Journal of Mobile Communication 3 (4): 68-74, 2009
ISSN: 1990-794X

Data Fusion and Spatial Dimension in the Gelocalization Scheme

[1]Fterich Souhir, [2]R. Belgacem Chibani and [2]Mohamed N. Abdelkrim
[1]Research Unit of Modeling, Analysis and System Command,
Institute Superieur d'Informatique et Multimedia de Gabes, 6072 Gabes, Tunisia
[2]Research Unit of Modeling, Analysis and System Command,
Ecole Nationale d'Informatique de Gabes, 6029 Gabes, Tunisia

Abstract: We are trying through this study, to expose data fusion process. In fact, Data Fusion approach is nowadays estimated as a great solution held for accuracy and precision problems applications. Fusion scheme is actually deemed as a new technology, which could be integrated in many scientific fields and could moreover offer many successful results. Many approaches proposed schemes and architecture steps for this concept. Thus, we are trying to propose an overview of these. We will also, as a second step, deal with the importance of space dimension in geographic applications. New research fields such as mobile location are included. In addition, temporal dimension and environment integration could be interesting and provide more accuracy for such applications. That's why, a particular attention will be paid for this point.

Key words: Fusion, fusion process, mobile location, architecture, fusion schemes, space dimension

INTRODUCTION

The evolution of technologies, the immensity of the flows of manipulated information and the multitude of the sources are mainly the characteristic features of the on-going era. Regardless of the diversity of the sectors, one is then facing a complex problem; the management of information. The problem is therefore, to know how to really manage similar quantities of information, while taking into account the diversity of resources (sensors, data bases, expertise), which translate the heterogeneity of the data in question. Indeed, one tries to improve the knowledge of the observed world so as to describe it as possibly well as it is rather than as one wishes it to be. It always should be noted that in addition to the problem of the data type, it is necessary to become aware of the levels of data treatment.

In this respective, the concept of data fusion has been emerged since few years. Thus, here comes the idea of combining various information in entry in order to improve either information in exit or the decision making. The fusion invaded various domains; henceforth, it applies to sectors as the satellite imagery, aerial, medical, robotics, radar, artificial intelligence and recently to mobile communication. Now one attends a considerable receding toward a better consideration of the geographical space surrounding the system in question. So, the integration of the component "spatial dimension" becomes, in this case, an essential stage in such systems. In the Fusion of data, the spatial component is reduced then to a data that one must seize in entry simultaneously to the others in order to better refine the output. It has been shown that information becomes better, while going through a strategy of Data Fusion. However, this option will be of as much use, when one depicts the observed process in >1 dimension. Besides, this is what will be appropriate for applications of geolocalization. Therefore, one will add and integrate a spatial dimension describing the reality to the temporal dimension translating the evolution of the process in order to better determine and describe the observed reality. The Fusion of data, thus encourages mastering the follow-up and the spatial behavior in order to perfectly apprehend the process of combination. The objective of the research lies in this thematic and in the immense framework of the located Fusion. Exactly, it is a matter of releasing the essential stages that possibly lead to the integration of the spatial component within an application of geographical nature as ours. Thus, integration of the spatial component during a simulation of any application first of all is summarized in the retention of a "spatial representation" permitting to describe the behavior of the system. All this supposes evidently well that there is a component of dynamic and non static nature. A state of the art of the approaches of the fusion of data will be elaborated throughout this study. In this, we call to mind the Bayesian principle and we will remain faithful to the theories of the beliefs and possibilities.

MATERIALS AND METHODS

Problem and context: If the simulated phenomena represent environmental systems such as the problem to

Corresponding Author: Fterich Souhir, Research Unit of Modeling, Analysis and System Command, Institute Superieur d'Informatique et Multimedia de Gabes, 6072 Gabes, Tunisia

localize a mobile terminal in mobile telephony, the integration then of "the spatial dimension" during the process of the fusion becomes an essential stage. However, a great number of geographical applications completely neglects to take into consideration and to model this dimension, which is a significant factor of the system. It is often due to the difficulties to identify and to model the geographical entities. As a matter of fact, the geographical space or the spatial environment is generally considered as a passive component of any simulation. Few reasoning and reflections, relating to the integration of the spatial dimension are established, when drawing the diagram of the modeling. Yet, there is an inseparable component for all systems acting and evolving in a geographical space. Hence, the context of this research is to take into account the spatial dimension within the procedure of data fusion altogether within the framework of the geolocalization of mobile sources. Therefore, we aim at mastering the follow-up and the behavior of the spatial dimension during a geographical simulation. In this, if we take this component into consideration so as to improve the decision making, we will gain a lot of interest.

Data Fusion: architecture and principle: In its architecture, the theory of data fusion proposes precisely four stages. Initially, L. Wald proposed the first two stages (Bloch, 2003): modeling and estimation. In fact, he estimated these two stages as the most important ones. Thereafter, two other stages were added to the first ones. So the architecture of the fusion proves to be as it is described by the four stage balance: modeling, estimation, combination, decision. The step of modeling successfully chooses the representation of data and external data by taking in consideration the additional and external ones (in a sense to take into account the spatial dimension in the representation of information). It is viewed to be important because it leads to the totality of the process. However, the estimation step: a stage of analysis of characteristics is proven to be non systematic, although its realization is often advantageous. It is the combination stage that practically summarizes the fusion approach. It is defined as the heart of the fusion. There is a framework in which the means and the techniques, permitting the alliance of the data coming from various sources (Rombaut, 1999), express themselves. Decision is the last stage of the fusion process. This stage rests systematically on criteria of decision whose choice depends directly on the choice of the formalism (choice of modeling and combination) (Rombaut, 1999; Martin *et al.*, 2004).

Accordingly, the objective of the research resides in this thematic. Exactly, we may suggest here a diagram that possibly permits to integrate the spatial dimension within an application to environment and geographical character as the studied case. Thus, several known approaches in the literature arrived to model the problem of uncertainty and imprecision of data and therefore to serve for the process of data fusion.

The Bayesian approach, the method of vote, the theory of the beliefs of Demspter Shafer and the Fuzzy approach described by Zadeh are among the most met theories at the time of modeling of the uncertain in the state of the art section. We are interested in describing the principle of data fusion as it is estimated by the approach of beliefs, probabilities or possibilities (Rombaut, 1999).

The theory of probability: The Bayesian inference is the most privileged method of combination in the theory of probability (Martin *et al.*, 2004). The main aim is the ability to model the information brought by a S_j source on a decision d_i, about an observation X (Fig. 1). This information, as it is described, is represented commonly by:

$$M_i^j(X)$$

In a probabilistic setting, the imperfections of the information themselves are modeled by distributing probabilities or statistical measures to estimate the quantity of information. So this approach allows us only to model the uncertainty of information (Bloch, 1995).

Modeling: It is enough to determine the chances to get information relating to the di decision knowing that we have to handle the S_j source. The stage of modeling then proves to be relying on the bases of conditional probabilities. We express this fact as follows:

$$M_i^j(X) = p(\frac{d_i}{S_j}) \quad (1)$$

Estimation: In this part, we try to consider the distributions of evoked and retained probabilities during the step of modeling. However, a problem often arises when determining these distributions. Indeed, two cases are introduced: the discreet case and the continuous one. In this respective, a valid approach for both cases has been proposed. It stems from the base of kppv (Bloch, 1996, 2003). Here we consider the points k_i, which are the closest neighbors to the S_j source in a basis of apprenticeship as:

$$k = \sum_{i=1}^{n} k_i$$

The estimation of the p (d_i/S_j) quantity estimates the term (k_i/k), where k_i is merely the number of the nearest points to S_j for which one decides d_i.

Bayesian combination: If we consider a singular source, the problem of estimation, which was previously met is meant to determine the p (d_i/S_j) quantity. Henceforth, since we have to deal with several sources, the problem is then how to estimate the p (d_i/S_1,..., S_m) quantity. This quantity translates a combination at the level of the information procured by all sources at hand. Bayes permits to rewrite this combination as;

$$p\left(d_i / S_1,\ldots,S_m\right) = \frac{p\left(S_1,\ldots,S_m / d_i\right)p\left(d_i\right)}{p\left(S_1,\ldots,S_m\right)} \quad (2)$$

Decision: Many criteria of decision have been kept in order to decide during the data fusion. The probabilistic approach came up with a large number of criteria which have the maximal hope, the maximum of similarity and the maximum to posterity. Actually, Martin *et al.* (2004) proposes to choose the maximum to posterity in order to keep and to adapt the decision that satisfies the Eq. 3:

$$p\left(d_k / S_1,\ldots,S_m\right) = \max_{i\in\{1,\ldots,n\}} p\left(d_i / S_1,\ldots,S_m\right) \quad (3)$$

In this study, we consider in a non-exhaustive way, a few works having employed the Bayesian approach for the fusion of information. First of all, we quote Kuncheva's work (Bloch, 2003; Rombaut, 1999) which compares the Bayesian naive approach to the maximum, minimum, average and the vote methods. These approaches are compared in the sense of the fusion of classifiers. The obtained conclusion on various databases shows good performances of the naive approach. However, the probabilities' major problem is that they represent mainly uncertainty and not really imprecision. This often brings about a confusion of the two concepts. Moreover, in the probabilistic stage of modeling, we focus on singletons, which represent the exhaustive and exclusive decisions. This means that the world must be closed and this is really impossible.

The theory of beliefs: In 1967 Dempster, the pioneer of the theory of beliefs, founded the basic principles as well as the main concepts of this thematic, which is also known as the theory of evidence. In its essence, it represents a formal degree of confidence in events allowing a good representation of knowledge. The theory of beliefs is inspired by the theory of probability.

In fact, it represents chances of appearance or confidence. Whereas, the theory of evidence manages much better the imperfections of information through an efficient scheme of the concepts such as imprecision and uncertainty. Indeed, the functions of beliefs are defined in all the subsets of the space of discernment and not simply in singletons as the probabilities, which reflect only the probability of adherence to a given class. Thus, it represents uncertainty. Nonetheless, the theory of beliefs allows representing imprecision and uncertainty at the same time by using functions of mass m, plausibility Pl and belief Cr. In this respect, data fusion had recourse to the approach of beliefs.

Modeling: We notice that D: D = $\{d_1,,,,d_n\}$ space of discernment and all the possible decisions, where each di indicates a hypothesis in favor of which a decision that can be made (typically a class in a problem of multisources classification). The rate of confidence in a given decision is expressed by means of the expressions of the function of mass, belief or plausibility. They are all defined in the subsets of the space of discernment D. In general one imposes m_j (ϕ) and a normalization as Bloch (1996):

$$\sum_{A\in 2^D} m_j(A) = 1 \quad (4)$$

Mass function: The function of mass m_j (A) characterizes the degree of belief in proposition given by the S_j source, and strictly in proposition A. A primordial and prominent concept brought by the theory of beliefs is reduced in considering the measurement of two decisions jointly. Here, the difference with the probabilities lies in the fact that an event A can be the union of two decisions d_1 and d_2. Due to this principle, the theory of beliefs allows us to model imprecision. Indeed, since there is a composite event, we are in a situation translating an imprecision case.

Belief function: This function measures all the belief in an event A; i.e., the intensity that the information provides by the source S_j sustains the proposition A. It translates the concept of an inferior probability known as minimum belief (Bloch, 1996).

Plausibility function: It is often known as a superior probability, it translates the maximal chances of the occurrence of a given proposition.

There is an approach which was evoked in requiring a setting of closed world which is defined as follows:

$$Pl_j\,(A) = \sum_{B\cap A\neq\emptyset} m_j\,(B) \quad (5)$$

Estimation: The estimation of the functions of mass is a difficult problem that does not have a universal solution.

The difficulty increases if one wants to affect masses to the compound assumptions. The simplest manner that one can imagine consists in calculating the masses on singletons in a S_j source by:

$$m_j(d_i) = M_i^j(X)$$

Where $M_j^i(X)$ is often estimated as a probability. Thus, all the masses in the remaining subsets of D are then null. Several functions of mass have been proposed in the literature.

We suggest here a particular estimation of the functions of beliefs through the functions with simple support. The principle of such functions affects the whole mass to a nonempty subset A and with the whole discernment D (Rombaut, 1999).

Estimation of the mass function: The functions of the mass are estimated in the following way:

$$\begin{cases} m_j(A) = s \\ m_j(D) = 1 - s \\ m_j(B) = 0, \forall B \in 2^D, B \neq A, B \neq D \end{cases} \quad (6)$$

Estimation of the belief function: The functions with simple support express the function of belief as follows:

$$\begin{cases} cr_j(B) = s, \text{ si } A \subseteq B, \text{ et } B \neq D, \\ Cr_j = 1, \\ Cr_j(B) = 0, \text{ si non} \end{cases} \quad (7)$$

Estimation of plausibility: Eventually, plausibility is described as following:

$$\begin{cases} Pl_j(B) = 1, \text{ si } A \cap B \neq \varnothing, \\ Pl_j(B) = 1 - s, \text{ si } A \cap B = \varnothing \end{cases} \quad (8)$$

Combination: The combination step known as the core of the technique of fusion. It permits to obtain information which would not be possible to obtain by using only one source. In case of tunable sources, the quality of information in terms of precision and certainty cannot be degraded and the quantity of information remains constant. However, when the sources are discordant, it is sometimes possible to combine the information but the quality is degraded. In this case, there is often a loss of information. Two types of combination are retained: conjunctive and disjunctive.

The context of the conjunctive combination evokes the orthogonal rule of Dempster-Shafer, which is also known as normalized form. The latter considers the work environment to be equivalent to a closed world. A second rule is also evoked, it is about the rule of combination of Smets called non-normalized form, completely in a reverse form of the first. It considers that the framework of the technique is open. The state of the art, also procures other laws of combination, the most known are those of Yager and Hedging.

Decision: This is the last step, in other words it's the choice of the subset D that maximizes a particular criterion. One needs to determine the resulting decision, and in this case, the result can be either a decision singleton d_i or a disjunction.

Various criteria of decision are retained as references in the theory of beliefs by having recourse always to calculations of the functions of mass and beliefs. We mention here the most met criteria in the literature in which there is the maximum of plausibility, belief and eventually the maximum of belief with rejection (Hégarat-Mascle *et al*., 1997).

The theory of beliefs can be applied to a large number of situations. Again, the functions of the mass pave us the way to obtain a rich modeling of imperfections of data. Nevertheless, it presents its own limits. The main defect of this approach is its complexity as it has an exponential growth with the size of the frame of discernment.

The theory of possibilities

The fuzzy: It is often associated with the use of the fuzzy subsets. It was essentially developed by Dubois and Prade. This theory is easily accessiible because it is rather intuitive. Both Didier Dubois and Henry Prade are a considerable reference in the literature. A great number of definitions as well as specificities according to the domain of application such as classification, reasoning by expert system, linear programming and then databases were suggested. Currently, other names such as Hans Bandemer, Wolfgang Nather and L.A. Zadeh intensely denote the state of the art (Bouchon-Meunier and Floue, 2003).

Modeling: While considering the formalism of the section, every S_j source offers information represented by M_i^j on the decision di when observing x. A first approach to define M_i^j is to consider the functions of adherence of fuzzy subsets corresponding to the accuracy of a decision:

$$M_j^i = \mu_i^j(x)$$

Where $\mu^j_i(x)$, expresses the degree of accuracy of the decision di which is taken for the observation x according to the S_j source. In terms of classification, it is the degree of adherence of x to the class C_i.

Estimation: The difficulty of the step of modeling lies in the estimation of the functions of adherence or the distribution of possibilities. Moreover, one has to determine a function F for a decision d_i and a source S_j such as:

$$\mu_i^j(x) = F_i^j(x)$$

The most used functions are those of normalization, monomodal or multimodal. But it is equally possible to determine them by using an algorithm of automatic classification such as C-average fuzzy or possibilities (Martin *et al.*, 2004).

Combination and decision: The steps of combination and decision refer to the procedures and principles used in the theory of beliefs. In this, we especially count on the concept of concordance of sources. Therefore, we mainly insist on the conjunctivitis of the combination.

RESULTS AND DISCUSSION

Issues, principle and choice of the spatial modeling: In order to integrate the spatial dimension in geographical applications of spatial character, the gait to follow to elaborate a modeling consists initially in making the following choices:

- A spatial representation of the reference
- A temporal scale of the modeling by considering the dynamics of the evolution of the environment
- A way of spatial reasoning and its integration in an application of geolocalization

The integration of the spatial component during a simulation of any application tends to adopt a "spatial representation" permitting to really describe the behavior of this element. It supposes well evidently that there is a component of dynamic and non static nature. However, the retained and adopted spatial representation for all applications must satisfy simulation objectives. A spatial component is reduced in a set of spatial entities that describes it. Consequently, the choice of a spatial representation depends directly on the model, which describes these entities. This lets us imagine that finally several structuring of space are possible. We have to retain primarily the continuous and discrete structuring. The definition of the environment or more exactly the geographical space, then tends to depict the receiving space: support of different behaviors of the model. The consideration of the spatial dimension in a given simulation is synonymous with the consideration of this component as a similar actor with any other one having to participate in the simulation of the model. We previously estimated that the environment is of static character; therefore, we were unaware of the importance of this component. However, the geographical space presents some movements and a number of interactions between its own elements. Therefore, stagnation cannot exist in such a context. That is why, we need to become aware of the dynamic character of the environment. Let us note that the dynamism of spatial dimension and the importance of its value during simulation depend closely on the selected temporal scale to describe the phenomenon and its elements as well as the desired objective. In the context for example, we must fix the temporal scale of the movements of owners of the mobile sources as the cell phones.

Integration of the spatial component: The perception, the behavior and the continuous check of the changes of the dynamism of spatial dimension are ensured by a set of components integrated into the system. That is how we can master the evolution of space. Later on, these components will allow us to consider their actions and their responses from their seizure and dynamics. The question which arises now is the following one: at the time of modeling a given phenomenon, would the integration of spatial dimension be always of an increasing importance? Indeed, the choice of the temporal scale proves to be profitable and essential now. Thus, it should be known that the decision about the need for integration of a spatial component as an actor of simulation is an immediate result function of such choice. Also, the tasks of perception, interaction and responses of the spatial components are simultaneously obtained through a sound spatial representation and an adequate choice of the temporal scale. That's why, a phase of spatial reasoning occurs and ensures the logical link which can exist between a phase of perception and possibly a phase of actions and responses. One then may define the spatial reasoning as the process allowing to analyze and generate, from spatial perceptions, the relative answers to such data. One also needs to bear in mind that the choice of the spatial reasoning must obey the requirements of the application as well as the targeted objective. One consequently retains two types of classes of information acting as a basis for all spatial reasoning.

Quantitative information: There is a set of retained measures as the measures of distances and orientations in relation to directions of reference.

Qualitative information: They consist in a certain number of evaluations of approximations (far, near) and of directions (right, north, etc.).

The concept of spatial entity, perception of the spatial dimension: To define the spatial component is to define the spatial entities that describe it as well as their properties. Particularly, it is necessary to give an increasing interest especially to the entities that participate explicitly or implicitly in the dynamics of the considered system and that influence, in the context of communication, the movement and mobility of the mobile sources.

One may presume, from the set of considered spatial entities, the active ones. The quantitative criteria, which most often correspond to the use of properties of the euclidean space and mainly the metrics are privileged tools in representation and in spatial thinking. For 10 years, we have thought about spatial representation, the complement of the quantitative and the emergence of qualitative criteria. The latter have been used in many fields such as the Systems of Geographical Information (SGI).

In this, we may define a qualitative relation between one or more spatial entities and a reference entity. The qualitative spatial thinking essentially uses geometrical objects for the concepts of proximity, orientation, topology, size and shape.

These qualitative relations of orientation and proximity can easily be manipulated by the human spirit than the quantitative information like distances or measures of angles. Qualitative information is however, partial and subjective, which triggers the problems of indecisiveness when expressing the different relations of orientation or proximity. In fact, these relations equally depend on the user and the context. That is what makes them more difficult to be modeled but this reflects better the perception and the representation that a human being can have from the space.

The influence of the context is indeed dominating in the capacity to position what surrounds us in the space. This assertion is confirmed by many works of researchers in cognitive sciences and in this context, it is a very important case in point (Gibson, 1979) and the data function scheme based on spatial dimension component is shown in Fig. 1.

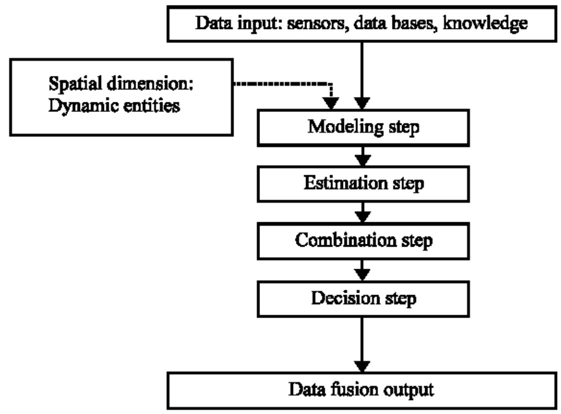

Fig. 1: Data function scheme based on spatial dimension component

CONCLUSION

We notice that appealing the data fusion approach is of an increasing interest. Precisely, no one nowadays can deny the enormous volume of information which we manipulate everyday. The diversity of sources of data: databases, external knowledge and expertise bring a multitude of varieties of information as we have already indicated. In order to improve the quality of information, a universal interest arouses and a particular attraction is felt towards the topic of fusion. The problem of management of the significant amounts of manipulated information is reduced so as to better guide the decision making. In the next decades, the research consists in finding the standards of these researches in order to master and exploit them in the problems of fusion with great complexity. In fact, many sectors are interested in this topic: in particular, the localization of the mobile sources. In this context, we are extremely interested in joining the already adopted data and the temporal dimension with a spatial component translating the behavior of the environment so that we refine the information more and more.

REFERENCES

Bloch, I., 1995. Fondements des probabilites et des croyances: Une discussion des travaux de Cox et Smets. Proceeding of the GRETSI Groupe Etudes du Traitement du Signal et des Images, Sept. 18-21, Juan-les-Pins, France, pp: 909-912.

Bloch, I., 1996. Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. Pattern Recogn. Lett., 17: 905-920.

Bloch, I., 2003. Fusion Informations en Traitement du Signal et des Images. Hermes Science Publication, Lavoisier, France, ISBN: 2-7462-0628-5.

Bouchon-Meunier, C. and M.L. Floue, 2003. Principes, Aide Decision. Hermes Science Publication, Lavoisier, France, pp: 149-204.

Hégarat-Mascle, S., I. Bloch and D. Vidal-Madjar, 1997. Application of dempster-shafer evidence theory to unsupervised classiffication in multisource remote sensing. IEEE Trans. Geosci. Remote Sens., 35: 1018-1031.

Martin, A., G. Sevellec and I. Leblond, 2004. Characteristics versus decisions fusion for sea-bottom characterization. Proceedings of the Colloque Caractérisation *in situ* Des Fonds Marins, Oct. 21-22, Brest, France, pp: 1-6.

Rombaut, M., 1999. Fusion de données images segmentées àl'aide du formalisme de dempster shafer. Proceeding of the GRETSI Groupe Etudes du Traitement du Signal et des Images, Sept. 13-17, Vannes, France, pp: 655-658.