On linguistic descriptions of image content

Isabelle Bloch
LTCI, Télécom Paris, Institut Polytechnique de Paris, France
isabelle.bloch@telecom-paris.fr

Résumé :

Un problème bien connu dans le domaine de l'image et de la vision par ordinateur est le fossé sémantique entre les informations numériques extraites des images et les concepts exprimés de manière symbolique. En raison de la nature des images et de la difficulté d'en extraire des caractéristiques pertinentes, le langage joue un rôle important. D'une part, les descriptions linguistiques des connaissances sur les images et les domaines peuvent être traduites en modèles et algorithmes formels pour guider la recherche d'images, la reconnaissance, la navigation et l'interprétation. D'autre part, la génération automatique de descriptions linguistiques des données est un domaine de recherche qui, bien que récent, évolue rapidement. Par exemple, la question de fournir, à partir de résultats de traitement d'images, une description de haut niveau dans le langage des experts du domaine n'a pas encore été beaucoup abordée, et peut s'inspirer des méthodes de résumé linguistique. L'objectif de cet article est de fournir un aperçu de l'état de l'art dans ce domaine, en montrant l'apport des ensembles flous.

Mots-clés :

Interprétation d'images, description du contenu des images, ensembles flous, fossé sémantique.

Abstract:

A well-known problem in image and computer vision is the semantic gap between the physical level of images, that is features extracted by image processing, and symbolic concepts. Due to the nature of images and the difficulty of extracting meaningful features from them, language plays an important role. On the one hand, linguistic descriptions of prior knowledge about the images and the domains can be translated into formal models and algorithms to guide image retrieval, recognition, navigation, understanding. On the other hand, automatic generation of linguistic descriptions of data is an increasing, though recent, field of research that evolves rapidly. For instance, the question of providing, from image processing results, a high level description in the language of the domain experts was not yet much addressed, and can get inspiration from methods for linguistic summarization. The aim of this paper is to provide an overview of the state of the art in this domain, demonstrating the usefulness of fuzzy sets.

Keywords:

Image understanding, description of image content, fuzzy sets, semantic gap.

1 Introduction

A well-known problem in image and computer vision is the semantic gap between the physical level of images, that is features extracted by image processing, and symbols expressed in a language. On the one hand, linguistic descriptions of prior knowledge about the images and the domains can be translated into formal models and algorithms to guide image understanding. On the other hand, automatic annotation, that is generation of linguistic descriptions of image content, is an increasing, though recent, field of research that evolves rapidly. We present in this paper an overview of the state of the art in this domain[1], focusing on the use of fuzzy sets.

The meaning of image interpretation and understanding adopted here is as follows. The basic step is to recognize (often after or together with a segmentation step) individual objects or structures present in an image. However, image understanding goes some step further and aims at global scene recognition, to obtain high-level descriptions of the objects in their context, including their spatial arrangement. When images are not static but dynamic such as in video processing, then further interpretation steps may include recognition of movements or changes, recognition of actions, gestures, emotions... Image understanding includes also semantic interpretation. Semantics is not present in the image itself, but requires some prior knowledge (for example expressed as formal models) to extract it. All this should then lead to verbal or linguistic descriptions of the image content. This definition includes, but may cover a larger scope, the purely logical view in [63], where the interpretation of an image is defined as a logical model of three sets of axioms (image, scene, depiction). Surprisingly enough, such views relating image understanding and linguistic models was developed in the 1960's, then less addressed, and is now renewed. As early as 1968, a survey of linguistic methods for picture processing, defined as analysis and generation of pictures by computers, with or without human interaction, was proposed in [48]. In [22], a linguistic approach for picture interpretation was proposed, as a pattern de-

[1] Note that we do not deal with the generation of (synthetic) images based on linguistic descriptions in this paper.

scription language. In [62], image understanding is defined as verbal descriptions of the image contents. The need for a semantic layer for spatial language was advocated again later in [7].

Importance of semantics related to images is acknowledged in several domains, including recognition, image understanding, cognitive vision, image retrieval, and image annotation. Although this question has been recognized since the early works in image understanding and computer vision, it was renewed in recent work on semantic image annotation and retrieval, and in recent work linking vision and language, as evidenced for instance by recent workshops on vision and language (for instance VL'16, ICCV19-CLVL), as well as special issues on this topic in journals such as Computer Vision and Image Understanding. The need for semantics of object representation and their epistemic justification was also highlighted in the context of machine learning in [74]. Several aspects related more generally to the philosophy of pattern recognition can also be found in a dedicated issue of Pattern Recognition Letters in 2015 [59]. One main problem is the semantic gap, close to the symbol grounding problem. We show that fuzzy methods provide useful tools to deal with both ontological concepts (often provided as linguistic terms) and concrete domains, and to establish links between them. Two main directions can be identified in connection with linguistic descriptions of images: (i) using linguistic descriptions expressed in a model to guide image interpretation, (ii) deriving linguistic descriptions of images based on image features. Our focus is on methods relying on fuzzy models (for representing vague knowledge, imprecision in images and in concepts, etc.), associated with symbolic and structural models. The suitability of fuzzy sets for representing image information and knowledge is now well known and is not further recalled here (see e.g. [13] for a review). Based on these representations, the first direction, from linguistic descriptions to image understanding, is reviewed in Section 2. This is related to spatial reasoning, where spatial information and knowledge have to be modelled, combined, and then integrated in a reasoning process to provide the final interpretation. The second direction, from image analysis to image content descriptions, in summarized in Section 3.

2 From linguistic descriptions to image understanding

As mentioned in the introduction, to go beyond individual object recognition, image understanding requires descriptions of the spatial organization of objects in images. In knowledge-based approaches, the models should then include such structural knowledge. Models have then to be combined with image information, to finally lead to scene understanding. These steps are summarized in this section, which is to some extent taken from [13].

Representations of structural information. Let us first summarize the main structural representations on which the interpretation methods described next rely. The main information contained in the images consists of properties of the objects (not detailed in this paper) and of relations between objects, both being used for pattern recognition and scene interpretation purposes. Relations between objects are particularly important since they carry structural information about the scene, by specifying the spatial arrangement of objects. These relations highly support structural recognition based on models, and global interpretation of the image. These models can be of iconic type, as an atlas, or of symbolic type, as linguistic descriptions, conceptual or semantic graphs, or ontologies.

Spatial relations are strongly involved in linguistic descriptions of visual scenes. They constitute a very important information to guide the recognition of structures embedded in a complex environment, and are more stable and less prone to variability (even in pathological cases) than object characteristics such as shape, size or appearence. Mathematical models of several spatial relations (adjacency, distances, directional relations, symmetry, betweenness, parallelism...) have been proposed in the framework of fuzzy sets theory, strongly relying on mathematical morphology operators (see [10] and the references therein for a review). For instance, the semantic of a relation such as *close to, to the right of* can be modeled as a fuzzy structuring element, and the dilation of a reference object by this structuring element provides the fuzzy region of space where the corresponding relation is satisfied. More details on fuzzy mathematical morphology can be found e.g. in [12]. Other approaches are based on directions (see e.g. [21] and the references therein).

These fuzzy representations can enrich ontologies and contribute to reduce the semantic gap between symbolic concepts, as expressed in the ontology, and visual percepts, as extracted from the images [38]. Ontologies [36] have been extended to deal with uncertainty and imprecision, using probabilistic or fuzzy approaches, in particular using fuzzy description logics (e.g. [51, 71]). Several spatial ontologies have been proposed, in various domains, such as in [7, 18, 26]. The ideas of linking ontologies expressing fuzzy spatial relations with images were used in particular in the segmentation and recognition methods described in [14, 23, 52, 76]: a concept of the ontology is used for guiding the recognition by expressing its semantic as a fuzzy set, for instance in the image domain or in an attribute domain, which can therefore be directly linked to image information. While the concepts and their use can be defined in a general way, the fuzzy sets expressing the semantics may involve some parameters depending on the context (for instance the notion of "close to" has a different meaning when speaking of brain structures in a medical image or of towns in a satellite image). These parameters can be learned for instance from annotated images [5] or for semantic annotation [60].

Similarly, such spatial relations are useful attributes in graphs and fuzzy graphs, and endow recognition and mining methods based on similarity between graphs with structural information [3, 20, 58], benefiting from the huge literature on fuzzy comparison tools (see e.g. [15]). Spatial relations can also be embedded in conceptual graphs and their fuzzy extensions, as in [76]. Another example is the hierarchical model with fuzzy attributes proposed in [49] for modeling objects (recognition is then based on a fuzzy measure between the model and image processing results).

Fusion. A lot of approaches for image processing and understanding, whatever their level, involve fusion steps. Information fusion becomes more and more important due to the increasing number of imaging techniques. The information to be combined can be obtained from several images, or from one image only, using for instance combination of several relations between objects or several features of the objects, or from images and a model, such as an anatomical atlas or a conceptual graph, or knowledge expressed in linguistic form or as ontologies. The advantages of fuzzy sets rely in the variety of combination operators, offering a lot of flexibility in their choice [9], that can be adapted to any situation at hand, and which may deal with heterogeneous information [30, 80]. The fusion process can be done at several levels of information representation, from pixel level to higher level. Local fusion is often limited because spatial information is not really taken into account, and working at intermediate or higher level (for instance combining several spatial relations to guide the understanding process) is more interesting and powerful. Examples can be found in various domains [23, 47, 52, 56, 64, 76].

Scene understanding. A survey of knowledge-based systems for image interpretation until 1997 can be found in [25]. Here we focus on more recent approaches, and using fuzzy formalisms. Scene understanding using fuzzy approaches mostly belongs to the domain of spatial reasoning, which can be defined as the domain of spatial knowledge representation, in particular spatial relations between spatial entities, and of reasoning on these entities and relations. This field has been largely developed in artificial intelligence, in particular using qualitative representations based on logical formalisms [2]. In image interpretation and computer vision it is much less developed and is mainly based on quantitative representations. Using fuzzy approaches can then be seen as halfway between purely quantitative and purely qualitative reasoning. A typical example in this domain concerns model-based structure recognition in images, where the model represents spatial entities and relations between them. Two main components of this domain are spatial knowledge representation and reasoning. In particular spatial relations constitute an important part of the knowledge we have to handle. Imprecision is often attached to spatial reasoning in images, and can occur at different levels, from knowledge to the type of question we want to answer. The reasoning component includes fusion of heterogeneous spatial knowledge, decision making, inference, recognition. Two types of questions arise when reasoning with spatial relations: (i) given two objects (possibly fuzzy), assess the degree to which a relation is satisfied; (ii) given one reference object, define the area of space in which a relation to this reference is satisfied (to some degree). It has been shown in [11] that the association of three frameworks in a unified way, namely mathematical morphology, fuzzy sets and logics, allows on the one hand matching two important requirements: expressiveness and completeness with respect to the types of spatial information to be represented [1], and on the other hand performing successful reasoning tasks for image understanding.

A common computational representation of structural information to guide image interpretation consists of a graph, where vertices represent objects or image regions (possibly with attributes such as shape, size, color or gray level), and edges carry the structural information (spatial relations between objects, radiometric contrast between regions...). Although this type of representation has become popular in the last 30 years [24], there are still a number of open problems regarding their efficient use for interpretation. One type of approach consists in deriving a graph from the image itself, based on a preliminary segmentation of the image into homogeneous regions, and to express the recognition as a graph matching problem between the image graph and the model graph, which, however, raises combinatorial problems [16, 24]. In [43, 57] an initial labeling of the image regions is performed, and spatial relations are used to refine this labeling or to extract the objects of interest.

All these approaches assume a correct initial segmentation of the images. However this is known to be a very difficult problem in image processing, for which no universally acceptable solution exists: the segmentation is usually imperfect and no isomorphism exists between the graphs to be matched. This leads naturally to the need to find an inexact matching, for instance by allowing several image regions to be assigned to one model vertex, or by relaxing the notion of morphism to the one of fuzzy morphism [20, 58]. As an example, in [27], an over-segmentation of the image is used, which is easier to obtain. Fuzzy relations can then be used to get the final labeling and interpretation [31]. A model structure is then explicitly associated with a set of regions and the recognition is expressed as a constraint satisfaction problem. Some methods rely on fuzzy graph comparison and matching, using genetic algorithm, estimation of distributions algorithms, or graph kernels involving spatial relations. One of the main issues in these methods is the design of an appropriate objective function, guaranteeing that it is optimal for the right solution, which is a difficult task. Still relying on a preliminary segmentation, some approaches have been proposed, for instance using ontologies [38, 53], with fuzzy extensions, besides other types of methods (grammatical or probabilistic ones). Other approaches combining segmentation results and fuzzy models of shapes and spatial relations were proposed, e.g. in [37] for medical images, or in [75] for seismic images, using fuzzy rules. Fuzzy Region Connection Calculus (RCC) [68] can also be used to identify objects based on their mereotopological relations, as done in the crisp case (e.g. [39, 40]).

To overcome the difficulty of obtaining a relevant segmentation, the segmentation and the recognition can also be performed simultaneously. For instance, the method proposed in [14, 23] consists in sequentially segmenting and recognizing each object of interest, in a pre-calculated order [32]. The objects that are easier to segment are considered first and taken as reference. Spatial relations to these reference objects encoded in the structural model, and formalized as fuzzy sets, are used as constraints to guide the segmentation and recognition of other objects. However the extraction of the first objects can be difficult if it is not sufficiently constrained, and due to the sequential nature of the process, the errors are potentially propagated. Backtracking may then be needed, as proposed in [32]. Similar approaches have been used for mobile robot navigation in [33], where linguistic descriptions of a scene, given by a human observer, are translated into fuzzy spatial regions. Another sequential approach was proposed in [72] for vessel tracking in MRI. Starting from an initial fuzzy classification, the authors apply fuzzy rules, involving both image information and geometrical characteristics, to track the vessels and handle the bifurcations.

These approaches can be formalized also as ontological reasoning [38], where both an ontology of the domain can be enriched by fuzzy spatial relations. A first step consists in extracting information from the domain ontology by querying it. The next step consists in actually segmenting (and recognizing) the structure of interest.

To overcome the problems raised by sequential approaches, while avoiding the need of an initial segmentation, another method, still relying on a structural model, but solving the problem in a global way, was proposed in [52]. A solution is the assignment of a region of space to each model object, that satisfies the constraints expressed in the model. A solution is obtained by reducing progressively the solution domain for all objects by excluding assignments that are inconsistent with the structural model. Constraint networks [65] constitute an appropriate framework both for the formalization of the problem and for the optimization. This approach was extended in [76] to fuzzy constraint satisfaction problems (extending [29]) do deal with more complex relations, or involving an undetermined number of objects, and applied to the interpretation of high resolution remote sensing images.

Besides recognition and segmentation, fuzzy spatial relations and more generally fuzzy spatial information have proved useful for other interpretation tasks, such as multiple object tracking [78, 79], graph kernels for machine learning [3], facial expression understanding [61], navigation in unknown environments in robotics [17, 28, 34], among others.

3 From image analysis to image content descriptions

Let us now consider the other way around, where the objective is to start from image features to derive descriptions of the image content in a way as close as possible to natural language. Usually referred to as image annotation, this task aims at identifying some terms (called "tags") that are associated with images to describe their content. These tags are most often related to the recognition of one main object in an image, or a few objects given as an unstructured list. However, in more recent work, new approaches emerged to provide descriptions as whole sentences, i.e. automatic image captioning. This refers to the typical "show and tell" approaches, that benefit from recent advances in machine learning (convolutional networks and deep learning), or use mostly clustering and probabilistic approaches [41, 66, 77]. Such approaches are also used to model queries in image retrieval (see e.g. the reviews in [35, 70]). As an example using fuzzy models, let us cite [6] where fuzzy multimedia ontologies were developed for semantic image annotation. Tags were identified based on consistency of candidate concepts, obtained from SVM classification, and tested using fuzzy description logic reasoning.

Interesting methods using structural representations such as graphs or grammars are worth to mention [54, 73]. For instance in medical imaging, attributed grammars are applied to results of image processing (e.g. detection, skeletonization) to provide a syntactic description of results.

Although a large majority of approaches rely on probabilistic models or learning methods, some of them, in particular structural approaches using graphs or grammars, could be enhanced by fuzzy components to deal with imprecision, vagueness, variability. Still a few fuzzy approaches have been proposed, as described next.

One problem with neural networks is that it may be difficult to understand which rules or reasoning processes they have learned. This question was answered in [45] where satellite image classification was performed using fuzzy neural networks, producing also the fuzzy rules that are actually used by the system, and that are understandable by domain experts, thus providing a description of the image and of how this description was obtained.

Fuzzy sets learned from neural networks were used in [42] in the domain of art image retrieval. The linguistic variables describe "fuzzy aesthetic semantics", in terms of action, relaxation, joy, fear, etc., associated with degrees of satisfaction. Using also a neural network, associated with a fuzzy classifier and an expert system reasoning on low level features, the work in [55] leads to descriptions of facial expressions.

Fuzzy rules were also exploited to generate simple linguistic descriptions of image content [4]. This approach was used for various applications, such as circular structures on Mars, traffic, human gait, medical images... In [8], a clustering and compression method was proposed to provide a small number of fuzzy rules having a linguistic meaning, which constitute fuzzy models that provide linguistic descriptions of low-level features in images. Applications in matching were developed.

At a more structural level, image descriptions involve spatial relations. For instance, in [19], linguistic features describing regions were obtained by fuzzy segmentation, fuzzy spatial relations and locations. From a set of predefined linguistic terms, a brief and accurate description of the whole image is then generated. In [46, 69], previous work by the authors on computation of relative direction was used to derive linguistic descriptions of relative positions in images, associated with a qualitative validity of the description. A typical example of result is "the building is perfectly to the right of the reference object; the description is satisfactory". Conversely, conceptual descriptions, in natural language, of visual scenes can be used to create an image matching these descriptions (an example can be found in [50]).

Another source of inspiration may come from work on summarization. For instance, in [67] the summarization of image databases was based on low-level features and fuzzy labels. While summarization was not much addressed until now for images, several works have emerged for time series and signals, see for instance the special issues on linguistic description of time series in [44], with several papers on automatic generation of linguistic descriptions of data, mapping from non linguistic to linguistic expressions, linguistic summarization. Although some ideas and methods could probably be exploited, the problem when dealing with images is quite different, and there is some work to do to really account for the spatial nature of images and for structural information and knowledge.

One issue in all these approaches is the validation of the obtained linguistic descriptions of the images. Most of the time, a simple comparison with the description provided by a human is performed. This assumes defining a common vocabulary and language, which raises the issue of the level of the description. For instance, describing a brain image may take different forms depending on whether it is intended for a wide public audience, for a patient or for an neurology expert, ranging thus from "an abnormal structure is present in the brain", to "a peripheral non-enhanced tumor is present in the right hemisphere". Providing descriptions with the required granularity can be formalized as an abduction process, where the interpretation is the best explanation of the observation (i.e. the image, or segmentation results), according to the available knowledge, expressed in some logics. The example in [81] exploits the two directions described in this paper. Starting from a linguistic description of the expert knowledge (here neuro-anatomy), a formal model is built (an ontology) and a knowledge base is derived in description logics, which will guide the image interpretation. At the end of the interpretation process, the result is expressed in the same logics, close to the expert natural language.

Acknowledgments The author would like to thank Professor Anca Ralescu for stimulating discussions.

References

[1] M. Aiello. *Spatial Reasoning, Theory and Practice*. PhD thesis, University of Amsterdam, February 2002.

[2] M. Aiello, I. Pratt-Hartmann, and J. van Benthem (Ed.). *Handbook of Spatial Logic*. Springer, 2007.

[3] E. Aldea and I. Bloch. Toward a better integration of spatial relations in learning with graphical models. In G. Ritschard H. Briand, F. Guillet and D. Zighed, editors, *Advances in Knowledge Discovery and Management*, pages 77–94. Springer, 2010.

[4] A. Alvarez-Alvarez and G. Trivino. Linguistic description of the human gait quality. *Engineering Applications of Artificial Intelligence*, 26(1):13–23, 2013.

[5] J. Atif, C. Hudelot, G. Fouquier, I. Bloch, and E. Angelini. From Generic Knowledge to Specific Reasoning for Medical Image Interpretation using Graph-based Representations. In *International Joint Conference on Artificial Intelligence IJCAI'07*, pages 224–229, 2007.

[6] H. Bannour and C. Hudelot. Building and using fuzzy multimedia ontologies for semantic image annotation. *Multimedia Tools and Applications*, 72(3):2107–2141, 2014.

[7] J. Bateman, J. Hois, R. Ross, and T. Tenbrink. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14):1027–1071, 2010.

[8] I. Baturone and A. Gersnoviez. Automatic extraction of linguistic models for image description. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2010.

[9] I. Bloch. Information Combination Operators for Data Fusion: A Comparative Review with Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1):52–67, 1996.

[10] I. Bloch. Fuzzy Spatial Relationships for Image Processing and Interpretation: A Review. *Image and Vision Computing*, 23(2):89–110, 2005.

[11] I. Bloch. Spatial Reasoning under Imprecision using Fuzzy Set Theory, Formal Logics and Mathematical Morphology. *International Journal of Approximate Reasoning*, 41(2):77–95, 2006.

[12] I. Bloch. Duality vs. Adjunction for Fuzzy Mathematical Morphology and General Form of Fuzzy Erosions and Dilations. *Fuzzy Sets and Systems*, 160:1858–1867, 2009.

[13] I. Bloch. Fuzzy sets for image processing and understanding. *Fuzzy Sets and Systems*, 281:280–291, 2015.

[14] I. Bloch, T. Géraud, and H. Maître. Representation and Fusion of Heterogeneous Fuzzy Information in

the 3D Space for Model-Based Structural Recognition - Application to 3D Brain Imaging. *Artificial Intelligence*, 148:141–175, 2003.

[15] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards General Measures of Comparison of Objects. *Fuzzy Sets and Systems*, 84(2):143–153, September 1996.

[16] H. Bunke. Recent Developments in Graph Matching. In *International Conference on Pattern Recognition, ICPR*, volume 2, pages 117–124, Barcelona, Spain, sep 2000.

[17] P. Buschka, A. Saffiotti, and Z. Wasik. Fuzzy landmark-based localization for a legged robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)*, volume 2, pages 1205–1210, 2000.

[18] R. Casati, B. Smith, and A.C. Varzi. Ontological Tools for Geographic Representation. In N. Guarino, editor, *Formal Ontology in Information Systems*, pages 77–85. IOS Press, Amsterdam, 1998.

[19] R. Castillo-Ortega, J. Chamorro-Martínez, N. Marín, D. Sánchez, and J. Soto-Hidalgo. Describing images via linguistic features and hierarchical segmentation. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2010.

[20] R. Cesar, E. Bengoetxea, I. Bloch, and P. Larranaga. Inexact Graph Matching for Model-Based Recognition: Evaluation and Comparison of Optimization Algorithms. *Pattern Recognition*, 38:2099–2113, 2005.

[21] M. Clément, C. Kurtz, and L. Wendling. Fuzzy directional enlacement landscapes for the evaluation of complex spatial relations. *Pattern Recognition*, 101, 2020.

[22] M. Clowes. On seeing things. *Artificial intelligence*, 2(1):79–116, 1971.

[23] O. Colliot, O. Camara, and I. Bloch. Integration of Fuzzy Spatial Relations in Deformable Models - Application to Brain MRI Segmentation. *Pattern Recognition*, 39:1401–1414, 2006.

[24] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(3):265–298, 2004.

[25] D. Crevier and R. Lepage. Knowledge-based image understanding systems: a survey. *Computer Vision and Image Understanding*, 67(2):160–185, 1997.

[26] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.K. Papastathis, and MG Strintzis. Knowledge-assisted semantic video object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10):1210–1224, 2005.

[27] A. Deruyver and Y. Hodé. Qualitative spatial relationships for image interpretation by using a conceptual graph. *Image and Vision Computing*, 27(7):876–886, jun 2009.

[28] D. Driankov and A. Saffiotti, editors. *Fuzzy Logic Techniques for Autonomous Vehicle Navigation*. Studies in Fuzziness and Soft Computing. Springer-Phisica Verlag, 2001. ISBN 3-7908-1341-9.

[29] D. Dubois, H. Fargier, and H. Prade. Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. *Applied Intelligence*, 6(4):287–309, 1996.

[30] D. Dubois and H. Prade. A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, 36:85–121, 1985.

[31] J.B. Fasquel and N. Delanoue. A graph based image interpretation method using a priori qualitative inclusion and photometric relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5):1043–1055, 2019.

[32] G. Fouquier, J. Atif, and I. Bloch. Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. *Computer Vision and Image Understanding*, 116(1):146–165, January 2012.

[33] J. Gasos and A. Ralescu. Towards a Linguistic Instructions-Based Navigation Support System Using Environment Information for Guiding Scene Interpretation. In *FUZZ-IEEE/IFES'95*, pages 1261–1266, Yokohama, Japan, 1995.

[34] J. Gasós and A. Saffiotti. Integrating Fuzzy Geometric Maps and Topological Maps for Robot Navigation. In *3rd International ISCS Symposium on Soft Computing SOCO'99*, pages 754–760, Genova, Italy, June 1999.

[35] A. Goodrum. Image information retrieval: An overview of current research. *Informing Science*, 3(2):63–66, 2000.

[36] T. R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.

[37] J. Hiltner, M. Fathi, and B. Reusch. An approach to use linguistic and model-based fuzzy expert knowledge for the analysis of MRT images. *Image and Vision Computing*, 19(4):195–206, 2001.

[38] C. Hudelot, J. Atif, and I. Bloch. Fuzzy Spatial Relation Ontology for Image Interpretation. *Fuzzy Sets and Systems*, 159:1929–1951, 2008.

[39] G. Landini, A. Galton, D. Randell, and S. Fouad. Novel applications of discrete mereotopology to mathematical morphology. *Signal Processing: Image Communication*, 76:109–117, 2019.

[40] F. Le Ber and A. Napoli. The design of an object-based system for representing and classifying spatial structures and relations. *Journal of Universal Computer Science*, 8(8):751–773, 2002.

[41] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[42] Q. Li, S. Luo, and Z. Shi. Semantics-based art image retrieval using linguistic variable. In *IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, volume 2, pages 406–410, 2007.

[43] Y. Liu, Y. Zhanga, and Y. Gaoa. Gnet: A generalized network model and its applications in qualitative spatial reasoning. *Information Sciences*, 178:2163–2175, 2008.

[44] N. Marin and D. Sánchez Eds. Special issue on linguistic description of time series. *Fuzzy Sets and Systems*, 285, 2016.

[45] L. Mascarilla. Fuzzy rules extraction and redundancy elimination: An application to remote sensing image analysis. *International Journal of Intelligent Systems*, 12(11-12):793–817, 1997.

[46] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera. Linguistic description of relative positions in images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 31(4):573–588, 2001.

[47] N. Milisavljevic and I. Bloch. Possibilistic vs. Belief Function Fusion for Anti-Personnel Mine Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1488–1498, may 2008.

[48] W. Miller and A. Shaw. Linguistic methods in picture processing: a survey. In *Fall Joint Computer Conference, part I*, pages 279–290. ACM, 1968.

[49] K. Miyajima and A. Ralescu. Modeling of natural objects including fuzziness and application to image understanding. In *Second IEEE International Conference on Fuzzy Systems*, pages 1049–1054, 1993.

[50] A. Mukerjee, K. Gupta, S. Nautiyal, M. Singh, and N. Mishra. Conceptual description of visual scenes from linguistic models. *Image and Vision Computing*, 18(2):173–187, 2000.

[51] G. Nagypal and B. Motik. A fuzzy model for representing subjective and vague temporal knowledge ontologies. In *Ontologies, Databases and Applications of Semantics*, Catania, Sicily, Italy, 2003.

[52] O. Nempont, J. Atif, and I. Bloch. A constraint propagation approach to structural model based image segmentation and recognition. *Information Sciences*, 246:1–27, 2013.

[53] B. Neumann and R. Möller. On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82 – 101, 2008.

[54] M. Ogiela and R. Tadeusiewicz. Syntactic reasoning and pattern recognition for analysis of coronary artery images. *Artificial Intelligence in Medicine*, 26(1):145–159, 2002.

[55] M. Pantic and L. Rothkrantz. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18(11):881–905, 2000.

[56] J.-N. Paoli, O. Strauss, B. Tisseyre, J.-M. Roger, and S. Guillaume. Spatial data fusion for qualitative estimation of fuzzy request zones: Application on precision viticulture. *Fuzzy Sets and Systems*, 158(5):535–554, 2007.

[57] G Th Papadopoulos, C Saathoff, HJ Escalante, V Mezaris, I Kompatsiaris, and MG Strintzis. A comparative study of object-level spatial context techniques for semantic image analysis. *Computer Vision and Image Understanding*, 115(9):1288–1307, 2011.

[58] A. Perchant and I. Bloch. Fuzzy Morphisms between Graphs. *Fuzzy Sets and Systems*, 128(2):149–168, 2002.

[59] M. Petillo, editor. *Special issue on "Philosophical Aspects of Pattern Recognition", Pattern Recognition Letters* **64**. 2015.

[60] R. Pierrard, J.-P. Poli, and C. Hudelot. Apprentissage de relations floues pour l'annotation sémantique expliquée avec peu de données. In M. Lefrançois, editor, *17èmes Rencontres des Jeunes Chercheurs en Intelligence Artificielle, RJCIA 2019, Toulouse, France*, pages 18–26, 2019.

[61] A. Ralescu and R. Hartani. Fuzzy modeling based approach to facial expressions understanding. *Journal of Advanced Computational Intelligence*, 1(1):45–61, 1997.

[62] AL Ralescu. Image understanding = verbal description of the image contents. *Journal of the Japanese Society for Fuzzy Theory and Systems*, 7(4):739–746, 1995.

[63] R. Reiter and A. K. Mackworth. A logical framework for depiction and image interpretation. *Artificial Intelligence*, 41(2):125–155, 1989.

[64] F. Rossant and I. Bloch. A Fuzzy Model for Optical Recognition of Musical Scores. *Fuzzy Sets and Systems*, 141:165–201, 2004.

[65] F. Rossi, P. Van Beek, and T. Walsh, editors. *Handbook of Constraint Programming*. Elsevier, New York, NY, USA, 2006.

[66] D. Roy. Learning visually grounded words and syntax for a scene description task. *Computer Speech & Language*, 16(3):353–385, 2002.

[67] R. Saint-Paul, G. Raschia, and N. Mouaddib. Prototyping and browsing image databases using linguistic summaries. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, volume 1, pages 476–481, 2002.

[68] S. Schockaert, M. De Cock, C. Cornelis, and E. Kerre. Fuzzy region connection calculus: Representing vague topological information. *International Journal of Approximate Reasoning*, 48(1):314–331, 2008.

[69] M. Skubic, P. Matsakis, G. Chronis, and J. Keller. Generating multi-level linguistic spatial descriptions from range sensor readings using the histogram of forces. *Autonomous Robots*, 14(1):51–69, 2003.

[70] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[71] U. Straccia. A fuzzy description logic for the semantic web. In E. Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.

[72] Yannis A Tolias and Stavros M Panas. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging*, 17(2):263–273, 1998.

[73] M. Trzupek. Linguistic approach in intelligent image content analysis of heart vessel structures. In *Complex, Intelligent and Software Intensive Systems (CISIS)*, pages 856–859, 2010.

[74] B. Tunç. Semantics of object representation in machine learning. *Pattern Recognition Letters*, 64:30–36, 2015.

[75] L. Valet, G. Mauris, P. Bolon, and N. Keskes. A fuzzy linguistic-based software tool for seismic image interpretation. *IEEE Transactions on Instrumentation and Measurement*, 52(3):675–680, 2003.

[76] M.-C. Vanegas, I. Bloch, and J. Inglada. Fuzzy constraint satisfaction problem for model-based image interpretation. *Fuzzy Sets and Systems*, 286:1–29, 2016.

[77] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2015.

[78] N. Widynski, S. Dubuisson, and I. Bloch. Integration of fuzzy spatial information in tracking based on particle filtering. *IEEE Transactions on Systems, Man and Cybernetics SMCB*, 41(3):635–649, June 2011.

[79] N. Widynski, S. Dubuisson, and I. Bloch. Fuzzy spatial constraints and ranked partitioned sampling approach for multiple object tracking. *Computer Vision and Image Understanding*, 116(10):1076–1094, August 2012.

[80] R. R. Yager. Connectives and Quantifiers in Fuzzy Sets. *Fuzzy Sets and Systems*, 40:39–75, 1991.

[81] Y. Yang, J. Atif, and I. Bloch. Abductive reasoning using tableau methods for high-level image interpretation. In *German Conference on Artificial Intelligence, KI2015*, volume LNAI 9324, pages 356–365, Dresden, Germany, 2015.