2006 IEEE International Conference on
Systems, Man, and Cybernetics
October 8-11, 2006, Taipei, Taiwan

Constructing Knowledge of the World in Crisis Situations using Visual Language

Siska Fitrianie, Dragos Datcu and Leon J.M. Rothkrantz

Abstract— **Knowing the situational information about the current world is the precondition for setting goals and domains of actions. Special events in the world can also be used to initiate new actions or interrupt ongoing processes. A prototype of a communication interface has been developed which enables users to create visual messages to represent concepts or ideas in mind. The messages are constructed using a spatial arrangement of visual symbols. The interface has been tested by observers in a simulated crisis situation. The system processes the incoming message to build a world model by the employment of ontology. A blackboard of a dynamic ad-hoc network shares the knowledge and updates the display of the mobile devices of the users within the same region of the crisis event.**

I. INTRODUCTION

The terrorist attack of 9/11 was a starting point of research realizations on crisis management in chaotic situations. The need to develop technology to support crisis management has never been more apparent. At Decis Lab (Delft Cooperation on Intelligent Systems), there is a project running on crisis management. We propose an architecture of a distributed system that is designed to handle non-deterministic environments of crisis management caused by the global wired-communication breakdown [23]. Our developed system is based on a mobile (wireless) ad-hoc network that connects all mobile devices, i.e. PDAs. A blackboard structure is used to share and distribute information.

The research described in this paper focuses on maintaining communication in crisis situations to have reliable and complete (up to date) information. In a crisis event, observation reports from all involved parties are essential to have a clear knowledge of the world for supporting critical decisions and actions. Such sensed data, what observers see, hear, smell and even feel or experience, is transformed into reports for others. Speech and text are common ways to report about events. However, the descriptive meaning of these modalities misses a direct link to multimodal observations.

In earlier work, we have investigated icons to represent concepts [9], i.e. objects, actions, or relations. The research was based on the view that human communication involves the use of concepts to represent internal models of humans themselves, the outside world and of things with which they are interacting [20]. As icons offer a potential across language barriers, any interaction using the icons is particularly suitable for language-independent context. Furthermore, besides direct manipulation on the icons allows us to have a faster interaction [15]. As pictorial signs, they can be recognized quickly and committed to memory persistently [10]. Therefore, icons can evoke a readiness to respond for a fast exchange of information and a fast action as a result [17].

According to [5], icons also form a language, i.e. a visual language, where each sentence is formed by a spatial arrangement of icons. The relation between icons and words can represent ambiguous meaning. We encounter this problem by defining the meaning of an individual icon with a predominant word or phrase that is created according to the metaphors appropriate for the context of the language. Using the same ontology, we expect the observers can focus to report only a relevant, objective and unambiguous description of situations that relates to the crisis event.

In contrast to the structure of a sentence in natural language, which is composed by word classes in a structured way, a visual language has a simultaneous structure with a parallel temporal and spatial configuration, e.g. the sign language syntax for deaf people [3], comic illustrations [6], diagrams. Inspired by this, our developed interface proposed a spatial arrangement of text-graphics symbols to represent concepts or ideas. This arrangement may not be in a linear order. Using icons, geometrical features, text and image, a user can share his/her world knowledge to others.

In the following sections we will give an overview of the visual language interface we develop. Further, we will concentrate on the 2D icon string construction and building the world knowledge based on user reports using visual language messages. Finally, we also present our test results.

II. VISUAL LANGUAGE INTERFACE

Crisis management, for events such as natural disasters, technology failures, aviation accidents, and acts of terrorism, relies upon geospatial information about the event itself, its causes, the people and infrastructure affected, resources available to respond, and more. It also relies upon teams of people who must collaboratively derive knowledge from geospatial information and coordinate their subsequent activities. The geospatial information is usually presented via maps, e.g. [16]. Our developed interface also employs maps to represent the location of a crisis event (Fig. 1).

A user can report about a crisis situation by specifying a location on a map where it occurs. The interface provides an observation form where the user can attach visual symbols that describe the situation on that location. The symbols can be icons, geometrical features, icon strings, text or images.

Man-Machine-Interaction Group, Delft University of Technology, Mekelweg 4 2628CD Delft, The Netherlands, {s.fitrianie, d.datcu, l.j.m.rothkrantz}@ewi.tudelft.nl

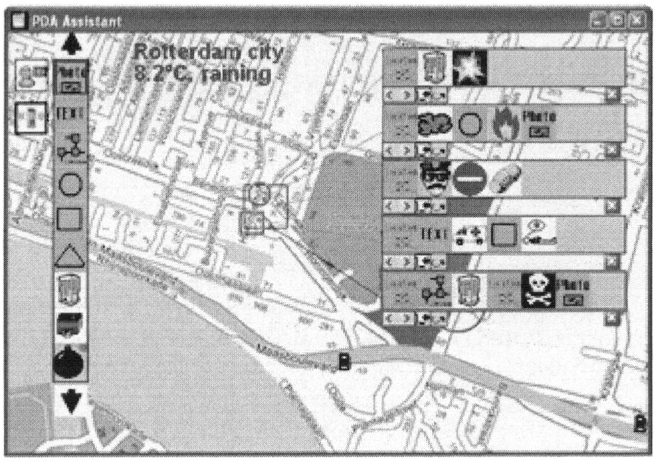

Fig. 1. The developed visual language interface.

The geometrical features, i.e. arrow, line, ellipse, rectangle, and triangle, can be used to represent the area on the map. They can also be used for highlighting or emphasizing an object, an event, or a location on the map. Besides providing icons for representing the observations, e.g. fire, ambulance, victim, etc, the interface also provides icons that represent text, an image (e.g. a photo taken from a camera) and an icon string. The user can select them and a pop-up window will appear where the intended information can be submitted.

To help a user to give an accurate and complete report, the interface provides: (1) menus for deleting and inspecting or altering the observation form; and (2) menus for zooming in/out and moving a map. When the user submits his/her observation(s), the system processes the data and sends the adapted world model around and shares it to other users.

III. TWO-DIMENSIONAL ICON STRING

The syntax analysis of a visual language does not reduce to classical natural sentence syntax. There exists a set of "topic" and "comment" relations, in which a comment explains about a topic [18]. Based on this, we propose a two-dimensional syntax structure, which enables the construction of a sentence in a 2D way. Conventional textual syntax structures are not considered 2D, since the parser processes them as 1D streams of symbols. The 2D syntax structure allows the sentence construction with visual expressions, i.e. spatial arrangements of symbols that represent concepts.

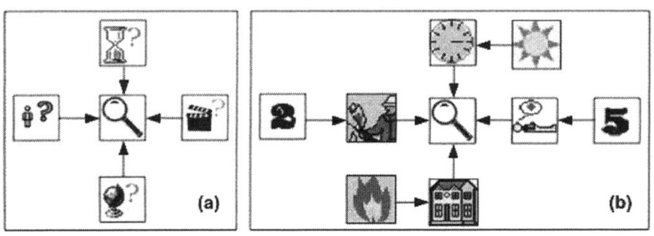

Fig. 2. (a) Hints for the concept "search" and (b) an example of 2D-icon string.

Our approach is inspired by a game's interaction language, Deikto [7]. A sentence, in Deikto, is constructed by an acyclic-graph of concepts, where the predecessor explains the successor concept, i.e. a word/phrase or a clause. To help the players, the game provides hints of what concept(s) can be filled in given a certain word class. Deikto follows a rigid grammar by assigning to each verb the parts of the sentence in its dictionary definition. This revolving verb approach is supported by Fillmore's case grammar [8] and Schank's conceptual dependency theory [22]. In case grammar, a sentence in its basic structure consists of a verb and one or more noun phrases. Each phrase is associated with the verb in a particular case relationship. The conceptual dependency defines the interrelationship of a set of primitive acts to represent a verb using rules. Instead of using word classes (e.g. noun, verb, adjective, etc), both approaches use thematic roles, e.g. agent, patient, instrument, etc.

Recent researches have been done in developing computer-based iconic communication, for example: CD-Icon [2], Sanyog [1], the Elephants memory [12], and VIL [19]. VIL, in particularly, is a (1D) icon language application that is also inspired by the case concept of Fillmore and the verb classification of Schank. This iconic application is also designed to allow people to communicate with each other by constructing sentences solely relying on icons.

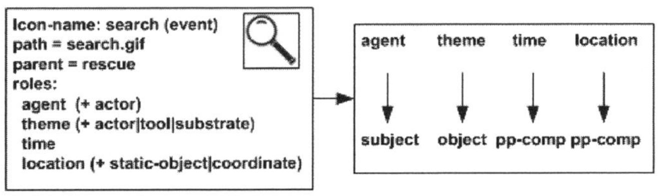

Fig. 3. Schematic view of the two components of a verb lexical definition: semantic types and a link to syntactic arguments.

On our developed icon language application, a user can select icons to form an acyclic-graph of icons that represents a sentence. An icon can be connected by an arrow to another icon, in which the former icon explains the latter. The user may start the string from anywhere she/he wants but as soon as a verb is selected, the structure of the sentence will be determined. Fig. 2 illustrates that two fire-fighters search five victims in the building at 15.00; the building is on fire. We define a case for every icon. For example: the icon "victim" contains number, location, status, name. Particulary for icons that represent verbs, we define their case based on the theory of [8][22]. For this purpose, we follow the frame syntactic analysis used for generating the VerbNet [14]. A lexeme has one or more sense definitions, which consist of a semantic type with associated thematic roles and semantic features, and a link between the thematic roles and syntactic arguments. The definition also defines required and optional roles. Fig. 3 shows a case for the concept "search". The vocabulary is stored in the system's ontology (see next section).

To help the user, on each selected icon, the interface displays attributes, which can be filled in (see Fig. 2). As

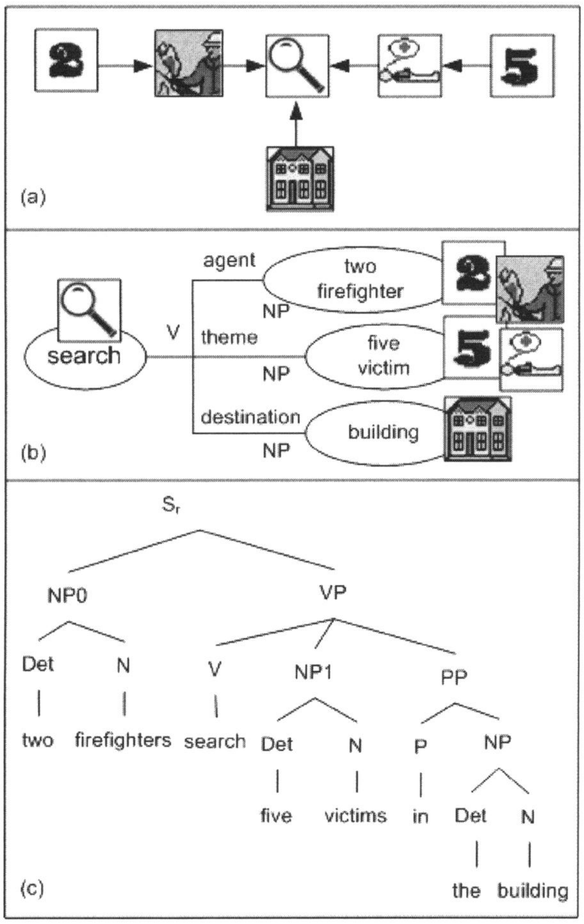

Fig. 4. An example of an input conversion: (a) example of a 2D icon string: "Two firefighters searches five victims in the building", (b) example of mapping the thematic roles to the basic syntactic tree defined by the case of the verb "search", and (c) example of a parse tree as the results of mapping the basic syntactic tree to the TAG trees

the icon is deselected, the hint will disappear to reduce the complexity. A hint icon can be selected and replaced by an icon that is grammatically correct to form a sentence. The approach gives a freedom to users to fill in the parts of a sentence, but at the same time the system can restrict the choices of icons which lead to a meaningful sentence. A user may attach icons that are not given by the hint by inserting a new node on a specific node on the sentence. This means that the new node explains the selected node.

Fig. 3 also shows the link between thematic roles and syntactic arguments for the concept "search". Using this information, the system can convert the iconic sentence into text based using Lexicalized Tree Adjoining Grammar [13] (see Fig. 4).

Based on the case of every icon in a 2D string, a parser processes a 2D stream of icons and maps the thematic roles of them into the basic syntactic tree on the VerbNet-based vocabulary. Transformation of VerbNet's syntactic frames are recoverable by mapping the 2D icon string onto elementary trees of TAG tree families. For this purpose, the parser exploits the system's ontology to have the syntactic argument of every icon in the sentence. The VerbNet structure has provided an explicitly constructed verb lexicon with syntax and semantics. Each verb case gives restricted choices of icons to form the sentence, i.e. by associating thematic roles to semantics features. The meaning of a TAG tree is just the conjunction of the meanings of the elementary trees used to derive it, once appropriate case elements are filled in. Therefore, the syntax analysis and natural language construction can be done simultaneously. By this way, the interface can give a direct feedback on the user's selections.

IV. KNOWLEDGE REPRESENTATION

The knowledge of the visual symbols is stored in the system's ontology represented in W3C-OWL [24]. Fig. 5 shows the taxonomy of the class WorldObject that refers to an entity that is involved in a crisis event. It consists of Dynamic-Objects and Static-Objects. The Dynamic-Object refers to the dynamic content of the world (i.e. objects, processes or events in focus, their status and user's belief), while the Static-Object refers to the static entities (e.g. a building, a building-part, a street, etc). We employ the ontology to store icons that represent nouns, verbs, pronouns, proper-nouns, adjectives, and adverbs, and their properties (i.e. attributes of a case). The icons are the instant of the WorldObject's subclasses, for example: the icon "fire-fighter" is an Actor, the icon "fire" is a Substrate, and the icon "search" is an event.

```
WorldObject
  Static-Object
  Dynamic-Object
     Actor
     Event
     Transportation
     Tool
     Substrate
     Preposition
     Geometrical Features
```

Fig. 5. The WorldObject taxonomy.

The relation between visual symbols and words can have an ambiguous meaning. Therefore, we designed our vocabulary by creating a verbal context that can link both visual and verbal thoughts together to form a symbol that can be remembered and recalled. The ontology represents context that binds verbal and visual symbols together.

Using our visual language interface, users will interact with a large number of icons. To help users to find their way around icons, we approached this by grouping related icons with the same set of concepts. Besides supporting a compact interface, this grouping is powerful way to hint where an icon can be found because the meaning of icons almost exists in the context of other icons [11]. Furthermore the interface also supports faster interaction by three approaches. First, it provides a next icon prediction tool. The tool predicts which symbols are most likely to follow a given segment of visual symbols. When a user selects one of the suggestions, it is automatically inserted into the interface. The probability of the prediction of a visual symbol is estimated with n-grams language modelling. To compute the multi-grams model, the data is collected during the interaction.

Second, the interfaces also provide a search engine based on a keyword what an icon represents. Finally, the interface for creating icon strings in particularly, provides a real time distinctive appearance of which icons can be selected next according to syntactical rules. Our developed interface is designed to prevent the users to make a syntactically incorrect sentence. An icon cannot be selected if it results a grammatically incorrect sentence. All icons are always displayed, however, only active icons can be selected. By this way, the interface could convince its users by offering accurate and reliable interpretation of inputted icon strings since mobile users cannot devote their full attention to operate the application.

The interface provides a direct feedback on each user interaction. Thereby, although some icons are still unknown, the users can learn them on trials. The interface provide ways to its users to return easily to the previous interaction state (e.g. using a delete or clear action), to the previous selection of icons, and to other groups of icons.

V. WORLD KNOWLEDGE CONSTRUCTION

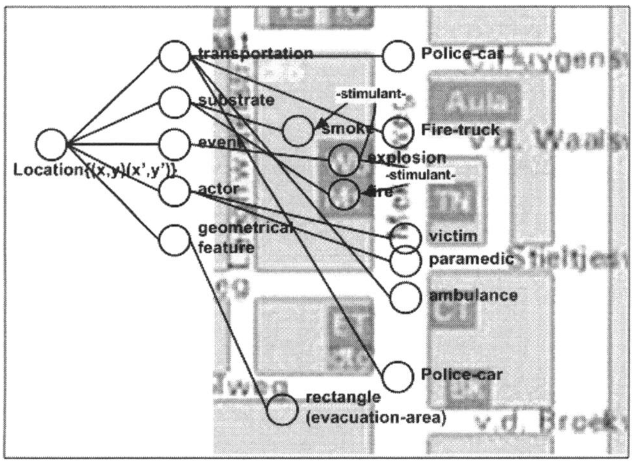

Fig. 6. Graph-based symbolic representation of a crisis event at a certain location.

We view a crisis event as a chain of temporal specific events and a group of dynamic objects (including actors with a specific role) in actions at a certain location in the world. This view represents the dynamic context of the world model, while the static context of the world model is actually the geographical information about the crisis location, e.g. building, street, etc. We represent the geospatial knowledge of the crisis event with topographical data using graphs for data modeling. A graph connects its nodes based on their approximated spatial coordinates in the world. Both contexts are organizes into two layers. Since both data sets are geo-referenced, the objects of both contexts have real-world locations and overlay one another.

The lower nodes of the graph represent (dynamic) objects, actions, or events in the world, that are represented by visual symbols on the user interface. They do not only contain a specific description of the individual, but also their current status (e.g. living condition, substrate condition, is-trapped, dynamic spatial state, etc), their temporal information (e.g. frequency and time point), and their spatial information (e.g. current location, origin, destination, path). The arcs can represent the hierarchy of groups of individuals or a relation between individuals (e.g. has, stimulant, result). At the root of the graph, a node describes a perspective location (the upper-left and bottom-right coordinates) of the crisis event. The illustration in Fig. 6 shows several facts, for example: the fire in a building is caused by an explosion, the rescue teams consist of: firefighters, polices, and paramedics, a save mitigation area has been provided.

The information of the static contexts of the world is retrieved from a geographic database (geodatabase). It includes the information about the blueprint of a building. An object in the geodatabase represents a feature of a real-world entity, e.g. a parcel, a building, a street, a river, or a streetlight.

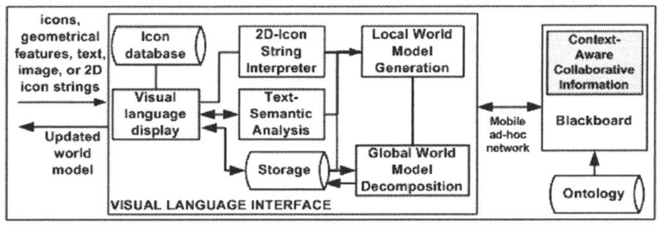

Fig. 7. Schematic architecture of the communication interface.

Fig. 7 shows the communication flows within our developed interface. Extracting the knowledge from a user's input takes three steps. First, our developed system translates the inputs into symbolic representations (local world model) by the use of the domain ontology. The Local World Model Generation extracts information into a symbolic representation using the Text-Semantic Analysis from input text and 2D-Icon String Interpreter from 2D icon strings. Fig. 8 shows the symbolic representation of the example on Fig. 6. All symbols are grouped based on their concept. The ontology provides a natural way to group its elements based on their concepts. For example: symbols of "firefighter" and "victim" are grouped under "actor" and a symbol of "fire-truck" is under "transportation"; both "actor" and "transportation" are grouped under "Dynamic-Object". This developed symbolic representation is sent to the blackboard via a mobile ad-hoc network.

Second, the Context Aware Collaborative Information System (explained in [9]) in the blackboard executes a report selection algorithm. It filters reports based on their prioritized subscriptions and the relevancy value. The relevancy value will be increased if more observers report the same situation at a certain location.

Finally, based on the new constructed global world model, the Global World Model Decomposition updates the user's display. Since the world model contains the spatial information of its objects and events, this process is almost straight-

```
Prespective-area{location=(345,234)(355,244)}{
  Dynamic-Object{
    actor{
      victims{id="1",number=5,location=(350,230)(356,233)}
    }
    event{
      explosion{id="2",begin-time=17.00,
        result="3", location=(350,230)(356,233)}
    }
    substrate{
      fire{id=3,begin-time=17.00,
           location=(350,230)(353,233), stimulant="2"}
    }
    transportation{
      fire-truck{id=4,
                begin-time=17.30,location=(356,233)(359,236)}
      police-car{id=5,
                begin-time=17.30,location=(343,230)(346,233)}
      ...
    }
  }
  Static-Object{
    building{
      building{location=(350,230)(356,233)}
    }
  }
}
```

Fig. 8. Symbolic representation of the world model.

forward. The interface only checks whether the knowledge has been stored in the form of text or 2D icon-strings, otherwise it will be represented by icons on the map. If the interface contains invalid information, the system will solve this ambiguity by reasoning. For example: if the global world model contains information "fire{location=(12,14)(15,17)}", but a small number of users send the same report around or near this location, the reasoning engine will select the former report to update these users' display. This reasoning is supported by a list of common sense knowledge. For example: an explosion results fire and fire results smoke.

VI. EVALUATION

We designed a user test that will allow us to test the usability, the correctness, and the completeness of our developed visual language interface at the same time. The usability determines whether the interface is easy to use, the correctness whether the algorithms work and the completeness to determine whether the interface has provided sufficient number of icons to express concepts in mind. Two groups of four people participated in the test. The tasks were created using scenarios presented in the form of sequence photographs of real crisis situations (see examples on fig. 9(a)). In separate rooms, each group with four participants were asked to report what they saw using the developed visual language and sent the report to each other. While performing the task, they were asked to think aloud. All activities were recorded and logged to be analyzed afterwards.

Smiths measurement of the sense of being lost in hypermedia [21] was adapted to measure disorientation of being lost in an visual language interface. As the term "hypermedia" is used to denote a hypertext document that is made up of interlinked pages or nodes, our visual symbol space also was out of interlinked symbols. Since all information can be displayed only by visual symbols, we thought that a visual language interface also might give cognitive overload and disorientation to its users. For a perfect search, the lostness rating should have been equal to 0.

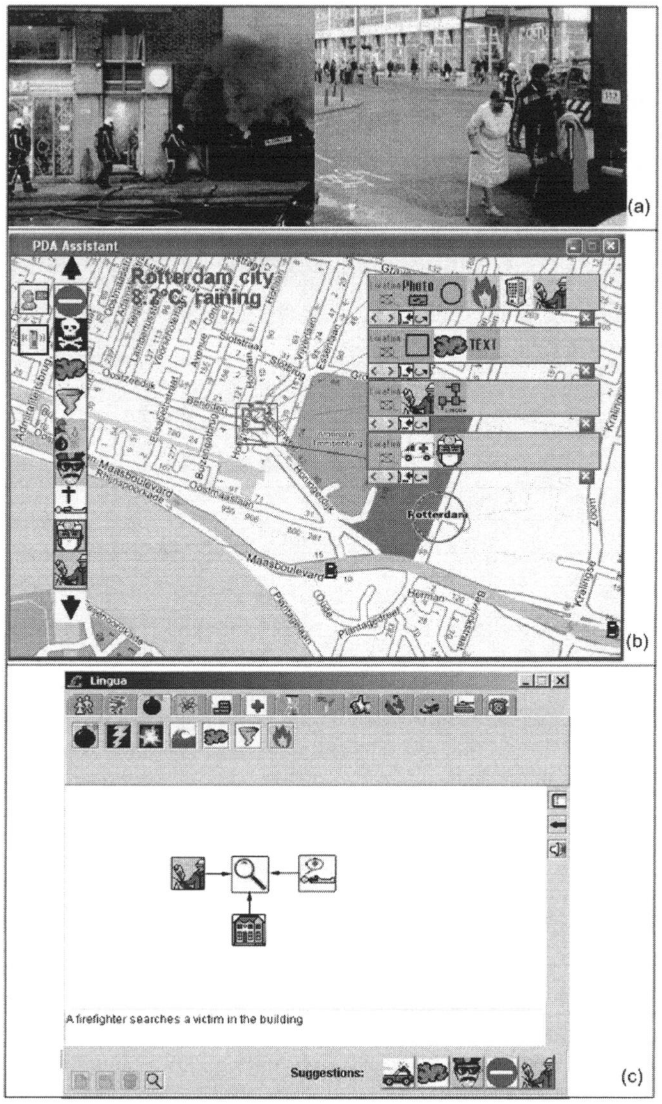

Fig. 9. Evaluation data: (a) Examples of photographs of a real crisis situation, (b) Example of visual language messages sent by a user, and (c) Example of a 2D string created by a user

Fig. 9 shows an example of user interaction with our developed interface. From the results, it showed that our target users were able to express their concepts and ideas in mind using all provided modalities. All participants could compose visual language messages for the given situations.

The experimental results also indicated that our target users still had problems in finding their desired visual symbols from the interface. This was indicated by the high number of the lostness rating of five sequence tasks (see fig. 10). Some of the reasons were referring to some problems in recognizing some symbols, such as: their size and colour contrast between elements in the image and between image and their background. Other reasons were: (1) the participants needed an adaptation time with the interface; (2) The time was also needed to find more relevant concept to represent their message; and (3) they should have rethought another concept that could fit within the problem domain due to limited vocabulary.

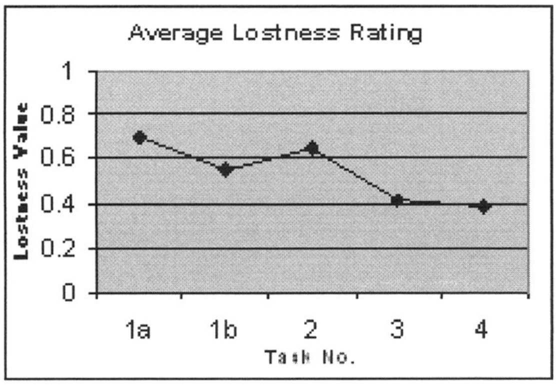

Fig. 10. Average lostness rating.

We also found that some participants took further assumptions from the given photographs. For example: on a photograph showed a firefighter entering a burning building, it was assumed that the firefighter entered the building to find (rescue) victims from fire. Some other users generated some arrangement of visual symbols, which lied outside the domain. However, we decided that there were not incorrect answers, except if the participant did not perform the task at all. Furthermore, we found that users tended to use text and send images only when they could not find a relevant concept from the provided symbols to represent their message.

Apart from these problems, we viewed decreasing lostness rating in terms of improvement of user performances. Our test results also indicated that the hints given while creating 2D-icon string helped the user to compose a complete report. The users also noticed that each time they sent a report, their incomplete reports were updated based on other people's reports in the network. We concluded that our test users only needed a small period of time to adapt to both interfaces.

VII. CONCLUSION AND RECOMMENDATION

As continuation of our research [9], a visual language interface has been developed that is applied for reporting observations on a crisis location. The current version of the interface allows users to describe situations using more modalities, i.e. the combination of icons, geometrical features, 2D-icon strings, text and image on map-based interface to represent concepts or ideas.

NLP has provided a method to interpret and convert visual language messages. We solved the problems of ambiguity and missing information that resulted by this type of messages using rule-based approach. We also proposed Lexicalized Tree Adjoining Grammar with visual symbols to convert icon strings into natural language text.

Using a graph for modeling of the world knowledge, we can represent the topographical data of the crisis event. By the employment of the domain ontology, the world knowledge shares a common semantic representation of the relevance, fidelity and within timeline reports from users.

Our experimental results showed that visual language interfaces can serve as communication mediator. However, future work should be done to gather data about how people might use the interface in their real crisis event and how they experience this. Therefore, the interface design can cover more user requirements in mobile context use.

VIII. ACKNOWLEDGMENTS

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.,

REFERENCES

[1] A. Basu, S. Sankar, K. Chakraborty, S. Bhattacharya, M. Choudhury and R. Patel , Vernacula Education and Communication Tool for the People with Multiple Disabilities, *Development by Design Conference*, Bangalore, 2002.

[2] C. Beardon, CD-Icon: an Iconic Language-Based on Conceptual Dependency, *Intelligent Tutoring Media*, vol. 3, 1992.

[3] D. Bevelier, D.P Corina and H.J. Neville, Brain and Language: a Perspective from Sign Language, *Neuron*, Cell Press, vol. 21, 1998, pp 275-278.

[4] D. Chandler, *Semiotics, the Basic*, Routledge, 2001.

[5] S. K. Chang, G. Polese, S. Orefice and M. Tucci, A Methodology and Interactive Environment for Iconic Language Design", *Intl. Journal of Human Computer Studies*, vol. 41, 1994, pp 683-716.

[6] N. Cohn, *Visual Syntactic Structures, Towards a Generative Grammar of Visual Language*, http://www.emaki.net/essays/, 2003.

[7] C. Crawford, *Erasmatron*, http://www.erasmatazz.com.

[8] C. J. Fillmore, *The Case for Case, in Universals in Linguistic Theory*, Emmon Bach and Robert Harms, Holt, Rinehart and Winston: New York, 1968.

[9] S. Fitrianie and L.J.M. Rothkrantz, Communication in Crisis Situation using Icon language, *Proc. of IEEE ICME 05*, July 2005.

[10] A. Frutiger, *Sign and Symbols, Their Design and Meaning*, van Nostrand Reinholt: New York, 1989.

[11] W. Horton, *The Icon Book*, John Wiley:New York, 1994.

[12] T.I. Housz, *The Elepahants Memory*, 1994–1996.

[13] A.K. Joshi, L. Levy, and M. Takahashi, Tree Adjunct Grammars, *Journal of the Computer and System Sciences*, vol. 10, 1975, pp 136-163.

[14] K. Kipper-Schuler, *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*, PhD thesis proposal, University of Pennsylvania, 2003.

[15] J. Kjeldskov and N. Kolbe, Interaction Design for Handheld Computers, *Proc. of APCHI02*, Science Press, China, 2002.

[16] N. Krahnstoever, E. Schapira, S. Kettebekov and R. Sharma, Multimodal Human-Computer Interaction for Crisis Management Systems, *6th IEEE Workshop on Applications of Computer Vision*, 2002, pp 203.

[17] S.W. Littlejohn, Theories of Human Communication, fifth editions, Wadsworth, 1996.

[18] K. Nakamura, *About American Sign Language*, Deaf Resource Library, http://www.deaflibrary.org, 2002.

[19] N.E.M.P Leemans, *VIL: A Visual Inter Lingua*, Doctoral Dissertation, Worcester Polytechnic Institute, USA, 2001.

[20] L.I Perlovsky, Emotions, Learning and Control, *Proc. of Intl. Symposium: Intelligent Control, Intelligent Systems and Semiotics*, 1999, pp 131-137.

[21] P. Smith, Toward a Practical Measure of Hypertext Usability, Interacting with Computers, Elsevier Science Ltd B.V., vol. 8, 1996, pp 365-381.

[22] R. Schank, *Computer Models of Thought and Language*, W. H. Freeman: San Francisco, 1973.

[23] B. Tatomir and L.J.M. Rothkrantz, Crisis Management using Mobile Ad-hoc Wireless Networks", *Proc. of ISCRAM 2005*, 2005, pp 147-149.

[24] W3C, *OWL: Ontology Web Language*, http://www.w3.org/TR/owl-guide/.