

Brain-Inspired Model for Multiagent Semantic Image Interpretation

Mihaela Luca¹ and Violeta Tulceanu²

¹*Institute of Computer Science, Romanian Academy, Iași Branch, Romania*

²*Faculty of Computer Science, "Al.I.Cuza" University of Iași, Romania*

E-mail: mihaela.luca@academiaromana-is.ro, violeta.tulceanu@gmail.com

(Received September 29, 2015)

We propose a semantic model of interpreting and expressing the contents of images based on a simplified mathematical model of human reasoning, derived from data captured by brain-computer interfacing. Our goal is to allow a mobile agent to capture images and transmit in abstract sentences of a formal language its basic observations to a human or to another intelligent agent. The model permits learning from and communication with other agents, as it is a high-level formalism, independent of agent representation of image. This enables sensor fusion and inter-agent interpretation of input. The visible universe and image content can be understood via the agent's personal observations and/or sentences sent by other agents. The model is applicable for smart home surveillance and street surveillance.

1. Introduction

Our model allows performing semantic queries over an image database. The semantic queries mean that the software can annotate the database so that it can perform complex interrogations over the recognized objects (for instance, "what is the size of the crowd marching towards X?" or "how has traffic varied in the Y time-span in Z location?"). The agents interacting with the database should employ a brain-inspired reasoning mechanism over the input they detect and obtain by querying the database. Thus, they can reach abstract reasoning over the knowledge stored in the database, and can send each other requests about their visible universes. Our empirical results from brainwave authentication [2, 4] have caused us to observe that emotions can be represented as words over an alphabet of semantic dimensions. This implies a non-commutative algebra. Our model particularizes and extends the generic one described by Wang [6]. We believe it can be extended to concepts in general, even abstractions. Also, it allows operating with unknown concepts at a "mold" level, even though the agent/user had no previous experience of any instance of the concept. This model applies to reasoning in swarms of heterogeneous sensor agents.

2. Application architecture description

For rapid information retrieval from large databases, we propose using the model in which the multi-dimensional attribute descriptors are mapped to discrete visual words [1], for each image. The mapping can be performed by using k-means clustering, vocabulary trees or random forests. When the mentioned semantic queries are performed, each attribute from a queried region is mapped to the corresponding visual word and a list of the most common visual words is built, that are eliminated for result accuracy. Once an image is mapped to the constituting visual words, images or corresponding video frames can be retrieved from the database.

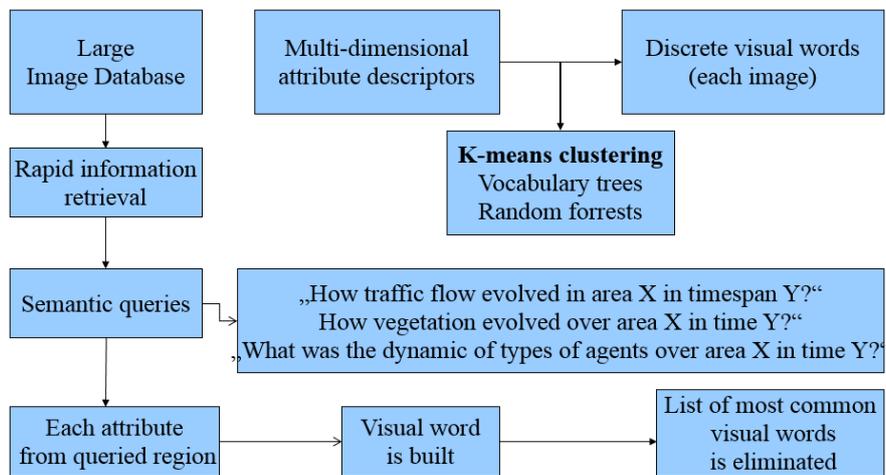


Fig. 1. Discrete visual words processing

We propose an extension of this solution, adding semantic annotations to the visual words, which allows performing complex queries over the database. Because of the constant expansion of the database, it will be necessary to use a hierarchic vocabulary tree, where attribute vectors are hierarchically clustered in a prototype tree. Our design implies support for vocabularies of millions of words, which allows using very specific individual words.

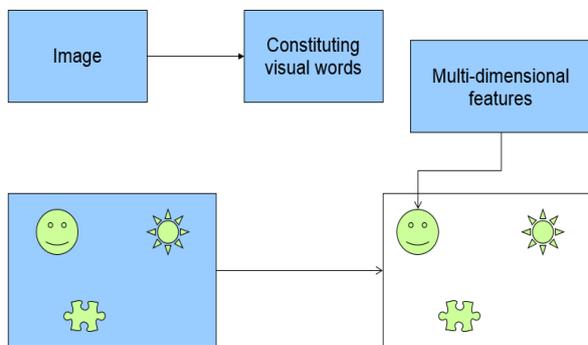


Fig. 2: Extracting constituting visual words from an image

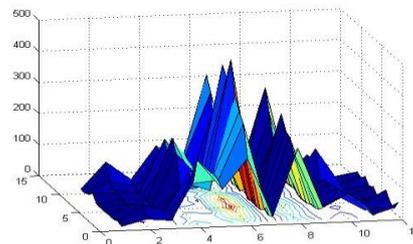


Fig.6: EEG plot for a mental word representation:

Ox: k-means centroid labels(alphabet) forming words over the given EEG channel for each emotion recorded (Oy); here we represent the weights of the centroids; see [2] for details on the model

By analyzing the above mentioned algorithms, we propose the application architecture that comprises decoupled communicating modules. The application relies on an attribute database, built by passing the collected data through the image processing component. The object detection module contains a background detection and elimination component, an edge detection component, a unification component that deals with occlusions, shades, discontinuities etc. The central module is the detector that contains an abstraction implemented by different classes, each specific to a certain object category. The object recognition module contains a classification module using a machine learning technique (SVM, neural networks, PCA) and following an ontological tree.

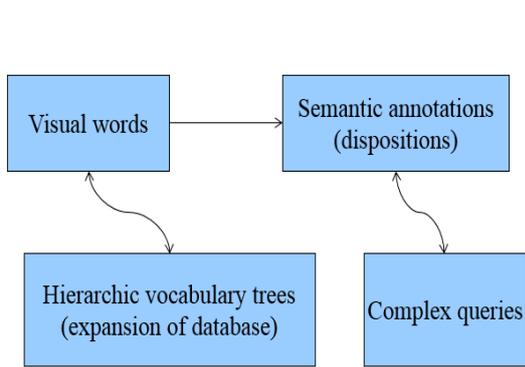


Fig.3: Hierarchic vocabulary trees by semantic annotations

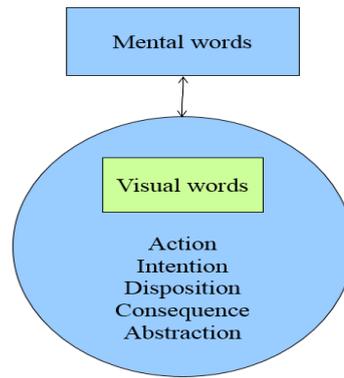


Fig.4: Relation between mental and visual words

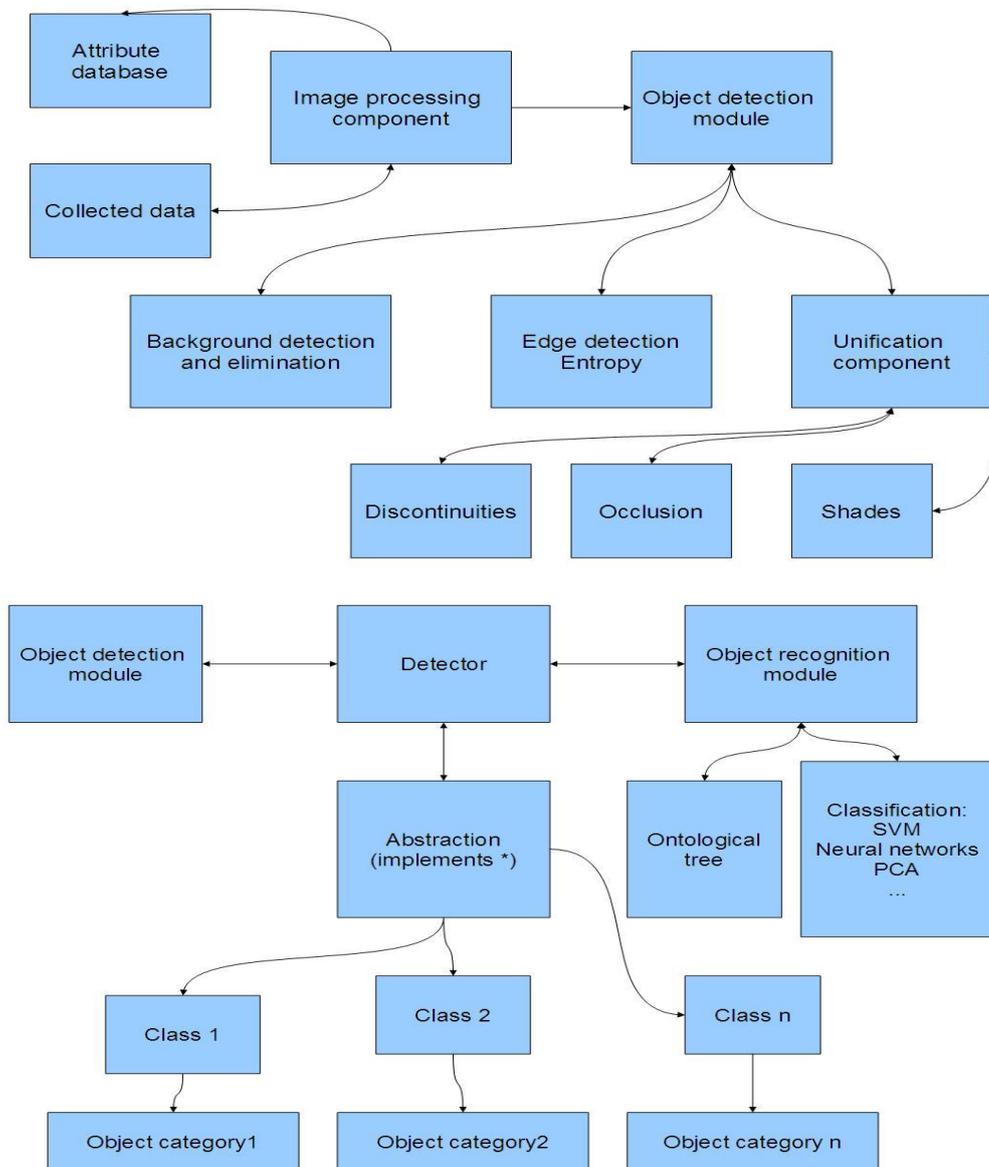


Fig.5: Application architecture description

3. Image Understanding: Brain-Inspired Reasoning and Meta-Cognition

The human mind expresses concepts according to semantic dimensions, “mental words” [4], making it possible to map concepts and trains of thought to algebraic formalisms (inference algebra [7]). We use dispositional explanation [5] for agent reasoning about other agents' output (actions or data) and inference algebra in order to express the action that is captured in the images, by composing the “visual words” with “mental words” in order to express complex sentences about the image. The agents exchange these sentences, expanding their knowledge by querying other agents and making judgments about their output or actions. We base our concept model on concept algebra [6].

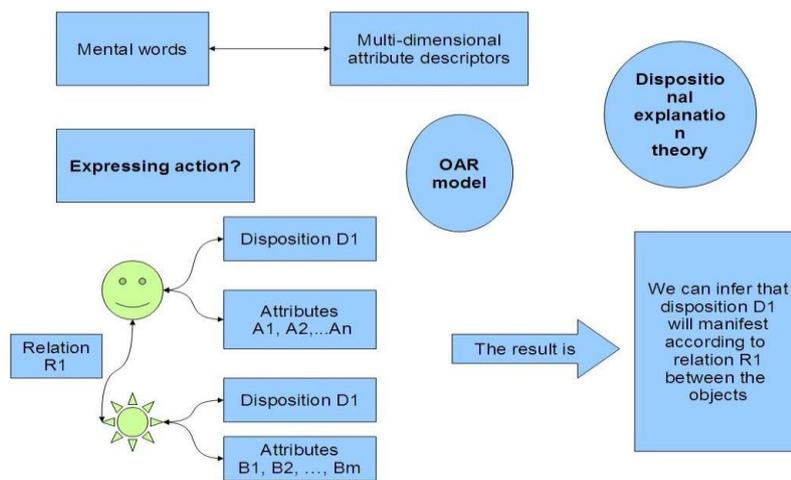


Fig. 7: Overlapping concept dispositions create relationships that activate their manifestation. Two objects in a context and in a relationship will inter-influence

The inference module combines dispositions of concrete concepts (visual words) and abstract concepts (verbs that express the dynamic between concepts) to express the complex relationship patterns by which the concepts link and impact one another. A context given by the union of a set of objects will be propitious to the manifestation of a certain abstract concept, that correspondingly affects concrete concepts and their dispositions. Words are collections of semantic dimensions activated over the human cortex, and correspond to cortical locations. There is an affinity between visual words and words as represented by the processing of the activation pattern they trigger over the human brain. It is possible to express concepts as represented by the human brain in a manner compatible and reducible to the one of expressing visual words or sound input. Thus, both visual and mental words can be represented as multi-dimensional attribute descriptors. They are collections of feature vectors, expressed in a uniform mathematical formalism. Each word is represented in a syntactic manner, representation specific, and most likely, incomprehensible to peers with different types of sensors. Semantic dimensions (D_1, D_2, \dots, D_n) are collections of multi-dimensional feature vectors of unequal lengths. A transducer is defined and employed in order to translate sentences from sets of representation-specific visual words and agent specific abstract words to mental words. This provides us with a high level of abstraction that is non-representation specific, and translatable to a universally comprehensible language. It is possible to reference an object or concept by it's higher category and operate with it even though the agent never encountered an instance of the concept. It is possible to operate with concepts that are beyond the sensory perception of given agents. Learning is possible as well as querying. The high level abstraction can be implemented by different specific agents.

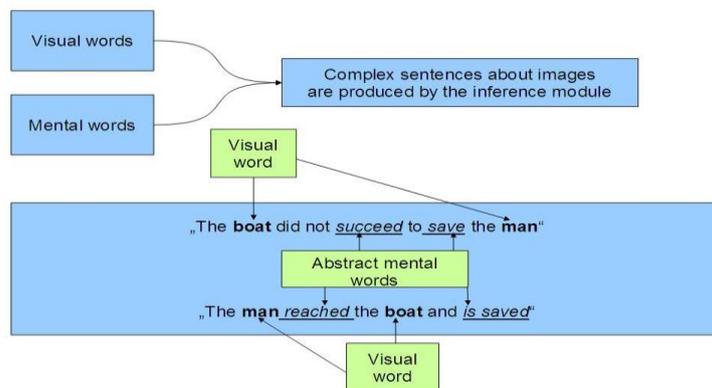


Fig.8: Building sentences to express how dispositions create relationships between concepts

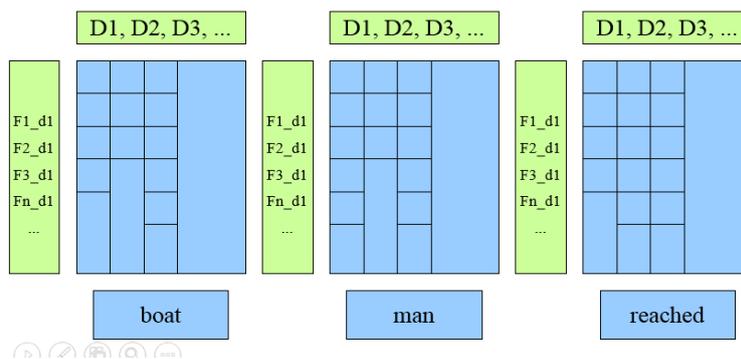


Fig.9: Sentence expressed at a syntactic level of representation as a collection of specific proprietary representations

4. Sensor fusion for image understanding

Let us take into consideration the following agents, pertaining to the same peer group: A1 (equipped with mobile LIDAR, A2 with thermocam, A3 that is a UAV with a camera, flying over the desert, A4 and A5 are the same as A3, but over the city, respectively, over a rural area, and A6 has UV specter vision). All agents implement the abstract interface. However, each has a different visible universe and a different knowledge base. Though an agent may not have an instantiation of a given concept, it has the high-level, abstract mold for learning that concept and referencing it to another agent For example, A3, a desert UAV, may have never experienced the concept of motorcycle before, but it can reference it according to the semantic dimensions it has in its mental words hierarchic vocabulary tree.

We follow with some interaction scenarios for the agents above:

- a) A3 needs to enhance its understanding of the invisible universe by gathering information from A2, that has a different perception and can cover blind spots;
- b) A3 needs to cover different areas, and need comprehensible information it can operate with even though it has never experienced the given areas;
- c) A3 needs to fill in a knowledge gap, and learn a new concept.

To fill in a lack of knowledge, agent A3 needs an instance of a given concept, therefore he might have more ways to proceed: a) asks nearest, most trustable, best reputation or most knowledgeable

agent; b) asks several agents and picks what most have as a main definition;

There can be several specific implementations of a concept, depending on area and specific tasks. A3 temporarily fills in its knowledge gap with information that is deign relevant, until it becomes obsolete or it gathers its own or newer information, and the old definition becomes secondary.

From our experiments in BCI, we observed that:

a) there are differences/variations in brain patterns over time and in different situations;

b) a mental word may have several definitions over time, that appear in different contexts and that are attached weights in order to designate their priority in being considered:

$$W = W_1 \vee W_2 \vee \dots \vee W_n;$$

c) over time, word synonym definitions change;

d) it is necessary to keep track of the dynamic of such changes, in order to know, in a reasoning sequence, what definitions the agent was operating with.

Thus, we must factor in synonym definition order, time and the introduction/elimination of definitions over time. As options, we have: a) dynamic timed I/O automata with weighs or branches (Nancy Lynch); b) non-commutative algebra; c) Hardy algebra; d) Hilbert spaces.

The time labels are necessary for retrospective inference and queries about past-tense actions or events. Trust mechanism can be based on:

a) peer group appurtenance; b) appurtenance and knowledgeability score (number of times it provided correct information over time); c) above and authority (location, responsibility, role etc provide a given agent with authority over the matter); d) above and network (the agent can be trusted to provide correct information because it has access to other agents that are authorities on the matter, though himself he is not); e) most common definition (majority is selected as most knowledgeable). The non-trustable might be: a) non-peer agents; b) potential contact or access with malicious intruder.

5. Conclusion

We propose an architecture that allows semantic queries over large image databases and further expression of the results in an abstract, formal manner that emulates human reasoning. The agents can reason about their data, about the knowledge extracted from the data and communicate with and judge upon other agents knowledge.

References

- [1] R. Szeliski, "Computer Vision: Algorithms and Applications", Springer, New York, 2010.
- [2] V. Tulceanu, "Brainwave authentication using emotional patterns", Int'l Journal of Advanced Intelligence Paradigms, 2015, to appear.
- [3] T. Barbu, M. Costin, A. Ciobanu. "An Unsupervised Content-based Image Recognition Technique", in Springer-Verlag - Series: Studies in Computational Intelligence: New Concepts and Applications in Soft Computing, Berlin, Eds. V. E. Balas, A. V. Koczy, J. Fodor, Vol. 417, pp. 157-164, 2013.
- [4] Violeta Tulceanu, "Comprehensive brainwave authentication using emotional stimuli", In Proceedings of the 20th European Signal Processing Conference, EUSIPCO 2012, Bucharest, Romania, pp.1772–1776 (August 27-31, 2012).
- [5] T. Dima, "Semiotic hypothesis of dispositional explanation," Balkan Journal of Philosophy, 2, pp. 131-139, 2013.
- [6] Yingxu Wang, "On Concept Algebra: A Denotational Mathematical Structure for Knowledge and Software Modeling", International Journal of Cognitive Informatics and Natural Intelligence, 2(2), pp. 1-19, (April-June, 2008).
- [7] Yingxu Wang, "Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning (I)", International Journal of Cognitive Informatics and Natural Intelligence, 5(4), pp. 61-82, (October-December, 2011).