Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 2249–2257

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Visual content representation and retrieval for Cognitive Cyber Physical Systems

Caterine Silva de Oliveira[a]*, Cesar Sanin[a], Edward Szczerbicki[b]

[a] *The University of Newcastle, University Drive, Callaghan, NSW 2208, Australia*
[b] *Gdansk University of Technology, Gdansk, Poland*

Abstract

Cognitive Cyber Physical Systems (C-CPS) have gained significant attention from academia and industry during the past few years. One of the main reasons behind this interest is the potential of such technologies to revolutionize human life since they intend to work robustly under complex visual scenes, which environmental conditions may vary, adapting to a comprehensive range of unforeseen changes, and exhibiting prospective behavior like predicting possible events based on cognitive capabilities that are able to sense, analyze, and act based on their analysis results. However, perceiving the environment and translating it into knowledge to be useful for the decision making process, still remains a challenge for real time applications due to the complexity of such process. In this paper, we present a multi-domain knowledge structure based on experience, which can be used as a comprehensive embedded knowledge representation for C-CPS, addressing the representation of visual content issue and facilitating its reuse. The implementation of such representation has been tested in a Cognitive Vision Platform for Hazard Control (CVP-HC) which aims to manage of workers' exposure to risks in industrial environments, facilitating knowledge engineering processes through a flexible and adaptable implementation.

Keywords: Decisional DNA; Set of Experience Knowledge Structure; Knowledge Engineering; Image Representation; Cognitive Systems; Cognitive Cyber Physical Systems

* Corresponding author. Tel.: +61-416-312-132.
E-mail address: Caterine.SilvadeOliveira@uon.edu.au

10.1016/j.procs.2019.09.400

1. Introduction

With the advance of Internet of Things, big data analysis, wearable devices and cloud computing, Cyber-Physical Systems (CPS) have been widely used. Applications such as smart grid, environment monitoring, surveillance, vehicular networks, and industrial control and automation applications make use of cognitive methods to improve the system performance and are known as Cognitive Cyber-Physical Systems (C-CPS). C-CPS contain nodes with cognitive capabilities that are able to sense, analyze the environment, and act based on their analysis results. In addition it contains the cyber physical element which enables interconnection of all elements of a given process [1].

In Cognitive Cyber-Physical Systems, knowledge and learning are central elements for the decision making process [2]. To be readily articulated, codified, accessed and shared across interconnected elements and platforms, knowledge must be represented in an explicit and structured way [3]. This is not a trivial task given the nature of knowledge itself, and the different source of information needed to be synthesized to generate and represent this knowledge. In addition, decision making can only be improved by human interference or if the system has autonomous learning capabilities. In both ways, past experiences are taking in consideration. Therefore, to build an advanced C-CPS, past experience must also be part of the represented knowledge.

However, C-CPS with those characteristics still remains a challenge [4]. To our acquaintance, numerous groups of research have proposed smart cognitive system applications [5]; nonetheless, most of these methodologies oversight the potential of using experience that grows continuously over real time operation and that might enrich the application with smartness while, at the same time, creating decisional fingerprints. This single concept allows the system knowledge growth through daily operation autonomously, just like human experience do in real life as well as facilitates its sharing across different platforms.

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 found, even though they present some principles for intelligent cognitive vision, none of them provide a unique standard that could integrate image/video modularization and virtualization, together with information from other sources (wearable sensors, machine signals, context, etc.) 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 systems) without limiting their operations to a specific domain and/or following a vendor's specification. Our representation supports mechanisms for storing, reusing and sharing experience gained during decision-making processes through a unique, dynamic, and single structure called Decisional DNA (DDNA). DDNA makes use of Set of Experience (SOE), which has been used for several different domains and has now been extended for the use of formal decision events related to image and video [7]. DDNA and SOE provide a knowledge structure that has been proven to be multi-domain independent; and therefore, suitable for use in Cognitive Cyber-Physical Systems [8, 9].

This paper is organized as follows: In Section 2, some fundamental concepts are presented with special focus on raw pixel, feature vector and annotations as well as SOEKS and DDNA. In Section 3 the Knowledge Representation using SOEKS and DDNA to address the visual content representation is introduced. In Section 4 the implementation of the proposed representation in a Cognitive Vision Platform for Hazard Control (CVP-HC) is explained. Finally, in Section 5 conclusions and future work are given.

2. Fundamental Concepts

In order to offer a more complete view, we briefly introduce concepts that have driven the proposed research as well as the technologies involved.

2.1. Raw Pixel Values

Recent years have seen a quick increase in the size of digital image collections. Daily, both military and civilian equipment generates gigabytes of images data [10]. However, this data cannot be easily accessed or made use of, unless it is organized so as to allow efficient browsing, searching, and retrieval [11].

Organizing visual information digitally to facilitate search and potentially reuse is a challenging task. This process may include a subset of procedures such as images collection, cleaning, standardization, structuring, indexing, and grouping. The difficulty to organize visual content increases exponentially as the dataset grows, and how exactly such data can be harnessed and structured for easy access remains a critical problem for large-scale platforms designers [12]. For Cognitive Cyber Physical Systems, the representation in this case plays a very important role as it must ensure preservation of visual content, and at the same time enabling its search and retrieval in real time.

In this context, uncompressed raw intensity value of each pixel is useful for machine learning and computer vision applications once they mostly need direct access to the pixel data without the burden of the complex computations required to determine the location of pixels within a compressed data stream [13]. This also simplifies reading and displaying the images by dealing with pixel data directly, speeding up retrieval process. Furthermore, raw intensities values can achieve surprisingly high recognition rates in computer vision applications when used directly as input of, for instance, Deep Neural Networks [14].

On the other hand, comparison of image samples based on raw pixel data can be very computationally expensive, and impractical for real time applications. Instead, dimensionality reduction is used to improve the computational efficiency. These steps are closely related: representing an image by a high-dimensional vector usually provides better exhaustive search results than with a low-dimensional one and is also better in preserving the visual information. However, high dimensional vectors are more difficult search efficiently. In contrast, a low dimensional vector is more easily compared with other vectors and can accelerate the searching process, but its discriminative power is lower and might not be sufficient for recognizing objects or scenes (loss of visual information) [15].

2.2. Vector of selected features

To reduce the dimension of the vectors for searching purposes, feature selection methods is usually used, such as, Principal Component Analysis (PCA) [16], supervised Linear Discriminant Analysis (LDA) [17], mRMR [18], among others. However, the best methodology for feature selection depends on the characteristics of the image and application. For instance, PCA and LDA are popular due to their simplicity and effectiveness. However, PCA and LDA may fail to discover essential data structures that are nonlinear [19]. For systems that aim generality and knowledge sharing this result in an impasse: the platform should enable the implementation of the method the suits the given application the best, but the subset of features generated by each methodology may differ from each other which compromise the comparison and reuse of that experience.

In fact, feature selection is a growing research area, and the competence of new methods in generating more compact and meaningful set of features is increasing every year. To demonstrate the growing interest of researchers in this topic along the years we have collected information about published papers in this field for the past 50 years that have been indexed by Google Scholar [20] (Fig. 1), which main topic was found to be Feature Selection.

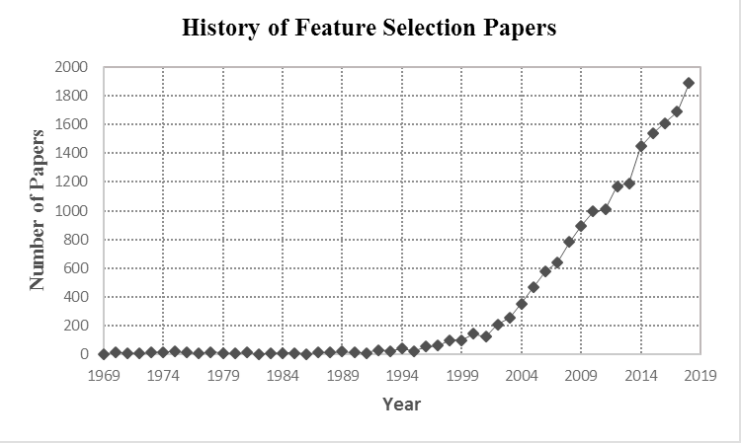

Fig. 1. History of papers indexed by Google Scholar with Feature Selection as main topic.

Fig. 1 only gives an estimation of this growing interest in Feature Selection field, but it is a reasonable evidence that notable advances can be expected in this field for the next years. For that reason, general systems may not adhere to any of those methodologies in specific, but instead, give support for their concurrence in the same application or facilitate the migration from one technology to another across different platforms.

2.3. Annotation Representation

High-quality labeled training datasets for supervised and semi-supervised machine learning algorithms are usually difficult and expensive to produce because of the large amount of time needed to label the data [21]. Many advances, such as data modeling, multidimensional indexing, and query evaluation, have been made along this research direction [22]. However, there are three major problems, especially when the size of image collections is very large. One is the enormous amount of labor required for manual image annotation [23]. Another difficulty, which is more essential, results from the rich content in the images and the subjectivity of human perception. That is, for the same image content different people may perceive it differently, especially if annotating images for different purposes and applications. The perception subjectivity and annotation impreciseness may cause unrecoverable mismatches in later retrieval processes and duplication of image content [24]. Finally, the last one is the lack of structured and standardized annotation that can be represented together with the image content (or easily coupled), giving support for the comparison and integration of different annotations into one single structure when a duplicated image content is found.

2.4. Set of Experience and Decisional DNA

The Set of Experience Knowledge Structure (SOEKS) is a knowledge representation designed to acquire and store formal decision events as experiences in a structured form. It is based on four key basic elements: variables, functions, constraints, and rules. Variables are in general used to represent knowledge in an attribute-value form, following the traditional approach for knowledge representation. Given that, the set of Functions, Constraints, and Rules of SOEKS are different ways of relating those variables. It is safe to say that variables are the central component of the entire knowledge structure. Functions define relations between a dependent variable and a set of input variables; thus, SOEKS uses functions as a way to create associations among variables and to build multi-objective goals. Likewise, constraints are functions that act as a way to limit the space of possibilities, or set of possible solutions and control the performance of the system in relation to its goals. Lastly, rules are relationships that operate in the universe of variables and express the condition-consequence connection as "if-then-else" and are used to represent inferences and associate actions with the conditions under which they should be implemented [8].

The Decisional DNA is a structure which has the capability of capturing decisional fingerprints of an individual or organization and has the SOEKS as its basis. Multiple Sets of Experience can be gathered, classified, organized and then grouped into decisional chromosomes, which accumulate decisional strategies for a specific area. The set of chromosomes comprise, finally, what is called the Decisional DNA (DDNA) of the organization [9].

3. Knowledge representation for visual content using SOEKS and DDNA

Choosing the most appropriate and effective formalization of knowledge for a given system is not trivial, and the consideration of using SOEKS can be founded in the fact that experience must considered if the objective is to mimic the human intelligence capabilities [2]. Therefore, SOEKS has been extended to the visual domain and used as carrier for decision making in Cognitive Cyber Physical Systems.

To preserve the visual content and also accelerate the retrieval process, in the proposed representation we include a minimally processed image pixel field *cvalue*, which contains the image pixel data $\boldsymbol{I_c}$, and is the starting value of the visual variable before being optimized:

$$cvalue \ni \boldsymbol{I_c} \tag{1}$$

To ensure efficiency in terms of search and indexing, and at the same time preserving generality, our representation

also has the optimized image *evalue* (which includes the set of selected features $\boldsymbol{I_e}$). The feature selection methodology and specifications is included in the header of the SOE for searching and indexing purposes:

$$evalue \ni \boldsymbol{I_e} \tag{2}$$

In this representation, images and the annotations (or detections), which are labeled bounding boxes coordinates of the objects or regions of interest (ROIs), are represented together. By coupling images and annotations, the sharing and reuse of these experiences may be facilitated, avoiding duplication of samples as well as promoting generation of new knowledge by the fusion of different annotations in one single image.

The framework representing the capture of visual content (from datasets or when the system is running in real time), its update for the generation of the optimized *evalue* as well as retrieval (in the presented case for user's feedback purposes) is represented in Fig. 3.

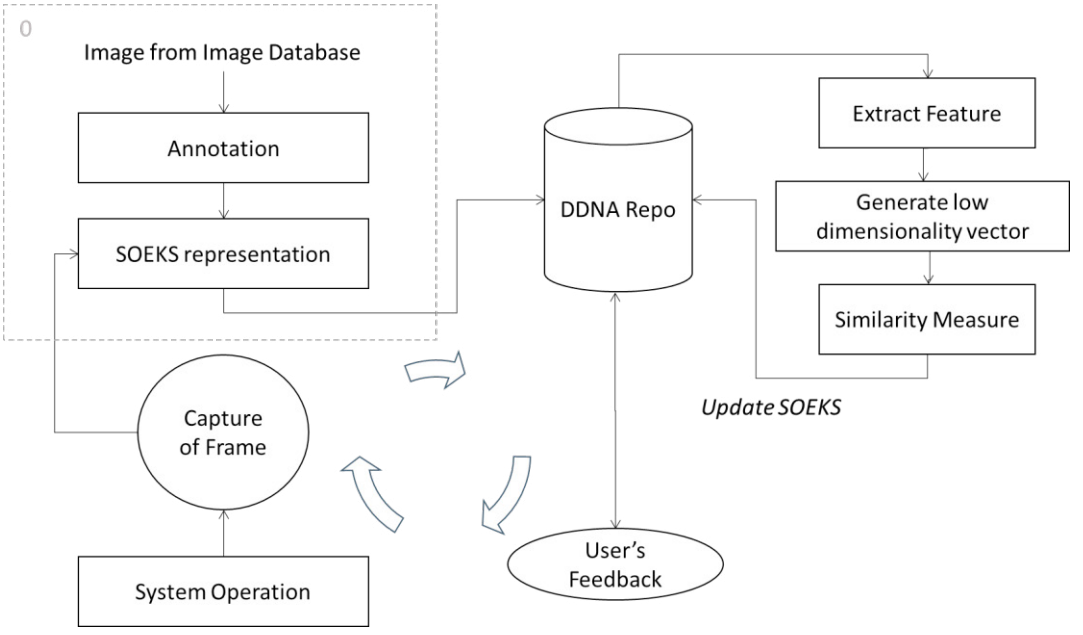

Fig. 3. Process of creation and update of experiences.

The explicitly represented knowledge in a SOEKS format may be used for three main purposes in general C-CPS. Firstly to promote code and knowledge reuse. Also, when facing a situation that has happened in the past the system may recommend a decision based on past experiences. In addition, in case of unexpected or unknown events where the given solution has not been found by user's feedback as an optimal solution, this new generated experience may be incorporated in the DDNA repository and used to retrain the system, thus improving the decision making process.

4. CVP-HC: Preciseness, uncertainty and creation of new knowledge

For the purpose of demonstrating the SOEKS representation for images, we embed this structure in a Cognitive Vision System (which could also be extended to Cognitive-CPS), enabling the representation of visual content together with contents from other domains (textual, sensorial etc), and giving support to its reuse. The knowledge retrieval is facilitated by taking into account the preciseness and uncertainty of each generated experience. These metrics are part of the SOEKS and used for calculation of similarity between new experiences gathered by the system and existing ones. By calculating similarity, we avoid replication of samples, improve the intelligence capabilities of the system by learning actively [25], and build a unique DDNA for each company's application as the system runs.

In addition, for the visual information in specific, the content of manual annotation is subject to human perception and application focus of interest. That is, for the same image content some people may perceive objects and regions that might have been unnoticed by others, especially if those entities are not relevant for system application in analysis. In this case, new knowledge may be generated by the fusion of similar image content but different annotated regions or objects.

The implementation of such representation and an example of how to define the preciseness and uncertainty of experiences composed by images and annotations has been tested in a Cognitive Vision Platform for Hazard Control (CVP-HC) that identify workers' exposure to risks in industrial environments, in special by the non-use of personal protective equipment. For this tests that have been performed, a set of 2500 images containing or not replicated samples and similar or not annotated regions have been represented as a SOEKS. For dimensionality reduction, PCA has been applied and a simple Euclidian distance [26] used to calculate the similarity between samples by a given precision. Results has shown gain in computational efficiency of search when a dimensionality reduction is applied, which justify the importance of having such selected features in the representation. On the other hand, time is the retrieving process is saved by keeping the original pixel values (once the transformation of the selected feature vector into pixel values are necessary for visualization purposes). Finally the capacity of the system to generate new and rich visual knowledge is achieved by including the annotation together with visual content in the same structure.

4.1. Dimensionality reduction search efficiency

PCA as any dimensionality reduction method is used to estimate how many components from the original pixel vector representation are needed to describe the giving data [16]. This can be determined by looking at the cumulative explained variance ratio as a function of the number of components of the dataset, as presented in Fig 4.

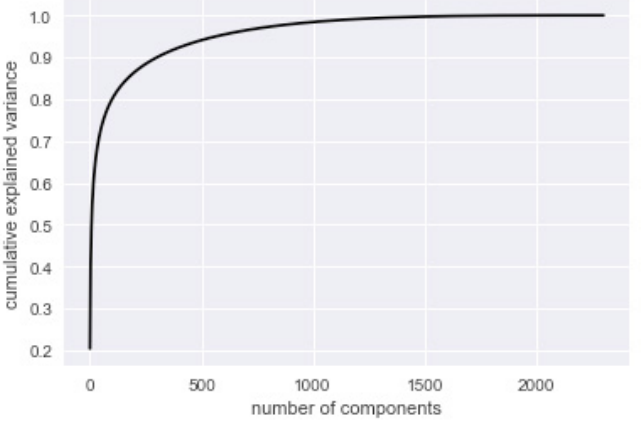

Fig. 4: Cumulative explained variance for the first N components of the sample dataset using PCA.

The values obtained quantifies how much of the total of a small 64 x 36-dimensional image is contained within the first N components. For example, we see that the first ~500 components contain approximately 95% of the variance (and a reduction of 80% of the initial image vector size), while over 2,000 components would be necessary to describe close to 100% of the variance (Fig. 4).This value gives a good estimate for uncertainty of that reduced representation, once it is an estimation of how much information from the initial raw data has been kept to describe that image. The chosen value of uncertainty (for instance 95%) is stored as part of the experience and will be used to define how many components will be kept in the optimized value for next collected experiences.

Using a simple Euclidian distance, we calculated computational costs in terms of time (Fig. 5-a) in relation to the reduced number of components as well as in relation to the cumulative explained variance (Fig. 5-b).

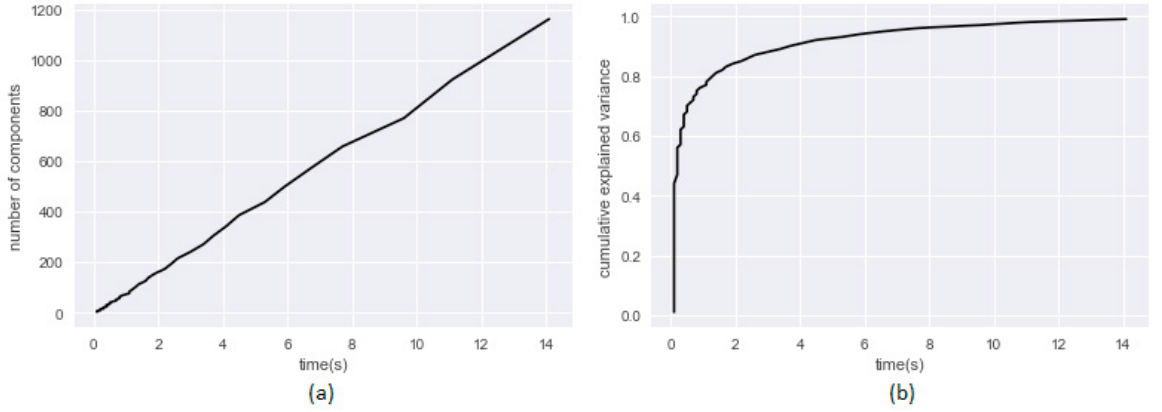

Fig. 5: Time processing in terms of number of components (a) and cumulative explained variance (b).

We can observe from Fig. 5 that, for a reduced vector composed by the 500 most relevant components (95% cumulative explained variance) when compared to the original vector size as results has shown a reduction of 30% of computational time.

4.2. Annotation and creation of new knowledge

In addition to the uncertainty for the calculation of the visual content similarity, preciseness is also a metric used to define to which extent an experience can be consider similar or equal to another. Similar experiences can be pruned to improve the intelligence of system by learning actively [25]. Moreover, by the pre-definition of preciseness, annotations can be joined into one single structure when the image inputs are considered highly similar or equal; therefore, resulting in the creation of new knowledge (Fig. 6).

Fig 6. Generation of new knowledge by the combination of annotations from highly similar image inputs.

5. Conclusions and Future work

The paper presents a knowledge representation for visual content in Cognitive Cyber Physical Systems (C-CPS). For the purpose of demonstrating the SOEKS representation for images, we embed this structure in a Cognitive Vision System, enabling the representation of visual content together with contents from other domains (textual, sensorial etc), and giving support to its reuse. The knowledge retrieval is facilitated by taking into account the preciseness and uncertainty of each generated experience. These metrics are part of the SOEKS and used for calculation of similarity between new experiences gathered by the system and existing ones. By calculating similarity, we avoid replication of samples, improve the intelligence capabilities of the system by learning actively, and build a unique DDNA for each company's application as the system runs (without compromising the platform generality).

In addition, for the visual information in specific, the content of manual annotation is subject to human perception and application focus of interest. In this case, new knowledge may be generated by the fusion of similar image content but different annotated regions or objects. Conflicting regions of interest (different annotations labels for same areas or objects) still has to be solved manually by the user.

The representation for images discussed in this paper will be evaluated for other sensorial data. Suitability of reusing experiences will also be explored for different case scenarios and the automation of decision making for conflicting annotations labels and regions will be further assessed.

References

1. Tang, K., Shi, R., Shi, H., Bhuiyan, M. Z. A., & Luo, E. (2018). Secure beamforming for cognitive cyber-physical systems based on cognitive radio with wireless energy harvesting. Ad Hoc Networks, 81, 174-182.
2. 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.
3. Brézillon, P., & Pomerol, J. C.: Contextual knowledge and proceduralized context. In Pro-ceedings of the AAAI-99 Workshop on Modeling Context in AI Applications, Orlando, Florida, USA, July. AAAI Technical Report (1999).
4. Vernon, D.: The space of cognitive vision. In Cognitive Vision Systems, 7-24, Springer, Berlin, Heidelberg (2006).
5. Zambrano, A., Toro, C., Nieto, M., Sotaquirá, R., Sanín, C., Szczerbicki, E.: Video semantic analysis framework based on run-time production rules – towards cognitive vision. J. Univ. Comput. Sci. 21(6), 856–870 (2015).
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. Sanin, C., E. Szczerbicki, E.: Set of experience: A knowledge structure for formal decision events. Foundations of Control and Management Sciences, 3, 95–113 (2005).
9. Sanin, C., Szczerbicki, E.: Decisional DNA and the smart knowledge management system: A process of transforming information into knowledge. In Techniques and tool for the design and implementation of enterprise information systems, ed. A. Gunasekaran, 149–175. New York: IGI Global (2008).
10. Rui Y, Huang TS, Chang SF (1999) Image retrieval: current techniques, promising directions, and open issues. J Vis Commun Image Rep 10(1):9–62
11. Jasmine, K.P., & Kumar, P.R. (2014). Localized Rgb Color Histogram Feature Descriptor for Image Retrieval.
12. Hsu, C. C., & Lin, C. W. (2018). Cnn-based joint clustering and representation learning with feature drift compensation for large-scale image data. IEEE Transactions on Multimedia, 20(2), 421-429.
13. Taatjes, D. J., Bouffard, N. A., Barrow, T., Devitt, K. A., Gardner, J. A., & Braet, F. (2019). Quantitative pixel intensity-and color-based image analysis on minimally compressed files: implications for whole-slide imaging. Histochemistry and cell biology, 1-11.
14. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.
15. Chen, B., Ho, D. W., Hu, G., & Yu, L. (2018). Distributed Dimensionality Reduction Fusion Estimation with Communication Delays in Cyber-Physical Systems. arXiv preprint arXiv:1802.03122.
16. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.
17. Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing, 18, 1-8.
18. De Jay, N., Papillon-Cavanagh, S., Olsen, C., El-Hachem, N., Bontempi, G., & Haibe-Kains, B. (2013). mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics, 29(18), 2365-2368.
19. Garrett, D., Peterson, D. A., Anderson, C. W., & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on neural systems and rehabilitation engineering, 11(2), 141-144.
20. Harzing, A. W. K., & Van der Wal, R. (2008). Google Scholar as a new source for citation analysis. Ethics in science and environmental politics, 8(1), 61-73.

21. Kothari, R., & Jain, V. (2002). Learning from labeled and unlabeled data. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 3, pp. 2803-2808). IEEE.
22. Jasmine, K. P., & Kumar, P. R. (2014). Localized Rgb color histogram feature descriptor for image retrieval. International Journal of Advances in Engineering & Technology, 7(3), 887.
23. Chun, Y. D., Seo, S. Y., & Kim, N. C. (2003). Image retrieval using BDIP and BVLC moments. IEEE transactions on circuits and systems for video technology, 13(9), 951-957.
24. Kherdikar, S. K., & Kulkarni, R. (2014). A novel approach for auto classification and grouping similar user query for image search. Internafional Journal of Computer Science & Informafion Technologies, 5(4), 4911-4915.
25. Carroll, J. M., & Mack, R. L. (1985). Metaphor, computing systems, and active learning. International journal of man-machine studies, 22(1), 39-57.
26. Wang, L., Zhang, Y., & Feng, J. (2005). On the Euclidean distance of images. IEEE transactions on pattern analysis and machine intelligence, 27(8), 1334-1339.