Post-print of: de Oliveira C.S., Sanin C., Szczerbicki E. (2020) Context-Aware Indexing and Retrieval for Cognitive Systems Using SOEKS and DDNA. In: Borzemski L., Świątek J., Wilimowska Z. (eds) Information Systems Architecture and Technology: Proceedings of 40th Anniversary International Conference on Information Systems Architecture and Technology – ISAT 2019. ISAT 2019. Advances in Intelligent Systems and Computing, vol 1050. Springer, Cham. DOI: 10.1007/978-3-030-30440-9_2

Context-Aware Indexing and Retrieval for Cognitive Systems using SOEKS and DDNA

Caterine Silva de Oliveira[1], Cesar Sanin[2], Edward Szczerbicki[3]

[1] The University of Newcastle, University Drive, Callaghan, NSW 2208, Australia

[2] Gdansk University of Technology, Gdansk, Poland

Abstract. Visual content searching, browsing and retrieval tools have been a focus area of interest as they are required by systems from many different domains. Context-based, Content-Based, and Semantic-based are different approaches utilized for indexing/retrieving, but have their drawbacks when applied to systems that aim to mimic the human capabilities. Such systems, also known as Cognitive Systems, are still limited in terms of processing different sources of information (especially when structured in different ways) for decision making purposes. This issue becomes significantly greater when past information is retrieved and taken in account. We address this issue by proposing an Ontology-Based Context-Aware Indexing and Retrieval using Set of Experience Knowledge Structure (SOEKS) and Decisional DNA (DDNA). SOEKS and DDNA allow the creation of a multi-modal space composed of information from different sources, such as contextual, visual, auditory etc., in a form of a structure and explicit experiential knowledge. SOKES is composed by fields that allow this experiences to participate in the processes of similarity, uncertainty, impreciseness, or incompleteness measures and facilitate the indexing and retrieval of knowledge in Cognitive Systems.

Keywords: Set of Experience Knowledge Structure (SOEKS), Decisional DNA (DDNA), Cognitive Systems, Ontology-Based Context-Aware, Image Indexing/Retrieval

1 Introduction

Cognitive Systems have gained substantial interest from academia and industry during the past few decade [1]. One of the main reasons for that is the potential of such technologies to revolutionize human life since they intend to work robustly under complex scenes, which environmental conditions may vary, adapting to a comprehensive range of unforeseen changes, and exhibiting prospective behavior like predicting possible visual events. The combination of these properties aims to mimic the human capabilities and create more intelligent and efficient environments [2]. However, perceiving the environment involves understanding the context and gathering visual and other sensorial information available and translating it into knowledge to be useful. In addition, past experiences also plays an important role when it comes to perception [3] and must also be considered as an important element in this process. Perceiving the environment such as humans do still remains a challenge, especially for real time cognitive vision applications, due to the complexity of such process that has to deal with indexing and retrieval of information in a much reduced amount of time [2].

In this paper we aim to address this issue by proposing an Ontology-Based Context-Aware Indexing and Retrieval [4, 5] using Set of Experience Knowledge Structure (SOEKS) [6] and Decisional DNA (DDNA) [7]. SOEKS and DDNA allow the building of a multi-modal space composed of information from different sources, such as contextual, visual, auditory etc., in a form of experiential knowledge, and is composed by fields that allow this experiences to participate in the processes of similarity, uncertainty, impreciseness, or in-completeness measures to facilitate the indexing and retrieval process.

This paper is organized as follows: In Section 2, some literature review on visual content indexing and retrieval is presented with special focus on content-based, context-based and semantic-based approaches. In Section 3 the proposed ontology-based context-aware approach for Cognitive Systems is presented and SOEKS and DDNA described. In Section 4 the general framework and explanation of the given knowledge indexing and retrieval is given for the case of Cognitive Vision Platform for Hazard Control (CVP-HC). Finally, in Section 5, conclusions and future work are given.

2 Literature Review

Image and video indexing and retrieval has been an active research area over the past few decades. There are a range of researches and review papers that mention the importance, requirements and applications of visual information indexing and retrieval approaches [8–11]. In this section, the methodologies are grouped into three main categories, Context-based, Content-Based, and Semantic-based visual content indexing/retrieval. A brief overview is provided for each of those classes, pointing out its drawbacks for application in cognitive systems when applied purely.

2.1 Context-based

Visual content searching, browsing and retrieval implementations are indispensable in various domains applications, such as remote sensing, surveillance, publishing, medical diagnoses, etc. One of the methods used for those purposes is the context-based approach [12]. Information that doesn’t come directly from the visual properties of the image itself can be seen as the context of an image. The context-based approach can be tracked back to 1970s. In context-based applications, the images are manually annotated using text descriptors, key words etc., which can be used by a database management system to execute image retrieval [13]. This process is used to label both image contents and other metadata of the image, for instance, image file name, image format, image size, and image dimensions. Then, the user formulates textual or numeric queries to retrieve all images that are satisfying some of criteria based on these annotations. The similarity between images is, in this case, based on the similarity between the texts.

However, there are some drawbacks in using this approach purely. The first limitation is that the most descriptive annotations must usually be entered manually. Considerable level of human labor is required in large datasets for manual annotation [13]. The second disadvantage of this method is that the most images are very rich in its content and can have more details than those described by the user [14]. In addition, the annotator may give different descriptions to images with similar visual content and different users give different descriptions to the same image. Finally, textual annotations are language dependent [15]. It can be overcome by using a restricted vocabulary for the manual annotations, but, as mentioned previously, it is very expensive to manually index all images in a vast collection [16].

2.2 Content-based

Content-based image retrieval (CBIR) has been introduced in the early 1980s [13]. In CBIR, images are indexed by their visual content (for instance, color, texture, shapes, etc) instead of their context. A pioneering work was published by Chang in 1984. In his research, he presented a picture indexing and abstraction approach for pictorial database retrieval [17]. This pictorial database comprised picture objects and picture relations. To construct picture indexes, abstraction operations are formulated to perform picture object clustering and classification.

Literature on image content indexing is very large [18], and commercial products have been developed using such approach. A common approach to model image data is to extract a vector of features from each image in the database (e.g. image color pixels) and then use a distance measurement, such as Euclidean [19], between those vectors to calculate the similarity between them [13]. Nonetheless, the effectiveness of this approach is highly dependent on the quality of the feature transformation. Often it is necessary to extract many features from the database objects in order to describe them sufficiently, which results in very high-dimensional feature vectors, which demand high storage capacity and increases computational costs. In addition, there is a gap between the high-level image and the low-level image, i.e. there is a difference between what image features can distinguish and what people perceives from the image [14].

2.3 Semantic-based

In order to overcome the limitations of Content-Based and Context Based approaches, Semantic-Based Image Retrieval (SBIR) has been proposed. SBIR can be made by extraction of low-level features of images to identify significant regions or objects based on the similar characteristics of the visual features [14]. Then, the object/region features will be used for semantic image extraction to get the semantics description of images. Semantic technologies like ontology offers promising approach to map those low level image features to high level ontology concepts [5]. Image retrieval can be performed based on the high-level concept (based on a set of textual words, which is translated to get the semantic features from the query).

For the semantic mapping process, supervised or unsupervised learning tools can be used to associate the low-level features with object concept. This procedures are combined with other techniques to close the semantic gap problem, such as using textual word through image annotation process [20]. Semantic content obtained either by textual annotation or by complex inference procedures are both based on visual content [21].

There are a number of papers that address the issue semantic mapping for images. One of the first was Gorkani and Picard [22], who used a texture orientation approach based on a multi-scale steerable pyramid to discriminate cities from landscapes. Yiu [23] applies a similar approach to classify indoor and outdoor scenes, using also color information as features. Wu and Zhu applies ontology to define high-level concepts. In their framework ontology and MPEG-7 descriptors are used to deal with problems arising from representation and semantic retrieval of images. The framework allows for the construction of incrementally multiple ontologies, and shares ontology information rather than building a single ontology for a specific domain not only between the image seekers but also between different domains [24]. For these implementations, the disadvantage is the computational complexity, which can be very high for largescale dataset [25].

3 Context-Aware Approach for Cognitive Systems

The human cognition capabilities is able to receive visual information from the environment and combine it with other sensory information to create perceptual experience [26]. The physical energy received by sense organs (such as eye, ears and nose) forms the basis of perceptual experience. In other words, the combination of sensory inputs are converted into perceptions of visual information (such as dogs, computers, flowers, cars and planes); into sights, sounds, smells, taste and touch experiences. According to Gregory [3], perceptual processes depends on perceiver's expectations and previous knowledge as well as the information available in the stimulus itself, which can come from any organ part of a sensory system. The processing all this information in a lapse of milliseconds makes the humans a very powerful “cognition machine”.

In this context, Cognitive Systems has emerged attempting to meet human capabilities. Cognitive Systems have been defined as “a system that can modify its behavior on the basis of experience” [27]. However, at the present, there is no widely agreed upon definition for cognitive computing. But in general we can say that the term “cognitive system” has been used to define a new solution, software or hardware that mimics in some ways human intelligence.

However, a system capable of processing all information available such as sensor data, visual content from cameras, input signals from machines and any other contextual information available to characterize setting in analysis and at the same time retrieving past experiences for the creation of perceptions still remains a challenge. One of the main difficulties encountered in this case is processing different sources of information that comes structured in different representations at once. This issue becomes significantly greater when past information is retrieved and used in this analysis. Therefore, a knowledge representation capable of building a multi-modal space composed of information from different sources, such as contextual, visual, auditory etc., in a form of experiential knowledge would be a very useful tool to facilitate this process.

3.1 Structuralized representation for visual and non-visual content

Choosing an appropriate image representation greatly facilitates obtaining methods that efficiently learn the relevant information about the category in short time and methods that efficiently match instances of the object in large collections of images [28]. Several researchers have identified that the starting point is to establish an image/video knowledge representation for cognitive vision technologies. However, among all proposed approaches, even though they present some principles for intelligent cognitive vision, none of them provide a unique standard that could integrate image/video modularization, its virtualization, and capture its knowledge [6]. Consequently, we propose to address these issues with an experience-based technology that allows a standardization of image/video and the entities within together with any other information as a multi-source knowledge representation (required for the further development of cognitive vision) without limiting their operations to a specific domain and/or following a vendor's specification. Our representation supports mechanisms for storing and reusing experience gained during cognitive vision decision-making processes through a unique, dynamic, and single structure called Decisional DNA (DDNA) [7]. DDNA makes use of Set of Experience (SOE) in an extended version for the use of storing formal decision events related to image and video. DDNA and SOE provide a knowledge structure that has been proven to be multi-domain independent [29].

Set of Experience Knowledge Structure (SOEKS) and Decisional DNA (DDNA). The Set of Experience Knowledge Structure (SOEKS) is a knowledge representation structure which has been created to acquire and store formal decision events explicitly. SOEKS is composed by four basic elements: variables, functions, constraints, and rules. Variable are the elementary component and it is used to represent knowledge in an attribute-value manner (fundamentally, following the old-fashioned approach for knowledge representation). Functions, Constraints, and Rules are different ways of establishing relationships among these variables. Functions define relations between a dependent variable and a set of input variables to build multi-objective goals. Constraints, on the other hand, act as a way to control the performance of the system in relative to its goals by limiting the possibilities or the set of possible solutions. Lastly, rules are used to express the condition-consequence connection as "if-then-else" and are used to create inferences and associate actions with the conditions under which they should be applied [6].

The Decisional DNA is a structure that has the ability to capture decisional fingerprints companies/organization and also individuals, and has the SOEKS as its basis. Multiple Sets of Experience or multiple SOEs can be gathered, classified, ordered and then grouped into decisional chromosomes. These chromosomes accumulate decisional strategies for a specific application. The set of chromosomes comprise, what is called the Decisional DNA (DDNA) [7].

4 Knowledge Indexing and Retrieval for a CVP-HC

Cognitive Vision Platform for Hazard Control (CVP-HC) being developed to manage risky activities in industrial environments [30]. It is a scalable yet flexible system, designed to work a variety of environment setting by changing its behavior accordingly in real time (context aware). We utilize in this section the knowledge being used to feed the platform to demonstrate the framework for indexing and retrieval of knowledge based on the approach proposed in this paper. More details about the representation of visual and non-visual content in the CVP-HC can be found in [31]. In this section will be focusing in the fields of SOEKS that are relevant for the indexing and retrieval process. Fig. 1 presents the overall architecture of the indexing and retrieval system.

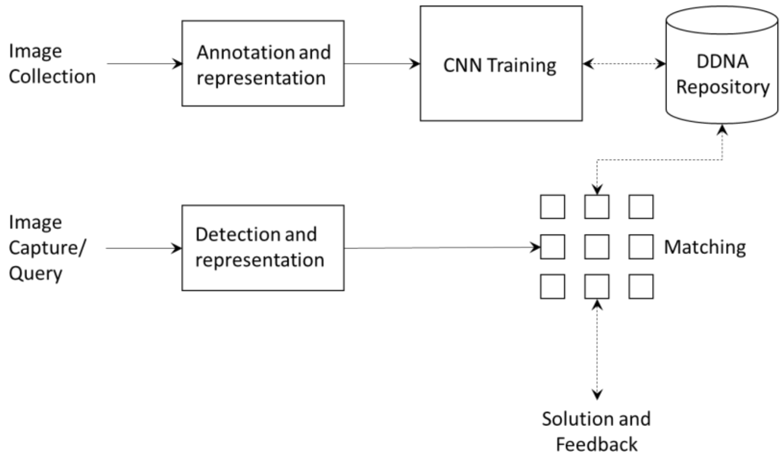

Fig. 1. Overall architecture of the indexing and retrieval system.

4.1 Weight, Range and Priority

Searching for visual similarity by simply comparing large sets of pixels (comparing a query image to images in the database, for example) is not only computationally expensive but is also very sensitive to and often adversely impacted by noise in the image or changes in views of the imaged content [32]. Variables of SOEKS includes fields that allow them to participate in the processes of similarity, uncertainty, impreciseness, or incompleteness measures to facilitate the indexing and retrieval purposes. These fields are: weight, priority, lower range and upper range [33], as shown in Fig 2.

Fig. 2. Fields of SOEKS that allow them to participate in the processes of similarity, uncertainty, impreciseness, or incompleteness measures.

Weight. Each variable has a weight related. Experts may be able to decide the percentage of the weight rate. However, subjective mistakes can be made by human beings, and it will influence decision making. This is a great challenge to find a better automatic and objective way instead of the experts [34]. In the CVP-HC the weights are initialized with the same value for all variables in interaction 0 (before first training) and automatically assigned as the contribution of each variable to the decision making process (variable importance) for next interactions. The summation of all weights for each experience is defined as 1 as shown below:

$$w_{v1} + w_{v2} + w_{v3} + \ldots + w_{vn} = 1 \qquad (1)$$

Therefore, this prediction is easy to be reused and transported in different systems. Hence, knowledge will be expended and shared with different users [35]. The quantity of sets of experience has a great impact on it. If the scope of existing experience is too short for the system to learn, the weight prediction will not be precise.
After those sets of experience are loaded in memory and used to train the first classifier, the generated weights will be allocated to the related variables. A loop is, then, used to assign collected weights to the variables of each experience of the training dataset. The combination of structured knowledge and weights helps addressing the issue of calculating how similar a present and past experience is by taking in account which information is more relevant to this analysis.

This also allows to enrich the system with more knowledge as it runs without compromising the calculation of similarity between experiences of different dimensionalities.

Range and Priority. Each variable has also a priority associated p_v. For the CVP-HC system, priorities are automatically assigned as:

$$p_v = (1 - d_{c\_ulv}) / w_v \qquad (2)$$

Where d_{c_ulv} is degree of correspondence of each variable to a chosen upper level value and goes from 1 (completely similar) to 0 (completely dissimilar) and w_v. the weight of that variable. The priority associated with each SOE is defined by the summation of each individual priority:

$$P_{soe} = p_{v1} + p_{v2} + p_{v3} + \ldots + p_{vn} \quad (3)$$

Where p_{vn} is the priority associated with the n^{th} element of the SOE. By having a priority associated each SOE, we can order the experiences facilitating the searching process.

In addition, in this approach new unique experiences are assigned a higher priority. By selecting experiences with higher priority during the training iterations (for learning purpurses), the system can increase specificity and creating a unique DDNA for that application. Furthermore, by analyzing the curve of SOE's priorities it is possible to infer if a big change in setting has happened and the system requires a new training iteration. Finally, the reusability of a system by another company/organization can be tested over the analysis of the set of priorities of the real time collected experiences with the ones that have been used to enrich and train the system. Fig. 4 shows the overall framework for calculation and update of the priorities in SOEKS over the training iterations.

5 Conclusions and Future Work

The paper presents an Ontology-Based Context-Aware approach to address the issue of indexing and retrieval of multi-modal space information systems, which processes a variety of sources such as contextual, visual, auditory, etc. at once. Such systems, also known as Cognitive Systems aim to mimic the human capabilities and therefore also make use of experiential knowledge. Our approach is an experience-based technology that allows a standardization of visual content and context together with any other information as a multi-source knowledge representation without limiting their operations to a specific domain and/or following a vendor's specification.

Our representation supports mechanisms for storing and reusing experience gained during cognitive vision decision-making processes through a unique, dynamic, and single structure called Decisional DNA (DDNA). DDNA makes use of Set of Experience (SOE) in an extended version for the use of storing formal decision events related to cognitive vision systems. The knowledge indexing and retrieval is facilitated by the use of *weight, ranges and priorities*. These fields are part of the SOEKS and allow the experiences to participate in the processes of similarity, uncertainty, impreciseness, or incompleteness measures.

The method discussed in this paper will be evaluated when the experience is enriched with other sensorial data. Suitability of reusing experiences will also be explored for different case scenarios.

References

1. Sanin, C., Haoxi, Z., Shafiq, I., Waris, M. M., de Oliveira, C. S., & Szczerbicki, E. (2018). Experience based knowledge representation for Internet of Things and Cyber Physical Systems with case studies. Future Generation Computer Systems.
2. Vernon, D.: The space of cognitive vision. In Cognitive Vision Systems, 7-24, Springer, Berlin, Heidelberg (2006).
3. Gregory, R. L. (1973). Eye and brain: The psychology of seeing. McGraw-Hill.
4. S. Malik and S. Jain, "Ontology based context aware model," 2017 International Conference on Computational Intelligence in Data Science(ICCIDS), Chennai, 2017, pp. 1-6. doi: 10.1109/ICCIDS.2017.8272632
5. U. Manzoor, N. Ejaz, N. Akhtar, M. Umar, M. S. Khan and H. Umar, "Ontology based image retrieval," 2012 International Conference for Internet Technology and Secured Transactions, London, 2012, pp. 288-293.
6. C. Sanin, E. Szczerbicki, Experience-based Knowledge Representation SOEKS. Cybernet Sys. 40(2) (2009) 99-122.
7. Sanin, C., Toro, C., Haoxi, Z., Sanchez, E., Szczerbicki, E., Carrasco, E.,. & Man-cilla-Amaya, L. (2012). Decisional DNA: A multi-technology shareable knowledge structure for decisional experience. Neurocomputing, 88, 42-53.
8. M. De Marsicoi, L. Cinque, S. Levialdi, Indexing pictorial documents by their content: A survey of current techniques, Image and Vision Computing 15 (1997) 119–141.
9. Y. Rui, T. Huang, S. Chang, Image retrieval Past, present, and future, in: International Symposium on Multimedia Information Processing.
10. Y. Rui, T. Huang, S. Chang, Image retrieval: current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation, pp. 39–62.
11. D.B. H. Muller, N. Michoux, A. Geissbuhler, A review of content-based image retrieval systems in medical applications clinical benefits and future directions, International Journal of Medical Informatics 73 (2004) 1–23.
12. Westerveld, T. (2000, April). Image retrieval: Content versus context. In Content-Based Multimedia Information Access-Volume 1 (pp. 276-284).
13. Raveaux, R., Burie, J. C., & Ogier, J. M. (2013). Structured representations in a content based image retrieval context. Journal of Visual Communication and Image Representation, 24(8), 1252-1268.
14. Alkhawlani, M., Elmogy, M., & El Bakry, H. (2015). Text-based, content-based, and semantic-based image retrievals: A survey. Int. J. Comput. Inf. Technol, 4(01).
15. H. Tamura and N. Yokoya, "Image database systems: A survey," Pattern recognition, vol. 17, 1984, pp. 29–43.
16. Oard, D.W. and Dorr, B.J. (1996) A Survey of Multilingual Text Retrieval. Technical Report UMIACS-TR-96-19, University of Maryland, Institute for Advanced Computer Studies.
17. S.H. Liu, S.K. Chang, Picture indexing and abstraction techniques for pictorial databases, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 6 (4) (1984) 475–483.
18. R. Datta, D. Joshi, J. Li, James, Z. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys 39 (2006) 2007
19. Danielsson, P. E. (1980). Euclidean distance mapping. Computer Graphics and image processing, 14(3), 227-248.

20. H.H. Wang, D. Mohamad, and N. Ismail, “Image Retrieval: Techniques, Challenge, and Trend,” International conference on Machine Vision, Image processing and Pattern Analysis, Bangkok, Citeseer, 2009.
21. N. Shanmugapriya and R. Nallusamy, “Anew content based image retrieval system using GMM and relevance feedback,” Journal of Computer Science 10 (2): 330-340, 2013.
22. Gorkani, M. M., & Picard, R. W. (1994, October). Texture orientation for sorting photos" at a glance". In International conference on Pattern recognition (pp. 459-459).
23. Yiu, E. C. (1996). Image classification using color cues and texture orientation (Doctoral dissertation, Massachusetts Institute of Technology).
24. Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. International Journal of Computer Vision, 27(2), 107-126.
25. Zin, N. A. M., Yusof, R., Lashari, S. A., Mustapha, A., Senan, N., & Ibrahim, R. (2018, June). Content-Based Image Retrieval in Medical Domain: A Review. In Journal of Physics: Conference Series (Vol. 1019, No. 1, p. 012044). IOP Publishing.
26. Bandura, A. (1989). Human agency in social cognitive theory. American psychologist, 44(9), 1175.
27. Hollnagel, E., & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. CRC Press.
28. J. Amores, N. Sebe and P. Radeva, "Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1818-1833, Oct. 2007.
29. Sanin, C., Toro, C., Haoxi, Z., Sanchez, E., Szczerbicki, E., Carrasco, E.,. & Man-cilla-Amaya, L. (2012). Decisional DNA: A multi-technology shareable knowledge structure for decisional experience. Neurocomputing, 88, 42-53.
30. de Oliveira, C. S., Sanin, C., & Szczerbicki, E. (2019). Visual Content Learning in a Cognitive Vision Platform for Hazard Control (CVP-HC). Cybernetics and Systems, 50(2), 197-207.
31. de Oliveira, C. S., Sanin, C., & Szczerbicki, E. (2019, April). Towards Knowledge Formalization and Sharing in a Cognitive Vision Platform for Hazard Control (CVP-HC). In Asian Conference on Intelligent Information and Database Systems (pp. 53-61). Springer, Cham.
32. Deserno, T. M., Antani, S., & Long, R. (2009). Ontology of gaps in content-based image retrieval. Journal of digital imaging, 22(2), 202-215.
33. Sanin, C., Szczerbicki, E.: Using XML for implementing set of experience knowledge structure. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3681, pp. 946–952. Springer, Heidelberg (2005).
34. Sanín, C. A. M. (2007). Smart knowledge management system. University of Newcastle.
35. Wang, P., Sanin, C., & Szczerbicki, E. (2012). Enhancing Set of Experience Knowledge Structure (SOEKS) WITH A Nearest Neighbor Algorithm RELIE-F. Information Systems Architecture and Technology, 13.