for fear that their identity will be disclosed.

2.3. User credibility and information reliability

With the explosion and diversity of the information sources, it is difficult to determine the veracity of any information given that it is done in an open and/or anonymous and not lucrative way, which is the case for participatory activities. However, this has repercussions on the quality of the decision-making process, which has led many researchers to propose some methods of Truth Discovery or Quality Assurance [13,14]. These two concepts aroused great interest in the field of participatory management.

Li Y, Gao J, Meng C, et al. [14] have classified three methods to explain the general principle of Truth Discovery: iterative methods, optimization-based methods and probabilistic graphical model based solutions (PGM). These methods and others have been successfully applied in crowdsourcing/crowdsensing applications to build mutual trust between the entities involved such [15] [16] and [17]. In the latter [17], the authors proposed a system based on fuzzy clustering to improve the quality of human computing in crowdsourcing applications. This system has been combined with a Trusted Access Control (TBAC) strategy to decide if a participant has access to a collaborative work or not.

3. The Framework overview

Taking into account the information received from the crowd is one of the major challenges in the e-participative activities. However, the ultimate objective is to find the simplest and the most appropriate way to do so while preserving the confidentiality of the information, especially in critical initiatives such as the reporting of crimes or the identification of suspects.

The purpose of this work will be to enable the authorities to develop, collect, analyze and interpret the data provided by the citizens about a crime situation. In this case, the use of an objectivity analysis turns out to be a necessity to take into account in the development of a crowdsensing-based application. This two-level analysis involves verifying the credibility of the participants, in addition to validating the veracity of the information they share. In order for the participants to demonstrate commitment and motivation, and not have this fear of disclosing their identities, a system will ensure the anonymization of their private information. The framework architecture and the request analysis and validation process will be presented below.

3.1. The architecture

The framework architecture consists of five components and their interactions are schematized in Fig. 1.

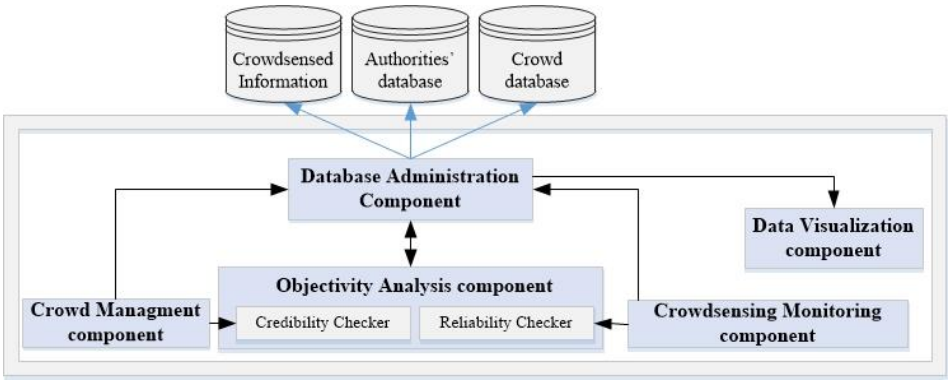

Fig. 1. The framework architecture

The Database Administration Component is linked to all other components and three databases: 'Authorities database' where the profiles of criminals are stored, 'Crowd Database' which contains all the participant identification data and finally, 'Crowdsensed Information DB' which is reserved for storing all data collected by the participants. Given the heterogeneity of the data (structured and unstructured) to be stored, we adopt a document-oriented database management system and more specifically MongoDB [18].

Containing two sub-components, the objectivity analysis component supports the validation of the participants' credibility and the verification of the collected data reliability. It is linked to the Crowd Management component that manages the participants' registration and authentication, and to the CrowdSensing Monitoring component which manages the collected data such as crime details, suspect locations and visuals (photos, videos, etc.).

All information about crimes and/or suspects are retrieved by the Data Visualization component and presented on digital maps that can be viewed by the authorities or the public through online user interfaces.

3.2. Request analysis and validation process

First, we would like to point out that the proposed framework is generic and is suitable for any type of crime or criminal incident that can be witnessed by a citizen (theft, violence and/or armed threat, physical aggression, etc.). Fig. 2 illustrates the process to be followed by a participant (a citizen) from the issuance of his request to its approval.

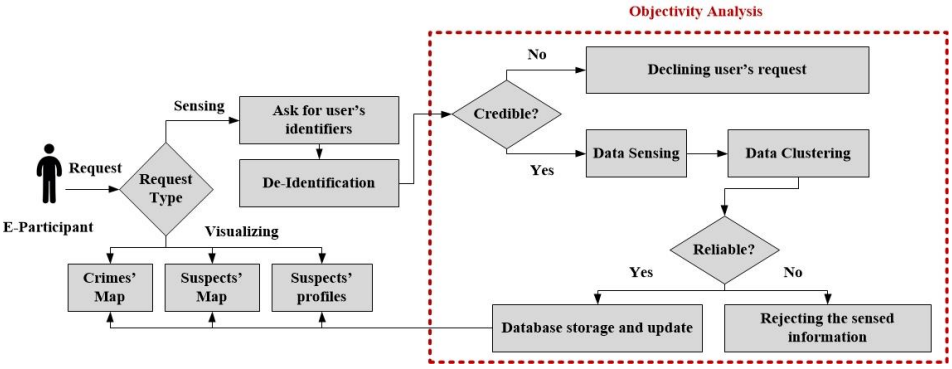

Fig. 2. The request analysis and validation

The participant information submission is carried out by an authentication to verify the user credibility. If the framework judges unfavorably his credibility, his request is automatically declined, if not, he can participate in the crime management by providing the information about a crime, the localization or the profiling of a suspect (image, video, etc.). The submitted information is subject to a verification before its approval.

4. The De-Identification Process

De-identification, also known as anonymizing data, is the process used to prevent a person's identity from being linked to information [19]. For example, data produced during research on human subjects could be de-identified to preserve the privacy of the participants. It is commonly used in the health field where de-identification is primarily focused on the protection of patient information [20,21].

The authorities have a database reserved for the personal crowdsensing information storage (national identifiers, address, etc.). This confidential information must be provided by a participant before the access enabling. Although the authorities can benefit from the relevant crowd information, they are wary of disclosing their details without a guarantee that they will not be intercepted. As a result, these information, known as Personally Identifiable Information (PII) is critical and vulnerable and needs to be de-identified. Fig. 3 below shows the process of de-identification followed after the registration or the authentication of a participant to allow him or not the access to the framework.

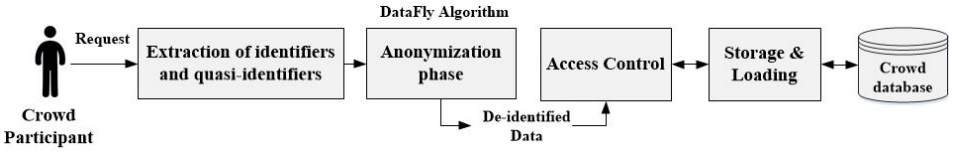

Fig. 3. The de-identification process

4.1. The extraction of identifiers and quasi-identifiers

To begin, we define the attributes that directly designate the identity of an individual (Identifiers) such as the name and the national identity card number and those who identify him indirectly (Quasi-Identifiers) such as ZIP code or the date of birth. We want to point out that a subset of some quasi-identifiers allows the identification of an individual. In other words, if the QIs are sufficiently well correlated, they can be combined to create a unique identifier. Table 1 below defines these attributes and their categories.

Table 1. Identifiers and Quasi-identifiers with their categories and PII

	Category	PII
Identifiers	Name	Fist Name, Last Name, Surname
	IDs	National Identity Card Number, Social Security Number, Driver's License Number, Passport Number
	Contact	Phone Number, Email Address
Quasi-Identifiers	Age	Birth Date (year, month, day)
	Gender	-
	Marital Status	Not Married (Single, Widowed,

	Divorced), Married
Contact	Country, City, Department, ZIP, Street Address
Job	Job, Institution or Organization

4.2. The anonymization phase

After identifying and classifying PII, we apply two strategies: pseudonymization [19] to suppress (replacing with *) or hide (replacing with symbols) identifiers and make individuals de-identified. Also, we proceed to a k-anonymization process [22] and treat the data by ensuring that at least k individuals have the same combination of QI values through generalization and/or deletion techniques. This reduces the probability of disclosure risk by 1/k. For example, the k-anonymization replaces some original data in the records with new range values and retains some unchanged values. The new combination of QI values prevents the identification of the individual and avoids the destruction of the data records.

Several methods have been established for this purpose [23]. In this paper, we use the DataFly algorithm [23] known for its minimal execution time and information loss. The DataFly algorithm is detailed below. It counts the frequency on the QID set (the set of quasi-identifiers) and if the k-anonymization is not yet satisfied, it generalizes the attribute having the most distinct values until the k- anonymization is satisfied.

```
The Datafly Algorithm
Input: {T: the table with private data, k: the k-
anonymity constraint, QID ={Q1,Q2, …,Qn}: the n quasi-
identifiers, suppThreshold: a suppression threshold};

Output: GT: the generalized table

start
  Compute the frequency count (FreqC) of T using QID

  While (FreqC<= k) do

     // Verify if the table T is ready for suppression

          If (FreqC <= suppThreshold) do

             Suppress tuples with FreqC <= suppThreshold

          Else

          Search for attributes t with FreqCt= max(FreqC)

          GT = Generalize table T using t

          End if
```

```
    Compute FreqC

End while

Return: GT
end.
```

An example of the de-identification achieved using the above-mentioned techniques is presented in the implementation section. Table 2 and Table 3 are respectively the private data table and the de-identified data table using some attributes.

Table 2. Identification Data Table

	Identifiers			Quasi-Identifiers			
	Name	NIC	Phone	Age	Gender	Marital Status	Zip Code
1	Name1	NIC1	+212666666666	24	M	Divorced	10242
2	Name2	NIC2	+212611111111	23	F	Single	10256
3	Name3	NIC3	+212600000000	35	M	Single	10440

Table 3. De-Identified Data Table

	Identifiers			Quasi-Identifiers			
	Name	NIC	Phone	Age	Gender	Marital Status	Zip Code
1	*	$*$	+2126$$$$$$$$	[20 :30)	M	Not Married	102**
2	*	$*$	+2126$$$$$$$$	[20 :30)	F	Not Married	102**
3	*	$*$	+2126$$$$$$$$	[30 :40)	M	Not Married	104**

5. The information reliability checking

In a previous work, we were interested in the suspect profiling data transfer, analysis and processing. However, in this paper, we detail the process of verifying the information provided by a crowd concerning a crime/suspect locations.

For this, a participant is required to mark on a map, the locations where he identified a crime or a suspect and the time of the identification. Thus, the information is not precise and can generate some problems in terms of precision. For example, two people may be in the same place at the same time but within a few meters. Therefore, instead of checking each location, we check the entire group of nearby locations. The first step, then, consists of grouping (clustering) these reported locations before applying the objectivity analysis.

5.1. Locations clustering

There are several clustering algorithms for partitioning the received location data into subgroups, or more formally into clusters. These "group" together similar observations (here localizations). To deal with this problem, we chose to use unsupervised learning methods, specifically the K-Means algorithm [24]. It aims at structuring all types of data into k groups in order to minimize a defined function. Fig. 4 illustrates this classification by showing the resulting clusters after applying the K-means algorithm.

Fig. 4. An example of the clusters obtained after applying the K-Means algorithm

5.2. The objectivity analysis algorithm

The objectivity analysis proposed in this paper is based on a simple and probabilistic algorithm in order to identify the most reliable information among a crime/suspect locations reported at a given moment. By setting a moment $T = t \pm \Delta t$, we define some parameters for reliability calculation as follows:

- S: the number of participants.
- L: set of information reported at the time T where $L = \{l_0, l_1 \ldots l_n\}$ and li is the ith cluster of localizations at time T and n is the number of clusters.

```
The information reliability analysis algorithm
Input: { the information L= {l0, l1… ln} and the number of
participants S };

Output: Identified reliable locations and their scores

Start.
For l0 to ln do
```

$RS(l_i) = |S_{li}|/|S|$ where $\sum Rs(l_i) = 1$ and S_{li} is the number of people who reported the location l_i

End For.

Return: the cluster l_i with the highest reliability score RS
end.

For each location l_i, the algorithm deduces the reliability score of the group l_i as a function of the current estimation. At the end of n iterations, it returns the cluster with the highest score, which indicates the most reliable information.

6. The implementation

In this section, three main online user interfaces will be presented: user registration interface and two data collecting interfaces that enables the e-participants to report information about a crime or a suspect. These interfaces are adapted for both web and mobile. For each case, we also present the storage process and the databases structure.

6.1. User registration

Before allowing a participant to report any information, he must first register in the framework by indicating the information presented in Fig. 5. These information are anonymized using the de-identification process presented earlier before being stored in the crowd database.

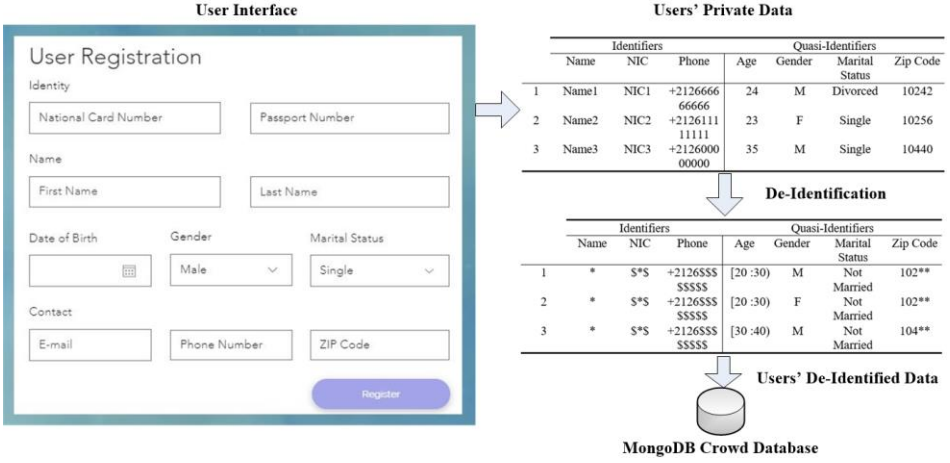

Fig. 5. User registration interface

6.2. Data collecting

The e-participants validated by the access control module are authorized to enquire information about a crime or a suspect to enable the authorities to progress in their decision-making process. These information are then structured to be suitable for storage in a document-oriented database such as MongoDB. In this section, we present the implemented user interfaces as well as the structure of the documents stored in a MongoDB database.

Crime description. Fig. **6** illustrates the web interface for loading the information about a crime and the structure of the crime document (JSON file) that is stored in the 'Crowdsensed Information Database' (MongoDB).

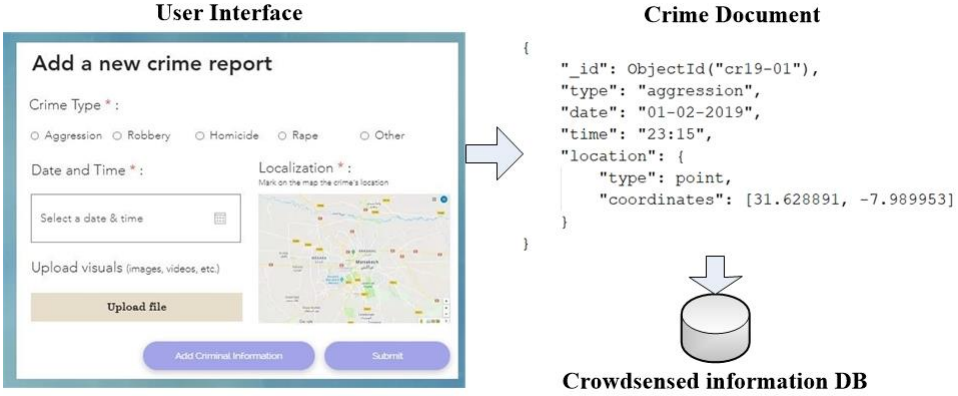

Fig. 6. Crime reporting

Suspect identification and localization. By clicking on the 'Add Suspect Information' button of the crime report form, a participant can add information to assist in identifiying or locating a suspect. Fig. 7 presents the form to which participants are redirected to share this data. This latter is structured in a JSON file before being stored in the MongoDB 'Crowdsensed Information Database'.

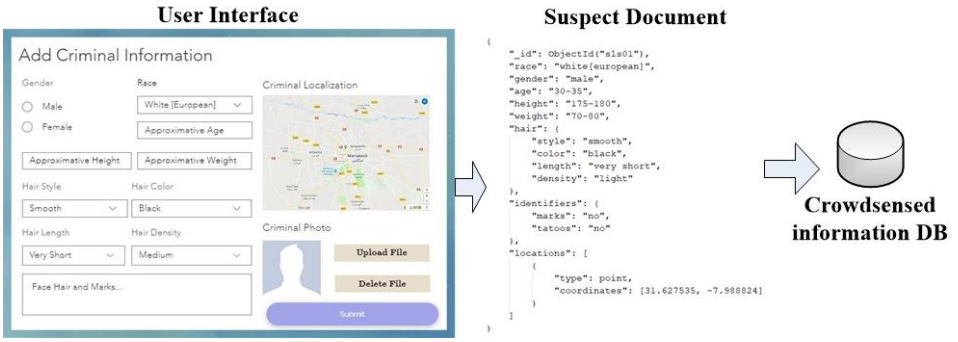

Fig. 7. Suspect identification and localization information

7. Case Study, Results and Discussion

A simulation of the proposed framework process was performed in an earlier work using the Anylogic simulator[1]. The simulation [25] helped to test the proposed framework in a virtual environment and to get a dataset to apply the objectivity analysis.

In this part, we present the results obtained after the processing and the analysis of the information reported by the e-participants. First, we will expose the results of the objectivity analysis (verification of the reliability of the reported locations). Then, the retained locations are used to generate a spatio-temporal map where all the reliable localizations are marked. We are interested in the case of suspects but this is also valid for the information on the crimes. Finally, we discuss the overall contributions of the proposed framework and its effectiveness.

7.1. Objectivity analysis results

The objectivity analysis and more particularly the information reliability make possible the data refinements for a better result of the suspect identifying process. Fig. 8 below is a diagram of two Matlab graphs plotting on the left all the information of a suspect location reported by the crowd and on the right those filtered and considered reliable by the OA module.

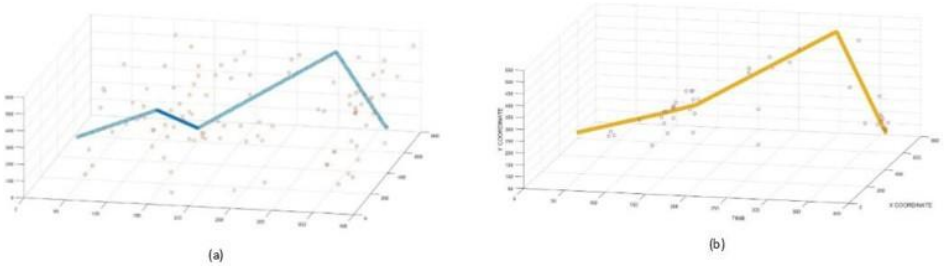

Fig. 8. Suspect localizations (a) before and (b) after applying the objectivity analysis

[1] www.anylogic.com

7.2. Suspect spatio-temporal map

Since the location information of a suspect are reported by the e-participants, the framework only retains the most reliable information (checked by the objectivity analysis component) to generate a digital map viewable only by the police officers (Fig. 9). This map allows tracking the movements of a suspect.

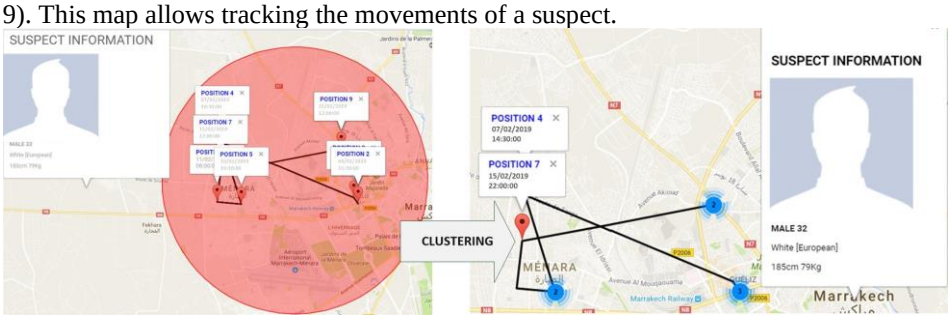

Fig. 9. Suspect movements spatio-temporal map

7.3. Discussion

The proposed framework differs from the systems discussed in the literature review in that it is based on both new technologies and the concept of crowdsensing. This combination is important for the acceleration of the investigation process. In addition, the proposed framework is not only intented for reporting crimes but also in identifying and tracking potential suspects. It also incorporates 3 important contributions: the verification of the e-participants' credibility before granting them the access to the different interfaces, the verification of the information reliability through an objectivity analysis and a de-identification process of users' private data which must encourage them to collaborate.

The experimental results obtained prove the effectiveness of the proposed framework since it allowed a significant reduction of 37.7% of the data that could hinder the investigation process or slow it down. Also, the generated spatio-temporal map is a good visual and support for the decision-makers to identify the risky areas (the dispersion of the crimes in a city) and follow the movements of a suspect.

The major limitations of this research work can be summarized in two major points: the lack of real data to test and validate the framework in question. Morocco doesn't currently make available to the public, data on crimes and/or suspects. Also, and as mentioned in the section 'Conclusion and Future Work', an integration of a reliable and secure reward system is a necessity. This will significantly increase the participants' motivation.

8. Conclusion and Future Work

The low rate of crime reporting in the cities to the authorities and the reduced number of the authorities' officials allocated to this task is a major impediment to the smooth

process of identifying and locating suspects. Therefore, adopting the concept of crowdsourcing for the information collecting and sharing between the citizens and the authorities can alleviate this problem. Also, the anonymization of e-participants can encourage them to collaborate. On the other hand, incorporating an objectivity analysis into the process can make it more relevant.

The implementation of this complex and spatiotemporal process by means of datamining tools and the storage of information in a document-oriented database (MongoDB) make it possible to have an appropriate infrastructure for taking spatiotemporal information into account relating to the identification and location of suspects.

In order to make this infrastructure more confidential and secure, we are considering the integration of the blockchain concept as future work.

References

1. Lease M, Alonso O. Crowdsourcing and Human Computation, Introduction [Internet]. In: *Encyclopedia of Social Network Analysis and Mining (ESNAM)*. Springer, 304–315 (2014). Available from: https://www.ischool.utexas.edu/~ml/papers/lease-esnam14.pdf.
2. Howe BJ. The Rise of Crowdsourcing. *Wired Mag.* 14(6), 1–4 (2006).
3. Guo B, Wang Z, Yu Z, *et al.* Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm. *ACM Comput. Surv.* [Internet]. 48(1), 1–31 (2015). Available from: http://dl.acm.org/citation.cfm?doid=2808687.2794400.
4. Cupido K, Ophoff J. A Model of Fundamental Components for an e-Government Crowdsourcing Platform. *Electron. J. e-Government*. 12(2), 141–156 (2014).
5. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. *Am. J. Prev. Med.* [Internet]. 46(2), 179–187 (2014). Available from: https://doi.org/10.1016/j.amepre.2013.10.016.
6. Tong H. A crowdsourcing based crime mapping system. (2014).
7. Mali P, Rahane V, Maskar S, Kumbhar A, Wankhade S V. Criminal Tracking System using CCTV. *Imp. J. Interdiscip. Res.* [Internet]. 2(7), 2454–1362 (2016). Available from: http://www.onlinejournal.in.
8. Asquith B james. Crime Intelligence 2.0: Reinforcing Crowdsourcing using Artificial Intelligence and Mobile Computing [Internet]. (2017). Available from: https://cloudfront.escholarship.org/dist/prd/content/qt39s3k7bw/qt39s3k7bw.pdf.
9. Pramanik MI, Lau RYK, Yue WT, Ye Y, Li C. Big data analytics for security and criminal investigations. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. (2017).
10. Hamilton M, Salim F, Cheng E, Choy SL. Transafe: A crowdsourced mobile platform for crime and safety perception management. *Int. Symp. Technol. Soc. Proc.* 2015-July (2015).
11. Shah S, Bao F, Lu C-T, Chen I-R. Crowdsafe: Crowd Sourcing of Crime Incidents and Safe Routing on Mobile Devices. In: *ACM SIGSPATIAL GIS'11.* , 521–524 (2012).
12. de Oliveira M, D'Orleans J, Caminha C, *et al.* Collective intelligence in law enforcement – The WikiCrimes system. *Inf. Sci. (Ny).* [Internet]. 180(1), 4–17 (2009). Available from: http://dx.doi.org/10.1016/j.ins.2009.08.004.
13. Pouryazdan M, Kantarci B, Soyata T, Song H. Anchor-Assisted and Vote-Based Trustworthiness Assurance in Smart City Crowdsensing. *IEEE Access*. 4, 529–541 (2016).
14. Li Y, Gao J, Meng C, *et al.* A Survey on Truth Discovery. *ACM SigKdd Explor. Newsl.* [Internet]. 17(2), 1–16 (2016). Available from: https://doi.org/10.1145/2897350.2897352.
15. Xu G, Li H, Tan C, Liu D, Dai Y, Yang K. Achieving efficient and privacy-preserving truth discovery in crowd sensing systems. *Comput. Secur.* [Internet]. 69, 114–126 (2017). Available from: https://doi.org/10.1016/j.cose.2016.11.014.

16. Huang C, Wang D, Chawla N. Towards time-sensitive truth discovery in social sensing applications. *Proc. - 2015 IEEE 12th Int. Conf. Mob. Ad Hoc Sens. Syst. MASS 2015*. , 154–162 (2015).
17. Folorunso O, Mustapha OA. A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments. *Appl. Comput. Informatics* [Internet]. 11(2), 116–129 (2015). Available from: https://doi.org/10.1016/j.aci.2014.07.001.
18. Sowmya R, Suneetha K R. Data Mining with Big Data [Internet]. In: *2017 11th International Conference on Intelligent Systems and Control (ISCO)*. , 246–250 (2017). Available from: http://ieeexplore.ieee.org/document/7855990/.
19. Khalil M, Ebner M. De-Identification in Learning Analytics. *J. Learn. Anal.* 3(1), 129–138 (2016).
20. Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. *J. Am. Med. Informatics Assoc.* 24(3), 596–606 (2017).
21. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J. Biomed. Inform.* 58, S11–S19 (2015).
22. Patil D, Mohapatra RK, Babu KS. Evaluation of generalization based K-anonymization algorithms. *Proc. 2017 3rd IEEE Int. Conf. Sensing, Signal Process. Secur. ICSSS 2017*. , 171–175 (2017).
23. Ayala-Rivera V, McDonagh P, Cerqueus T, Murphy L. A Systematic comparison and evaluation of k-Anonymization algorithms for practitioners. *Trans. Data Priv.* 7(3), 337–370 (2014).
24. Arora P, Deepali, Varshney S. Analysis of K-Means and K-Medoids Algorithm for Big Data. In: *Physics Procedia*. (2016).
25. El Alaoui El Abdallaoui H, El Fazziki A, Ennaji FZ, Sadgal M. A gamification and objectivity based approach to improve users motivation in mobile crowd sensing. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. (2018).

ELALAOUI ELABDALLAOUI Hasna: Is a computer science engineer, graduated from the National School of Applied Sciences Marrakesh/Morocco year 2014. After acquiring scientific and technical knowledge in the computer and information systems field; especially in designing, modeling and implementing software solutions from both the architectural and administrative perspectives, she then integrated the Computing Systems Engineering Laboratory of CADI AYYAD University since January 2016 to prepare her PhD thesis. Her research interests include e-government applications, crowdsourcing, image processing, mobile applications, etc.

ELFAZZIKI Abdelaziz: Received the M.S. degree from the University of Nancy, France, in 1985, and the Ph.D. degree in computer science from CADI AYYAD University in 2002. He has been with CADI AYYAD University since 1985, where he is currently a Professor of computer science. He has been responsible for the master's degree program in information system engineering since 2006. He was the Director of the Computer Systems Engineering Laboratory between 2011 and 2015. He has co-authored several papers on agent-based image processing, and is the main Author of over 20 papers in software engineering and data analytics field. His research interests are related to software engineering, decision support, big data, data analytics, crowdsourcing, and e-government. In the MDA field, he has been involved in agent-based systems, service-oriented systems, and decision support systems.

ENNAJI Fatima Zohra: Is a computer science engineer, graduated from the National School of Applied Sciences Marrakesh/Morocco at 2014 and finished her PhD in 2019. She joined in 2015 the Computing Systems Engineering Laboratory of CADI AYYAD University. Her thesis includes many research interests like social media, sentiment analysis, data mining, Big Data Analytics, crowdsourcing, social CRM, etc. She participated in many international conferences and published many articles in international journals.

SADGAL Mohamed: is professor of computer science at Cadi Ayyad University, Morocco, and researcher on computer vision with the Vision team at the LISI Laboratory. His research interests include object recognition, image understanding, video analysis, multi-agent architectures for vision systems, 3D modelling, virtual an augmented reality, among other topics. Before Marrakech, he was in Lyon (France), working as Engineer in different computer Departments between 1988 and 1994. He obtained a PhD in 1989 from Claude Bernard University, Lyon, France.

Received: April 27, 2018; Accepted: September 12, 2019