

Method for Automated Georeferencing and Integrating Printed Maps in GIS for Collecting Addresses

Mohamed El Imame Malaainine¹, Hassane Rhinane¹, Lahcen Baïdder¹, Omar Bachir Alami²

¹Faculté des Sciences Ain Chock, Université Hassan II, Casablanca, Maroc

²Ecole Hassania des Travaux Publics, Casablanca, Maroc

Email: mohamed.malaainine@gmail.com, h.rhinane@gmail.com, lbaidder@gmail.com, alami.ehtp@yahoo.com

Received November 1, 2012; revised December 5, 2012; accepted January 7, 2013

ABSTRACT

Addresses are important data for urban applications. About 80% of the information local authorities use have a geographic component that is generally related to addresses. Addressing systems efficiency depend on the quality of addresses locators. There are several methods to collect data. Surveys from the field are essential: GPS and pre-printed maps can be used to achieve this goal. GPS surveys from the field may be a solution, but it remains practical only for limited areas. To insure an accepted accuracy, GPS methods need special considerations that are time and money consuming. For Casablanca's addressing locators, an alternative approach was adopted to collect 400,000 points. It took two months, 200 operators and 3500 printed maps to cover a study area of 1226 km². This paper is to develop an optimized approach based on automated procedure for reintegrating printed maps in a geographic information system (GIS). It saves georeferencing time from 5 min to just seconds per document. It insures, more importantly, an accuracy that is between 20 cm to 1 m for scales that are between 1/500 and 1/2500. It ensures maps' integration, independently of base map and coordinates system by introducing the notion of Georeferencing Code (GC).

Keywords: GIS; Addressing Reference; Georeferencing; Geodatabase; Raster Data; NAC; GC

1. Introduction

Addresses are an important data for urban applications. About 80% of the information local authorities use have a geographic locations, and most of those are related to addresses [1,2]. It is very important to have an accurate database of addresses. The emergency dispatching applications for example, can not work without such database [3]. Unfortunately, having a reliable data, especially in developing countries, is often a big challenge.

In morocco, some researches took an interest in urban phenomena such as detecting slums [4] or urban heat islands [5], but rare are the researches that discuss this matter. Even those that touched this subject [6] didn't emphasis on addressing locators schema and approaches to get data from the field. On other hand, an interesting work was realized by Davis and Fonseca [7] on the certainty of locations produced by an address geocoding system, where they present a conceptual schema for addressing database that it is flexible enough to accommodate different contexts, including the Moroccan case. Their work extends Simpson's and Yu's work on postal codes to records of any kind [8]. An important matter is still to study: It concerns the method to acquire data, to build a reliable reference database. Traditional methods that involve GPS surveying have a very low quality price ratio.

Moreover, using printed maps if not managed carefully, may lead to data loss or low quality. It also presents the big challenge of georeferencing and dependence to base maps and coordinates references.

The objective of this work is to present a method to gather data from the field, based on automation of georeferencing maps of scales that are between 1/500 and 1/2500, achieving accuracy better than 1 m. The case study concerns urban addresses of Casablanca.

2. Material and Methods

2.1. Case Study Area

Casablanca is Morocco's economic capital. It is also its largest city, with more than 3 million inhabitants with a total area of 1140.54 km² [4] (**Figure 1**). It is located on the Atlantic coast, about 100 km south of the administrative capital (Rabat). The Wilaya of Greater Casablanca counts two prefectures (Casablanca and Mohammedia) and two Provinces (Nouceur and Mediouna) for a total of 17 municipalities, 10 urban and seven rural [4].

2.2. Available Data and Procedures

In order to collect 400,000 addresses, representing the majority of addresses locators in Casablanca and Moha-

media prefectures, 3500 maps were generated (Figure 2). Each map represents a sub-zone, and contains necessary data for collecting addresses from the field: routes, routes limits, plots, points of interests, neighbourhoods and sub-zone's boundaries [9].

Here below, Figure 3 shows the project's operations

process flow:

The metropolis is subdivided into sub-zones. Every one of them is printed in A3 format map with the sub-zone number in the header of the document (Figure 2). Once the work in the field is done, and both geometry and necessary details for the matching are drawn, the



Figure 1. Case study area.

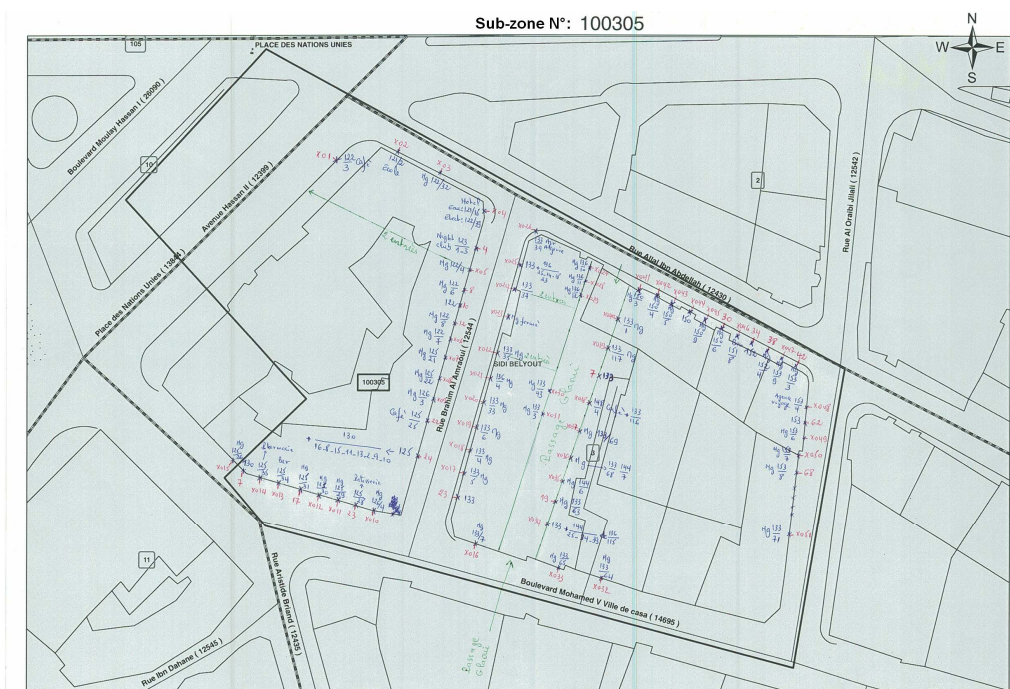


Figure 2. Sub-zone sample.

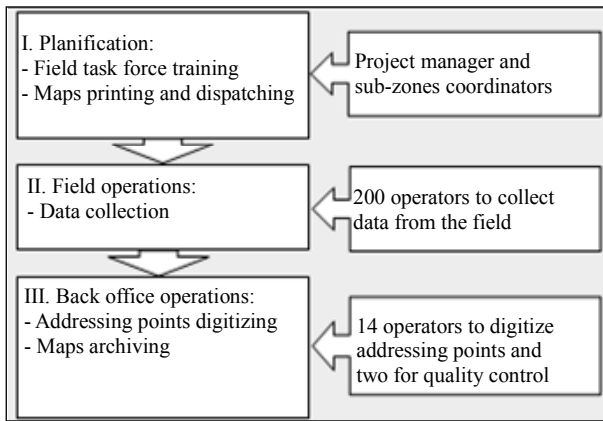


Figure 3. Project's operations process flow.

maps are returned to the processing task force in the back-office for digitizing. After the quality control operations, the map is finally scanned and archived.

The hard document is the only reference for the quality control and verification of the digitized data. That is why it should be scanned and preserved. It should even be georeferenced to simplify its reuse. This would optimize the operation of digitizing data (Addresses locators in our case) [10].

3. Method of Automated Georeferencing

By georeferencing a raster map we mean to translate, rotate and to scale a digitalized copy of the original map so its (significant) points do align with their corresponding points on a referential (vector) map [11].

The operation of georeferencing maps that were generated from our geographical information system is one process among others, of an entire procedure. It starts from the printing and ends with the georeferencing of the digitized maps. In order to optimize the last step, we should rethink the first one.

The georeferencing of scanned maps depends on recognizable reference points on the document and in the geographical information system. But data changes in the system. In our case, base map data and sub-zones changed in many cases compared with the paper maps. The next work flow in (Figure 4) presents the proposed procedure and necessary parameters.

3.1. Printing and Automated Georeferencing Methods

The GIS offers the possibility to insert automatically some references on every printed map. We are using this possibility to generate a code that indicates the coordinates of the fore corners of the map's frame, to use it for georeferencing. This will enable us to be independent of any possible updates in the database, between the maps' printing and their use; But also to perform this operation

in any other system (Even if it doesn't have the same data that figures in the map).

In order to optimize the georeferencing code in the maps generated from our GIS, we are considering the Natural Code Area (NAC) alphanumeric codes, developed by Geographic Products Inc. [12]. NAC uses a Base 30 numeral system to encode position. The primary grid for the globe is subdivision of the flattened globe into a 30 by 30 grid. Each grid is referenced by a unique identifier that is described by a base 30 "number". This means that each grid reference is one a single identifier. This process is recursive and quickly converges under 1 m resolution (Table 1).

The previous table shows that six digits NAC ensures a resolution around 5 cm, which is sufficient for our addressing database. For different needs, applications may use lengthier NAC [14].

The Georeferencing Code (GC) is the concatenation of NAC(X1), NAC(Y1), NAC(X2), NAC(Y2), NAC(X3), NAC(Y3), NAC(X4) and NAC(Y4) where (X1, Y1), (X2, Y2), (X3, Y3) and (X4, Y4) are the coordinates of the

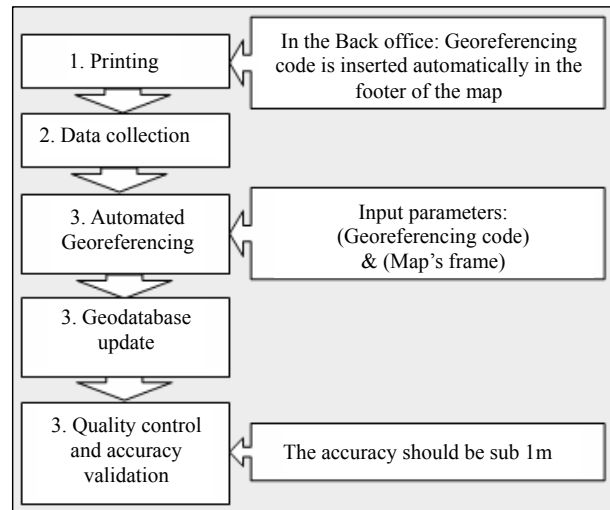


Figure 4. Procedure for printed maps use.

Table 1. Convergence of the NAC resolutions [13].

Number of NAC digits	Resolution in degrees	Resolution in meters
1	12	1,333,440
2	0.4	44,448
3	0.013333	1481.6
4	0.000444	49.38667
5	1.48×10^{-5}	1.646222
6	4.94×10^{-7}	0.054874
7	1.64609×10^{-8}	0.00182914

map’s frame corners.

Here bellow, **Figure 5** shows the GC calculation algorithm:

Table 2 presents parameters for printing and geo-referencing sub-zone number 235301:

The corresponding Georeferencing Code is:

GBS2BQNKV6WDGBRSNVNKTVPGBS2BQNK
TVPCGBRSNVNKV6WD.

Figure 6 illustrates the printed sub-zone with the GC in the footer of the map.

The result of georeferencing routine is presented in **Figure 7**. We consider 10 control points to compare coordinates between printed maps and reference geodata-base. **Table 3** shows the results of this comparison. The maximum difference observed between georeferenced map and reference database is 28.4 cm. The correlation coefficient between the two coordinates’ series is 0.9999999998.

3.2. Method Validation: Accuracy Discussion

3.2.1. The Target Accuracy

Our addressing database statistics for Mohammedia town

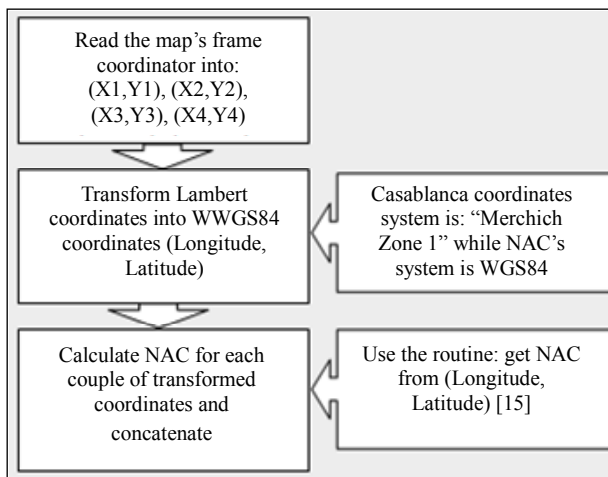


Figure 5. GC calculation algorithm.

Table 2. Printing parameters for sub-zone N° 235301.

Reference Coordinate	Coordinate system		
	Merchich Zone I	WGS84	NAC
X1	288360.1735	-7.678952099	GBS2BQ
Y1	333074.9451	33.57486963	NKV6WD
X2	288044.1927	-7.682357531	GBRSNV
Y2	333081.8515	33.57260272	NKTVPC
X3	288354.6853	-7.678952099	GBS2BQ
Y3	332830.5669	33.57260272	NKTVPC
X4	288038.6963	-7.682357531	GBRSNV
Y4	332823.6604	33.57486963	NKV6WD

show that 79.57 % of addresses locators are 5 meters far from the nearest different point address. 98.98% of addresses locators are 1 m far from the nearest different address.

We consider 1 m as a reasonable tolerance for our addressing database. This means that any method of data collection should have an accuracy that is sub 1 m.

3.2.2. The GPS Accuracy Discussion

Global Positioning System (GPS) that is a satellite based navigation and surveying system for determination of precise position and time, using radio signals from the satellites, in real time or post-processing mode [16]; Is commonly considered as the best data collection tool from the field. This fact is contestable for two main reasons: cost and accuracy. These two factors are related as shown in **Table 4**.

Various errors are in GPS system: Orbital errors, Clock errors, Ionospheric errors, Multipath errors, Tropospheric errors, Receiver noise, Relativistic corrections, Dilution of Precision (DOP), etc. [17]. The ionospheric delay is the main problem in achieving millimeter level positioning [18]. These errors bias the receiver user’s position to be ±15 m from the actual coordinates [19].

3.2.3. Printed Maps with Georeferencing Code Method Accuracy

In our method, the accuracy of digitized data depends on the *scale of the printed map, the printer resolution, the Scanner resolution, the georeferencing precision and the graphic error* (of marks on the map and mouse cursor on the screen).

Scanning resolution is normally expressed as pixels-per-inch (ppi) or dots-per-inch (dpi). Printing resolution is also measured in dots-per-inch and depending on the type of printer the dpi can range anywhere from 72 dpi to 3000 dpi [20].

Supposing that we use the same resolution “R” for printing and scanning the map, the final document resolution is R (expressed in dpi) and the map’s scale is 1/S, the Devices’ Error (DEr) is then:

$$DEr = \left[\left(2.54 \times 10^{-2} \right) / R \right] \times S \quad (1)$$

The graphic error is the difference between the marked point on a map with a pencil, and its actual position [21]. Its average value depends on the operator. In general, it can be estimated to 0.1 mm.

With the map scale equals to 1/S. the Graphic Error (GpEr) is:

$$GpEr = 0.1 \times 10^{-3} \times S \quad (2)$$

Using the affine polynomial transformation based on the fore points of control as presented in **Figure 6**, the accuracy of georeferencing depends on the Graphic Error

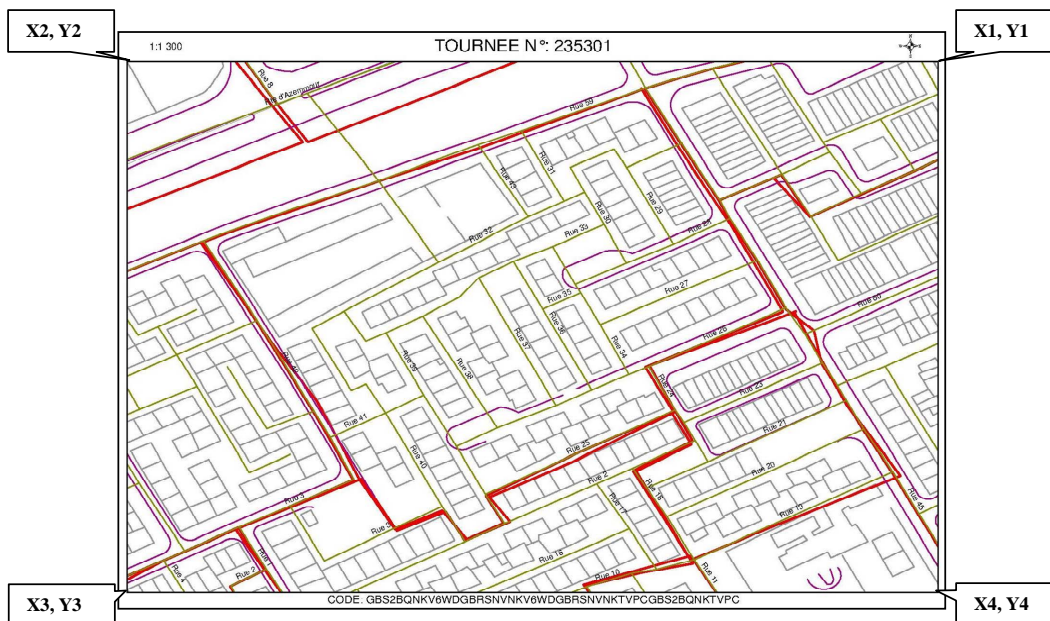


Figure 6. Printed “Sub-zone” with a generated georeferencing code.

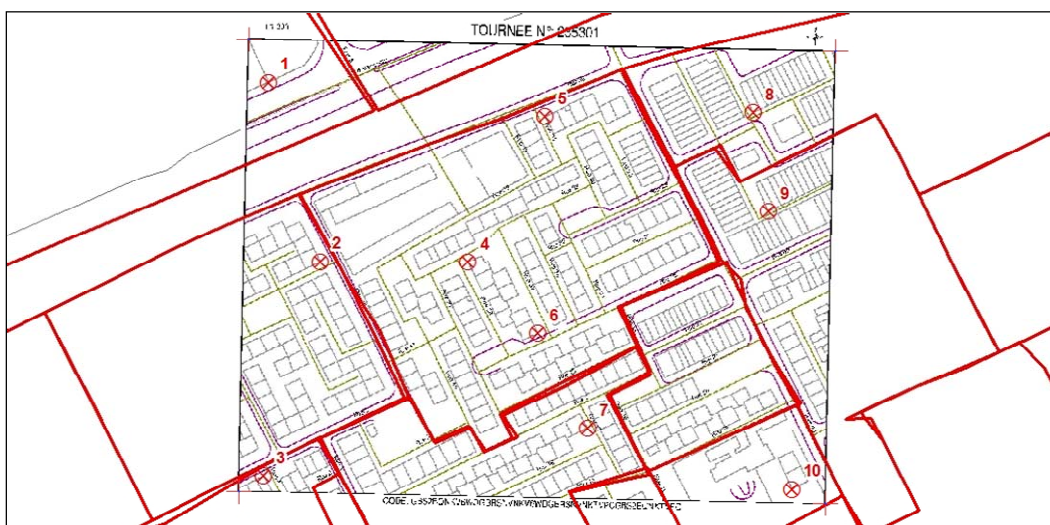


Figure 7. Georeferenced raster with “georeferencing code method”.

Table 3. Comparative table between calculated and reference coordinates of control points.

Control point	Database coordinates (m)		Map's coordinates (m)		Difference (m)	
	X	Y	X'	Y'	X'-X	Y'-Y
1	288055.00	333057.78	288055.05	333057.87	0.047	0.092
2	288083.00	332958.09	288082.89	332958.25	-0.111	0.160
3	288052.50	332838.91	288052.46	332838.75	-0.037	-0.160
4	288162.59	332957.88	288162.70	332957.74	0.111	-0.136
5	288204.31	333038.88	288204.37	333038.71	0.062	-0.160
6	288200.09	332918.41	288199.81	332918.44	-0.284	0.037
7	288226.84	332865.34	288226.61	332865.58	-0.234	0.234
8	288316.19	333040.94	288316.10	333041.02	-0.086	0.086
9	288324.47	332986.19	288324.28	332986.30	-0.185	0.111
10	288337.41	332831.66	288337.17	332831.82	-0.234	0.160

Table 4. GPS receiver classifications*.

	Class I	Class II	Class III	Class IV
Price	\$100 to \$800	\$800 to \$3000	\$3,000 to \$10,000	\$20,000 to \$50,000
Static accuracy [#]	1.5 to 3.5 m	Less than 1 m	10 to 30 cm	Less than 5 cm

*Adapted from work by Sullivan and Ehsani, Ohio State University; and work by Calven and Glen, The University of Georgia (may 2009). [#]How much the reported position varies while the unit is stationary.

and Database Precision (DPr) (Our database precision for instance is 0.01 m). The Georeferencing Error (GfEr) is:

$$\text{GfEr} = 2 \times (\text{GrEr} + \text{DPr}) \quad (3)$$

The Overall Accuracy (OAcc) of digitized points based on our method is:

$$\text{OAcc} = \text{DEr} + \text{GrEr} + \text{GfEr} \quad (4)$$

From Equations (1), Equations (2), Equations (3) and Equations (4) we conclude:

$$\text{OAcc} = \left\{ \left[\frac{2.54 \times 10^{-2}}{R} \right] + (3 \times 10^{-4}) \right\} \times S + (2 \times \text{DPr}) \quad (5)$$

where R is printing & scanning resolution, 1/S is the map scale and DPr is the database precision.

The last formula shows that the preponderant parameter in the accuracy of our method is the map's scale. In our case, we have a database precision of 0.01 m, devices' resolution of 300 dpi and a scale that is between 1/500 and 1/2500. Our overall accuracy is then between 20 cm and 1 m.

4. Conclusion and Perspectives

This paper described the development of a method of automated georeferencing and integrating printed maps in a GIS for collecting addresses from the field. Through this study, we showed limitations of traditional methods based on GPS, especially their limited extent, high cost and varying accuracy (5 cm to 15 m). We introduced the notion of Georeferencing Code, based on the National Area Code (NAC). A method that guarantees accuracy better than 1 m for maps scales under 1/2500. Georeferencing time is reduced from 5 min to less than a minute per map. It also guarantees georeferencing independently of database's content modification.

Even if the method we presented gives an alternative solution that is more accurate and costs less. It still remains limited. The process includes human intervention to perform processings that can be fully automated.

This research may provide a basis for more advanced work on raster maps georeferencing automation and also data collection from the field. Standardizing the use of bi-dimensional barcodes (QR code for example) to store Georeferencing Code and necessary metadata on every printed map can be an interesting application.

REFERENCES

- [1] J. Davis, "Address Base Creation Using Raster-Vector Integration," URISA Annual Conference, Atlanta, Georgia: URISA, 1993.
- [2] P. Eichelberger, "The Importance of Addresses—The Locus of GIS," URISA 1993 Annual Conference, Atlanta, 1993.
- [3] G. Derekenaris, "Integrating GIS, GPS and GSM Technologies for the Effective Management of Ambulances," *Computers, Environment and Urban Systems*, Vol. 25, No. 3, 2001, pp. 267-278. [doi:10.1016/S0198-9715\(00\)00025-9](https://doi.org/10.1016/S0198-9715(00)00025-9)
- [4] H. Rhinane, A. Hilali, A. Berrada and M. Hakdaoui, "Detecting Slums from SPOT Data in Casablanca Morocco Using an Object Based Approach," *Journal of Geographic Information System*, Vol. 3, No. 3, 2011, pp. 217-224. [doi:10.4236/jgis.2011.33018](https://doi.org/10.4236/jgis.2011.33018)
- [5] H. Rhinane, A. Hilali, H. Bahi and A. Berrada, "Contribution of Landsat TM Data for the Detection of Urban Heat Islands Areas Case of Casablanca," *Journal of Geographic Information System*, Vol. 4, No. 1, 2012, pp. 20-26. [doi:10.4236/jgis.2012.41003](https://doi.org/10.4236/jgis.2012.41003)
- [6] A. El Garouani, R. A. Barry, S. El Garouani and A. Lahrach, "Geospatial Database Template for Urban Management in Fez (Morocco)," *Journal of Geographic Information System*, Vol. 4, No. 4, 2012, pp. 335-340. [doi:10.4236/jgis.2012.44038](https://doi.org/10.4236/jgis.2012.44038)
- [7] C. Davis and F. Fonseca, "Assessing the Certainty of Locations Produced by an Address Geocoding System," *GeoInformatica*, Vol. 11, No. 1, 2007, pp. 103-129. [doi:10.1007/s10707-006-0015-7](https://doi.org/10.1007/s10707-006-0015-7)
- [8] L. Simpson and A. Yu, "Public Access to Conversion of Data between Geographies, with Multiple Look Up Tables Derived from a Postal Directory," *Computers, Environment and Urban Systems*, Vol. 27, No. 3, 2003, pp. 283-307. [doi:10.1016/S0198-9715\(02\)00018-2](https://doi.org/10.1016/S0198-9715(02)00018-2)
- [9] M. Malaainine, L. Slaoui, H. Rhinane and L. Baïdder, "Méthode de Saisie en Masse des Points d'Adresses Géolocalisés de la Grande Ville de Casablanca," *First International Congress on GIS and Land management (SIG-GT 2010)*, Casablanca, 20-21 May 2010, p. 101.
- [10] M. Malaainine, L. Baïdder and O. Bachir Alami, "Standards pour l'Intégration Automatique des Plans Issus d'un Système d'Information Géographique: Étude de cas des Plans d'Enquête des Points d'Adresses de la Grande Ville de Casablanca," Taza GIS_DAYS, Maroc, 2012.
- [11] M. Simunek, "On Building and Processing of Large Digitalized Map Archive," *Journal of Systems Integration*, Vol. 2, No. 3, 2011, pp. 3-16.

- [12] S. Coetzee, A. K. Cooper and J. Ditsela, "Towards Good Principles for the Design of a National Addressing Scheme," *25th International Cartographic Conference (ICC 2011)*, Paris, 4-8 July 2011,
- [13] T. Chan, G. Frankish and S. Farrell, "A National Grid Cell Data Infrastructure: Significance of a Hierarchical Grid System and Cooperative Research Centre," *GSDI 12 World Conference*, Singapore, 19-22 October 2010.
- [14] P. Sivakkolundhu, S. Chatterjee, J. John and N. Sharma, "Translating and Geocoding Addresses," US Patent No. 0161334, 2011.
- [15] H. Kirchner, D. Glöckner, G. Bieber and S. Gabrecht, "Addressing Geographic Objects of Unique Location Areas," *Informatik Berichte*, Berlin, 2005.
- [16] B. W. Parkinson, "Global Positioning System: Theory and Applications—Vol. I and II," American Institute of Aeronautics and Astronautics, Washington DC, 1996.
- [17] S. Raghunath, B. L. Malleswari and K. Sridhar, "Analysis of GPS Errors during Different Times in a Day," *International Journal of Research in Computer Science*, Vol. 2, No. 1, 2011, pp. 45-48.
- [18] I. Adeniyi, "Ionospheric Error in GPS Applications," GP1, 46(01062) Poster Session, Plymouth, 2005.
- [19] J. D. Bossler, "Manual of Geospatial Science and Technology," Taylor & Francis Inc, London, 2002
- [20] Birchman and Miller, "Preparing Images for Distribution," *Engineering Design Graphics Journal*, Vol. 64, No. 1, 2000, p. 39.
- [21] A. Caillemer and P. Planques, "Topographie, Photogramétrie," Technip, Paris, 1983.