[94] A. Iosifidis, E. Marami, A. Tefas, I. Pitas, K. Lyroudia, The MOBISERV-AIIA eating and drinking multi-view database for vision-based assisted living, J. Inf. Hiding Multimedia Signal Process 6 (2011) 254–273.
[95] N. Kumar, P. Belhumeur, A. Biswas, D. Jacobs, W. Kress, I. Lopez, J. Soares, Leafsnap: A computer vision system for automatic plant species identification, in: European Conference on Computer Vision, 2012.
[96] J. Krause, M. Stark, J. Deng, L. Fei-Fei, 3d object representations for fine-grained categorization, in: International Conference on Computer Vision, 2013.
[97] T. Berg, J. Liu, S. Lee, M. Alexander, D. Jacobs, P. Belhumeur, Birdsnap: Largescale fine-grained visual categorization of birds, in: Computer Vision and Pattern Recognition, 2014.
[98] S. Wu, A. Oelemans, E. Bakker, M. Lew, A comprehensive evaluation of local detectors and descriptors, Signal Process. Image Commun. 59 (2017) 150–167.
[99] W. Li, H. Pan, L. Pengyuan, X. Xie, Z. Zhang, Medical image retrieval method based on texture block coding tree, Signal Process. Image Commun. 59 (2017) 131–139.
[100] X. Jiang, M. Simon, Y. Yang, J. Denzler, Multi-marker tracking for large-scale X-ray stereo video data, Signal Process. Image Commun. 59 (2017) 140–149.
[101] B. Yang, X. Shang, S. Pang, Isometrick hasing for image retrieval, Signal Process. Image Commun. 59 (2017) 117–130.
[102] Z. Xia, X. Feng, J. Lin, A. Hadid, Deep convolutional hashing using pairwise multilabel supervision for large-scale visual search, Signal Process. Image Commun. 59 (2017) 109–116.
[103] Q. Wei, B. Sun, J. He, L. Yu, Bnu-lsved 2.0: Spontaneous multimodal student affect database with multi-dimensional labels, Signal Process. Image Commun. 59 (2017) 168–181.

tion and description, in: Computer Vision and Pattern Recognition, 2015.
[72] Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., Toderici, G., 2015. Beyond short snippets: Deep networks for video classification, in: Computer Vision and Pattern Recognition.
[73] N. Srivastava, E. Mansimov, R. Salakhutdinov, Unsupervised learning of video representations using lstms, in: International Conference on Machine Learning, 2015.
[74] Z. Li, E. Gavves, M. Jain, C. Snoek, Videolstm convolves, attends and flows for action recognition, 2016, arXiv:1607.01794.
[75] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, D. Parikh, Making the V in VQA matter: Elevating the role of image understanding in visual question answering, in: Conference on Computer Vision and Pattern Recognition, 2017.
[76] D. Agrawal, S. Das, A. El Abbadi, Big data and cloud computing: Current state and future opportunities, in: Proceedings of the 14th International Conference on Extending Database Technology, ACM, New York, NY, USA, 2011, pp. 530–533.
[77] N. Tsapanos, A. Tefas, N. Nikolaidis, I. Pitas, A distributed framework for trimmed kernel k-means clustering, Pattern Recognit. 48 (2015) 2685–2698.
[78] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM 51 (2008) 107–113.
[79] M. Zaharia, M. Chowdhury, M. Franklin, S. Shenker, I. Stoica, Spark: cluster computing with working sets, in: 2nd USENIX conference on Hot topics in cloud computing, 2010, pp. 10–10.
[80] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 594–611.
[81] M. Everingham, L. Van Gool, C. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.
[82] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Zitnick, P. Dollar, Microsoft COCO: Common objects in context, 2015, arXiv:1405.0312.
[83] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Zitnick, D. Parikh, VQA: Visual question answering, in: International Conference on Computer Vision, 2015.
[84] N. Mostafazadeh, I. Misra, J. Devlin, M. Mitchell, X. He, L. Vanderwende, Generating natural questions about an image, in: 54th Annual Meeting of the Association for Computational Linguistics, 2016.
[85] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Techn. Report, University of Massachusetts, Amherst, 2007.
[86] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 faces in-the-wild challenge: The first facial landmark localization challenge, in: International Conference Computer Vision Workshops, 2013.
[87] N. Erdogmus, S. Marcel, Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect, in:Biometrics: Theory, Applications and Systems, 2013.
[88] H. Ng, S. Winkler, A data-driven approach to cleaning large face datasets, in: IEEE International Conference on Image Processing, 2014.
[89] J. Xu, D. Vazquez, A. Lopez, J. Marin, D. Ponsa, Learning a part-based pedestrian detector in virtual world, IEEE Trans. Intell. Transp. Syst. 15 (2014) 2121–2131.
[90] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, Hmdb: A large video database for human motion recognition, in: International Conference on Computer Vision, 2011.
[91] H. Kim, A. Hilton, Influence of colour and feature geometry on multi-modal 3d point clouds data registration, in: International Conference on 3D Vision, 2014.
[92] F. Heilbron, V. Escorcia, B. Ghanem, J. Niebles, Activitynet: A large-scale video benchmark for human activity understanding, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[93] Y. Jiang, Z. Wu, J. Wang, X. Xue, S. Chang, Exploiting feature and class relationships in video categorization with regularized deep neural networks, 2015, arXiv preprint arXiv:1502.07209.

[44] A. Iosifidis, A. Tefas, I. Pitas, Approximate kernel extreme learning machine for large scale data classification, Neurocomputing 219 (2017) 210–220.
[45] W. Schmidt, M. Kraaijveld, R. Duin, Feedforward neural networks with random weights, in: Advances on Neural Information Processing Systems, 1992.
[46] Y. Pao, G. Park, D. Sobajic, Learning and generalization characteristics of the random vector functionl link net, Neurocomputing 6 (1994) 163–180.
[47] A. Rahimi, B. Recht, Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning, in: Advances on Neural Information Processing Systems, 2008.
[48] G. Huang, Q. Zhu, C. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[49] A. Iosifidis, Extreme learning machine based supervised subspace learning, Neurocomputing 167 (2015) 158–164.
[50] A. Iosifidis, A. Tefas, I. Pitas, M. Gabbouj, A review of approximate methods for kernel-based big media data analysis, in: European Signal Processing Conference, 2016.
[51] A. Cochocki, R. Unbehauen, Neural Networks for Optimization and Signal Processing, first ed., John Wiley & Sons, Inc., New York, NY, USA, 1993.
[52] G. Zhang, Neural networks for classification: a survey, IEEE Trans. Syst. Man Cybern. C 30 (2000) 541–462.
[53] G. Hingon, S. Osindero, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006).
[54] S. Haykin, Neural Networks and Learning Machines, third ed., Prentice Hall, 2008.
[55] S. Kiranyaz, T. Ince, A. Iosifidis, M. Gabbouj, Progressive operational perceptrons, Neurocomputing 224 (2017) 142–154.
[56] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.
[57] A. Ng, J., D., Building high-level features using large scale unsupervised learning, 2012, arXiv:1112.6209.
[58] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1915–1929.
[59] N. Passalis, I. Tefas, Neural bag-of-features learning, Pattern Recognit. 64 (2017) 277–294.
[60] G. Cao, A. Iosifidis, K. Chen, M. Gabbouj, Generalized multi-view embedding for visual recognition and cross-modal retrieval, IEEE Trans. Cybern. (2017). http://dx.doi.org/10.1109/TCYB.2017.2742705.
[61] A. Iosifidis, A. Tefas, I. Pitas, View-invariant action recognition based on artificial neural networks, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 412–424.
[62] P. Nousi, A. Tefas, Deep learning algorithms for discriminant autoencoding, Neurocomputing 226 (2017) 325–335.
[63] G. Cao, A. Iosifidis, M. Gabbouj, Neural class-specific regression for face verification, IET Biometrics (2017). http://dx.doi.org/10.1049/iet-bmt.2017.0081.
[64] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[65] J. Dong, Q. Chen, S. Yan, A. Yulle, Towards unified object detection and semantic segmentation, in: European Conference on Computer Vision, 2014.
[66] R. Girshick, 2015. Fast r-cnn, arXiv:1504.08083.
[67] S. Ren, K. He, R. Girshick, J. Sun, 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, arXiv:1506.01497.
[68] J. Redmon, A. Farhadi, Yolo9000: Better, faster, stronger, 2016, arXiv:1612.08242.
[69] M. Waris, A. Iosifidis, M. Gabbouj, Cnn-based edge filtering for object proposals, Neurocomputing 266 (2017) 631–640.
[70] B. Fernando, E. Gavves, J. Mogrovejo, A. Ghodrati, T. Tuytelaars, Rank pooling for action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 773–787.
[71] J. Donahue, L. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recogni-

[19] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, 3d object recognition in cluttered scenes with local surface features: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 2270–2287.
[20] C. Aytekin, H. Possegger, T. Mauthner, S. Kiranyaz, H. Bischof, M. Gabbouj, Spatiotemporal saliency estimation by spectral foreground detection, IEEE Trans. Multimedia (2017). http://dx.doi.org/10.1109/TMM.2017.2713982.
[21] N. Passalis, A. Tefas, Learning neural bag-of-features for large-scale image retrieval, IEEE Trans. Syst. Man Cybern. 99 (2017) 1–12.
[22] J. Li, N. Allinson, A comprehensive review of current local features for computer vision, Neurocomputing 71 (2008) 1771–1787.
[23] I. Laptev, On space–time interest points, Int. J. Comput. Vis. 64 (2005) 107–123.
[24] A. Yilmaz, O. Javed, M. Shah, Object trackiing: a survey, ACM Comput. Surv. 38 (2006) 1–45.
[25] M. Enzweiler, D. Gavrila, Monocular pedestrian detection: Survey and experiments, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 2179–2195.
[26] A. Iosifidis, A. Tefas, I. Pitas, Activity-based person identification using fuzzy representation and discriminant learning, IEEE Trans. Inf. Forensics Secur. 7 (2012) 530–542.
[27] P. Turaga, R. Chellappa, V. Subrahmanian, O. Udrea, Machine recognition of human activities: A survey, IEEE Trans. Circuits Syst. Video Technol. 18 (2008) 1473–1488.
[28] A. Iosifidis, A. Tefas, I. Pitas, Multi-view human action recognition: A survey, in: International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2013.
[29] J. Blat, A. Evans, H. Kim, E. Imre, L. Polok, V. Ila, N. Nikolaidis, P. Zemčík, A. Tefas, P. Smrž, A. Hilton, I. Pitas, Big data analysis for media production, Proc. IEEE 104 (2016) 2085–2113. http://dx.doi.org/10.1109/JPROC.2015.2496111.
[30] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, 1998.
[31] B. Scholkopf, A. Smola, Learning with Kernels, MIT Press, 2001.
[32] S. Kiranyaz, M. Gabbouj, Dynamic and scalable audio classification by collective network of binary classifiers framework: An evolutionary approach, Neural Netw. 34 (2012) 80–95.
[33] A. Iosifidis, A. Tefas, I. Pitas, Dynamic action recognition based on dynemes and extreme learning machine, Pattern Recognit. Lett. 34 (2013) 1890–1898.
[34] A. Iosifidis, A. Tefas, I. Pitas, Active classification for human actions, in: IEEE International Conference on Image Processing, 2013.
[35] C. Williams, M. Seeger, Using the nyström method to speed up kernel machines, in: Advances on Neural Information Processing Systems, 2001.
[36] P. Drineas, R. Kannan, M. Mahoney, Fast monte carlo algorithms for matrices ii: Computing a low-rank approximation to a matrix, SIAM J. Comput. 36 (2011) 158–183.
[37] A. Iosifidis, M. Gabbouj, Nyström-based approximate kernel subspace learning, Pattern Recognit. 57 (2013) 190–197.
[38] Y. Lee, Y. Huang, Reduced support vector machines: A statistical theor, IEEE Trans. Neural Netw. 18 (2007) 1–13.
[39] K. Zhang, J. Kwok, B. Parvin, Prototype vector machine for large scale semisupervised learning, in: International Conference on Machine Learning, 2009.
[40] A. Iosifidis, M. Gabbouj, Class-specific nonlinear projections using class-specific kernel spaces, in: IEEE International Conference on Big Data Science and Engineering, 2015.
[41] V. Mygdalis, A. Iosifidis, A. Tefas, I. Pitas, Large-scale classification by an approximate least squares one-class support vector machine ensemble, in: IEEE International Conference on Big Data Science and Engineering, 2016.
[42] A. Iosifidis, M. Gabbouj, On the kernel extreme learning machine speedup, Pattern Recognit. Lett. 68 (2015) 205–210.
[43] A. Iosifidis, M. Gabbouj, Scaling up class-specific kernel discriminant analysis for large-scale face verification, IEEE Trans. Inf. Forensics Secur. 11 (2016) 2453–2465.

terms of repeatability and performance on large-scale image search. Medical image analysis is the focus of [99]
and [100]. In [99], an efficient image retrieval model based on iterative texture block coding tree is proposed. [100] proposes an approach for efficiently addressing the problem of manual annotation of large-scale medical video collections, and specifically focuses on X-ray stereo video data. Authors propose an efficient multi-marker tracking method that is able to highly accelerate the annotation process. Efficient distance calculation in Big Media Data based on hashing is the focus of [101] and [102]. Specifically, [101] proses a new hashing technique that can efficiently minimize distances in both the input and binary (hashing) space, leading to better hashing description. [102] proposes a hashing method that can be directly applied on raw image data and, thus, optimize both image description and representation for hash-based distance calculation. Finally, a new large-scale data set for spontaneous and multimodal affect analysis is introduced in [103].

References

[1] R. Pearson, M. Gabbouj, Nonlinear Digital Filtering with Python: An Introduction, first ed., CRC Press, 2015.
[2] R. Maini, H. Aggarwai, Study and comparison of various image edge detection techniques, Int. J. Image Process. 3 (2009) 1–11.
[3] B. McCane, K. Novins, D. Crannitch, B. Galvin, On benchmarking optical flow, Comput. Vis. Image Underst. 84 (2001) 126–143.
[4] A. Saxena, J. Schulte, A. Ng, Depth estimation using monocular and stereo cues, in: International Joint Conference on Artificial Intelligence, 2007.
[5] G. Finlayson, E. Trezzi, Shades of gray and color constancy, in: Color Image Conference, 2004.
[6] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.
[7] Y. Boykov, M. Jolly, Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images, in: International Conference on Computer Vision, 2001.
[8] C. Aytekin, S. Kiranyaz, A. Iosifidis, M. Gabbouj, Recent advances in salient object detection: Towards object recognition in big media data, Futura 35 (2016) 80–92.
[9] E. Hjelmas, B. Low, Face detection: A survey, Comput. Vis. Image Underst. 83 (2001) 236–274.
[10] S. Zafeiriou, C. Zhang, Z. Zhang, A survey on face detection in the wild: Past, present and future, Comput. Vis. Image Underst. 138 (2015) 1–24.
[11] D. Triantafyllidou, P. Nousi, A. Tefas, Fast deep convolutional face detection in the wild exploiting hard sample mining, Big Data Res. (2017).
[12] C. Aytekin, A. Iosifidis, M. Gabbouj, Probabilistic saliency estimation, Pattern Recognit. 74 (2018) 359–372.
[13] C. Aytekin, A. Iosifidis, S. Kiranyaz, M. Gabbouj, Learning graph affinities for spectral graph-based salient object detection, Pattern Recognit. 64 (2017) 159–167.
[14] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Comput. Surv. 35 (2003) 399–458.
[15] X. Tan, S. Chen, Z. Zhou, F. Zhang, Face recognition from a single image per person: A survey, Pattern Recognit. 39 (2006) 1725–1745.
[16] Z. Zeng, M. Pantic, G. Roisman, T. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 39–58.
[17] C. Shan, S. Gong, P. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image Vis. Comput. 27 (2009) 803–816.
[18] F. Patrona, A. Iosifidis, A. Tefas, N. Nikolaidis, I. Pitas, Visual voice activity detection in the wildy, IEEE Trans. Multimedia 18 (2016) 967–977.

frameworks in order to handle large-scale datasets. For example, a distributed approach that can work with any serial clustering algorithm entails using the serial algorithm on data subsets, then merging the clusters [77].
The MapReduce distributed programming model [78] was invented by Google, specifically for handling extremely big datasets. It was inspired by the corresponding map and reduce primitives offered by functional programming languages, such as Lisp. In general, it consists of two steps. The first one, is applying a function on all the elements separately (Map) and the second one is collecting the results, using a commutative and associative operation (Reduce). An advantage of this model, over using a standard Message Passing Interface system, is that the programmer does not have to handle the low-level details of an implementation, e.g., data distribution to the worker nodes, faulttolerance, or load balancing, as tools for writing high-level programs for clusters do exist. While it may not provide a suitable solution for every possible problem, the MapReduce model particularly lends itself to problems that involve running simple operations on a large number of elements.
The Apache Spark cluster computing framework [79] builds upon Hadoop, in order to improve computation speed and is also compatible with HDFS. Its advantages over Hadoop include its ability to create and operate on more complex Directed Acyclic Graph (DAG) scheduling for tasks than Hadoop's two-stage MapReduce DAG scheduling and the ability to cache data in the distributed memory.
Nowadays, the big impact on Big Data analytics has been caused by the extensive use of Graphics Processing Units (GPU) for parallel computing. Indeed, the GPU processing allowed for training very deep neural networks for solving various learning tasks with very largescale datasets even in one workstation that has multiple GPUs available.
Several programming frameworks have been released in order to make the training and deployment of such deep learning models easy (e.g. TensorFlow, Caffe, Theano, Torch, etc.).

5. Data collections

While the size of data being available everyday becomes enormously big, their practical value for applying machine learning models is limited. This is due to the fact that most of this data is released without annotation and even unstructured. For this reason, the collection of large and annotated data sets is of significant importance. This is due to the importance of exploiting domain knowledge during the model selection and training process. Existing datasets target the problems of generic object and scene analysis [80–82], visual question generation and answering [83,84], facial image analysis [85–88], person detection [89], human action recognition [90–94], as well as datasets targeting applications involving media data analysis in other scientific fields [95–97].

6. The special issue

In this context, the current Special Issue on Big Media Data Analysis includes works on generic image description, medical image and video analysis, distance calculation acceleration and data collection. Specifically, an analysis of local image description is provided in [98]. The authors provide an experimental analysis of many (standard and more recent) local descriptors in

However, in the case of Big Media Data Analysis, the application of standard kernel-based learning is difficult. This is due to the high space and time complexities of standard kernel based learning approaches, which typically are quadratic and cubic functions of the cardinality of the training data. Thus, for Media Data Analysis problems involving hundreds of thousands (or even millions) of samples, the use of standard kernel-based learning is impractical. However, they can be applied in the context of Big Data by exploiting Divide and Conquer strategies [32–34]. According this, the Big Data problem is divided, in some optimal manner, to smaller sub-problems where the application of standard kernel-based learning can be efficiently applied.

In order to make the application of kernel-based learning approaches in large problems possible, approximate learning schemes have been proposed. The main idea behind these schemes is to keep as much information as possible, while considerably reducing the size of the model.

In order to do so, three approaches have been proposed, i.e. based on low-rank approximation of the corresponding kernel matrix [35–37], based on a reduced kernel space definition [38–44] and based on a randomized kernel space definition [45–49]. A review of these three approximate kernel-based learning approaches can be found in [50].

Other types of models, widely adopted in Big Media Data Analysis problems, are those based on iterative optimization. The most important such models are those exploiting neural network topologies [51–55].

Neural network models have received enormous attention during the last years due to their ability to be applied on raw (image/video) data and learn data representations of increased level of abstraction, a paradigm usually refereed to as Representation Learning [56–58].

Instead of adopting human-made data representations, Representation Learning defines the optimal (according to a given criterion) representation based on training data, thus, achieving state-of-the-art performance in numerous research problems, including object detection and recognition [59], image and text retrieval [60] and face and action recognition [61–63].

Two neural network topologies that have been widely used in Big Media Data Analysis problems are Convolutional Neural Networks (CNNs) [56] and Recurrent Neural Networks (RNNs) (usually implemented by using the Long-Short Term Memory network architecture [64]). CNNs, taking as input raw image/video data and optimizing both the data representation and classification tasks in a combined way, have been shown to achieve excellent performance in many Media Data Analysis problems, including object detection/recognition and scene analysis [65–69] and activity recognition [70]. RNNs are more suitable in modeling data in which the time dimension contains significant information, like image and video analysis [71–73], action recognition [74] and visual question answering [75].

4. Big Data management and analytics

Distributed computing can provide the means to handle problems on very large datasets that would otherwise be almost impossible to solve [76]. It provides virtually limitless memory and processing power.

Provided that a task can be split into many independent subtasks, then it can theoretically be performed in a reasonable amount of time, regardless of the data size, given enough processing units. Thus, many machine learning and pattern recognition algorithms can be implemented in distributed computing

ing and Pattern Recognition, and Big Data Management and Analytics. Since it involves processes belonging to all these research topics, their distinction within the concept of Big Media Data Analysis E-mail addresses: alexandros.iosifidis@eng.au.dk (A. Iosifidis), tefas@aiia.csd.auth.gr (A. Tefas), pitas@aiia.csd.auth.gr (I. Pitas), moncef.gabbouj@tut.fi (M. Gabbouj).
is unclear. In the following, we provide a comprehensive discussion on selected topics of these fields, followed by the introduction of the works included in this Special Issue.

2. Image/video analysis and computer vision

The amount of images and videos available every day is growing at an exponential rate. In order to be successfully used for analysis, such images and videos need to be pre-processed for noise removal and enhancement [1]. Depending on the task under consideration, several image processing steps need to be applied, including the calculation of edges [2] and optical flow [3], depth and/or disparity estimation (in the cases where depth sensors or stereo cameras are available) [4] and color normalization [5].
In order to efficiently process large collections of images, data need to be reduced by focusing on a part of it which is more important for the given task. Image segmentation is used in order to split the image in parts according to their similarity [6,7]. At a higher level, semantic image segmentation identifies locations in an image that are important according to the task semantics [8]. In this context, face detection identifies image locations depicting human faces [9–11], while salient object segmentation identifies image locations that are plausible to the human eye [12,13]. Given such semantic image locations, higher level tasks like face recognition [14,15], facial expression and action recognition [16–18] and object recognition [19] can be applied. These concepts have been also extended in the analysis of videos, by applying a spatio-temporal analysis [20]. Large scael image and video retrieval [21] is another task closely related to big media analysis since the datasets used on this task are usually huge.
Image/video locations description, e.g. by exploiting local descriptors on local neighborhoods of Interest Points [22,23], is another processing step towards the application of high level image analysis and computer vision tasks, like object tracking [24] and human behavior analysis. Human behavior analysis includes the tasks of human detection [25], identification [26] and the recognition human actions [27,28]. It has been heavily researched during the last two decades due to its importance in many application scenarios involving Big Media Data, like video surveillance, security, human–computer interaction and entertainment [29].

3. Machine learning and pattern recognition

Big Media Data Analysis inevitably involves the understanding of visual content. For many Media Data Analysis problems, it has been shown that the use of linear models leads to inferior performance, compared to nonlinear ones. This is why, during the last decades when large annotated data sets were not available, research in these problems was primarily focused on the application of nonlinear models, like kernel-based learning [30,31].
Kernel-based learning approaches are still now widely adopted in many small- and medium-scale classification problems due to their excellent performance, theoretical foundation and easy implementation.

Signal Processing: Image Communication

Alexandros Iosifidis ▪
Department of Engineering, Electrical & Computer Engineering, Aarhus University, Denmark
Department of Informatics, Aristotle University of Thessaloniki, Greece
Anastasios Tefas ▪
Department of Informatics, Aristotle University of Thessaloniki, Greece
Ioannis Pitas ▪
Department of Informatics, Aristotle University of Thessaloniki, Greece
Moncef Gabbouj ▪
Laboratory of Signal Processing, Tampere University of Technology, Finland

abstract
In this editorial a short introduction to the special issue on Big Media Data Analysis is given. The scope of this Editorial is to briefly present methodologies, tasks and applications of big media data analysis and to introduce the papers of the special issue. The special issue includes six papers that span various media analysis application areas like generic image description, medical image and video analysis, distance calculation acceleration and data collection.
Keywords
Big Media Data, Data analytics, Machine learning, Statistical learning, Deep learning

1. Introduction

Recent advances in consumer electronics, such as digital cameras, smartphones and depth sensors, as well as the daily use of social media and image/video sharing platforms have led to an explosion of digital media data, i.e. images and videos, captured every day. The analysis of such large sets of data (usually referred to as Big Data) has the potential to reveal patterns, trends and rules that would have been impossible to observe with smaller datasets used only few years ago. This potential of Big Data has led to increased interest from both the scientific community and industry. On the one hand, processing and analyzing such large collections of data generate new challenges that need to be appropriately addressed while, on the other hand, successful handling and analysis of Big Data leads to better prototypes and products.

Two properties of Big Media Data make the application of standard state-of-the-art pattern recognition methods prohibitive, are its cardinality and dimensionality. Moreover, most of such data is released without annotation and even unstructured making the application of standard data processing approaches in this context is prohibitive. In this context, there is an increasing interest in devising new methodologies that are able to efficiently process and analyze large collections of data, as well as in collecting large and annotated data sets that can be used in order to adapt generic Machine Learning methodologies using domain knowledge and train models to be used in real applications.

Big Media Data Analysis is a multi-disciplinary research field, including topics of classical Image and Video Analysis, Computer Vision, Machine Learn-