International Journal of Engineering & Technology, 7 (2.33) (2018) 195-199

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



State of the art in image processing & big data analytics: issues and challenges

S. Vahini Ezhilraman [1] *, Sujatha Srinivasan [2]

[1] Research Scholar,School of Computing Sciences, VISTAS, Pallavaram, Chennai
[2] Asst. Prof & Head, Dept. of Information Technology,School of Computing Sciences, VISTAS,Pallavaram, Chennai
**Corresponding author E-mail: vahiniezhil19@gmail.com*

Abstract

Image processing, in the contemporary domain, is now emerging as a novel and an innovative space in computing research and applications. Today, the discipline of "computer science" may be termed as "image science", why because in every aspect of computer application, either science or humanities or management, image processing plays a vital role in varied ways. It is broadly now used in all the industries, organizations, administrative divisions; various social organizations, economic/business institutions, healthcare, defense and so on. Image processing takes images as input and image processing techniques are used to process the images and the output is modified images, video, or collection of text, or features of the images. The resultant output by most image processing techniques creates a huge amount of data which is categorized as Big-data. In this technique, bulky information is processed and stored as either structured or unstructured data as a result of processing images through computing techniques. In turn, Big Data analytics for mining knowledge from data created through image processing techniques has a huge potential in sectors like education, government organizations, healthcare institutions, manufacturing units, finance and banking, centers of retail business. This paper focuses on highlighting the recent innovations made in the field of image processing and Big Data analytics. The integration and interaction of the two broad fields of image processing and Big Data have great potential in various areas. Research challenges identified in the integration and interaction of these two broad fields are discussed and some possible research directions are suggested.

Keywords: *Image Processing; Bigdata Analytics; Data Mining; Feature Extraction.*

1. Introduction

The technology of Image processing encompasses by highly utilizing the computer proficiency to analyze the digital images i.e. the images generated using a computer. Image processing is used in numerous ways in many of the important technological-related fields like Oceanography, currency recognition, Medical imaging, remote image transmission, fake-note deduction, Satellite imaging etc. The Digitized image is analyzed and manipulated to improve the image's eminence. Separation of images at present is a most domineering phase in image processing which is popularly called as 'Image Segmentation'. The process of Image segmentation involves separating the required objects from an existing image. Thus the process of image segmentation is desirable to analyse some specific objects from an image instead of examining an intact image. In the present scenario, Image segmentation plays a vital role in deducting the fake-notes, recognizing the original currencies etc. Interestingly it is worth noting that many of the image processing techniques were involved in separating the distinct color levels of the images and making them into two-dimensional signals. Image processing is also used for processing those distinct color levels into three-dimensional signals with the third-dimension. The resultant signals or features extracted by processing the images are analyzed to incur some knowledge and take decisions.

Big Data mainly encompasses 3Vs viz., volume, velocity, and variety. All the images are basically processed by huge amount bites such as Petabytes (1,024 terabytes) or Exabyte's. Usually, Petabytes consists of billions of trillions records. Big Data analytics comprises of applying data mining techniques to Big Data. Processing huge amounts of data require large computing power both software and hardware.

Different kinds of techniques are currently used in the fields of image processing and Big Data analytics. However, the interaction and integration of the two broad fields need more insight to better explore and exploit the benefits of the two techniques. Thus, the present study aims to focus of the recent research progresses in the two broad fields of image processing and Big Data analytics to bring out the importance of their interaction and integration. The Second Section of this paper is designed as it gives an outline on image processing techniques and big-data analytics. The Third Section reviews the state of the art in the two broad fields. Also, recent developments in the integration of the two fields are discussed. The Discussion part dealt with the research gaps and challenges identified from the review of literature and suggest some future research directions for further exploration. Finally, the last Section is the conclusion of this paper.

2. An overview of image processing and big data analysis techniques

Image processing involves applying computing techniques to processing images. Images are taken as input and the properties of the images are changed to enhance the image or features are extracted to make them less complex to study. The following section gives an overview of different image processing techniques.

Copyright © 2018 S. Vahini Ezhilraman, Sujatha Srinivasan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.1. An overview of image processing techniques

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.
There are many techniques involved in the image processing system, of which very few are popularly in use.

2.1.1. Visualization

Visualization is to communicate a message through images, diagrams, or animation. One of the visualization techniques is visual imagery. There are two methods of image visualization. The first method is abstract visualization and the second technique is model-based scientific visualization. In abstract visualization, 2D or 3D techniques are used. Whereas, another method using digitally constructed images (real object) is directly obtained from source of data like medical images.

2.1.2. Image restoration

Image restoration is to clear noise and recover the resolution loss. Otherwise, we say that image restoration recovers the original image from a given degraded image. Software used image restoration Adobe Photoshop, CS3 Extended, GIMP, Paint and NET.

2.1.3. Retrieval of images

The process of retrieving the image in a huge database system is referred as Image retrieval. Different searching techniques involved in this area such as Content-based Image Retrieval, Content-based Visual Information Retrieval, Document-based Image Retrieval, Human-oriented Image Retrieval and so on.

2.1.4. Image recognition

Image recognition is nothing but recognizes the image object. For example: Consider the collection of image as an input and the output is to recognize the object. Some of the open source software is used to recognize the image.

2.1.5. Image enhancement

Image enhancement is different from the process of restoration of images. Image Enhancement's principal aim is to give an excellent image with high quality. The more enhancement concepts are sometimes called as image editing. E.g., filtering, contrast adjustment, de-blurring, morphological etc.

2.1.6. Image pre-processing

The techniques of pre-processing of Images are generally categorized into two basic categories such as

i) Static Thresholding and.
ii) Dynamic Thresholding. As like the Image Enhancement, the method of Image pre-processing largely involved in the process of enhancing the images by modifying the pixel values either its brightness or its contrast for its visual impact which may be occurred due to blurriness which capturing from low conventional/digital cameras or from the images obtained from the satellite pictures.

Apart from the above techniques of image processing, various new and advanced systems of image processing using Big Data tools are also in use. An organization can be handling their multitasking challenges, build their enterprise data hub, content management, managing their open source data, etc. by using those Big Data tools.

2.2. Usage of big data

For handling a large number of datasets, Big Data is one of the effective fields in the existing data handling system. This field includes data, images, and videos. For the subsequent utilization of Big Data to a greater extent, one can prefer to use Big Data analytics. Commonly Analytics is implemented in both structured and unstructured data. The mandatory things to implement these data are hardware, software and algorithms should be applied for analyzing the original data. In fact, nowadays it is more helpful for the business peoples, Healthcare institutions and Military organizations to analyze the hidden pattern as well as to know the unknown correlation. The big sizes of images used in Big Data and data analytics are together called as Image data analytics.

2.3. Importance of this study

This study aimed to survey the existing scholarly, as well as innovative works done on the field of Big Data with special importance, is given to the Image processing with the help of using Big Data. Sources for this survey comprised of the scholarly articles published in electronic media (internet), peer-reviewed journals of national and international, various kinds of published sources such as proceedings of the seminar volumes, etc. For to know such recent innovations using Image Processing in the study of Big Data, the Research articles published from the years between 2008 and 2016 are surveyed for a better understanding of its progress in the study area.

3. A brief survey on image and data analytics

In the latest development on Big Data and Image processing, there are many innovative up comings were contributed by different authors. The reveals that there is much scope left for the future researchers to find more recommendations for the development of the study.

3.1. State of the art in image processing

Authors like Wang et al. (2008) [1] projected an algorithm for contour extraction and de-noising. According to them, those algorithms should be used for feature extraction, smoothing of images, reconstruction of images, enhancing the quality of images, etc. Apart from that, canny edge detection, median filtering, contour tracing and wavelet transform are combined by making use of those algorithms. Their experiments reveal that the algorithm is capable of processing images captured with poor quality containing spatial redundancy and high noise.
Du et al. (2009) [2] proposed a registration of medical image with a mixed-type approach under the standards of DICOM (Digital Imaging and Communications in Medicine). The method of Mutual Information (MI) is used as the similarity measure for image matching. This method provides the performers with high accuracy, speed and with less computational complexity.
Elakkia and Narendran (2016) [3] broadly discuss the usage of image processing. According to them, Image segmentation here refers to divide up the image into a number of regions, and each of which is homogeneous in some sense like its intensity, texture, color, etc. In the present day development, image segmentation has also an important role to play in the field of medical imaging such as diagnosing of cancer, detection of tumor at early stage etc.
In Digital Image Matting, the pre-processing method is an important step which wishes to enhance more pixels from an unidentified region of the input of Trimap (three-region mask). Yao (2017) [4] proposed that, even though this step has no relation with the well-known matting equation, but it compares colour differences between those known pixels and unknown pixels.
Thus the above-discussed surveys are basically concentrated in the field of image processing and its allied fields. Along with this, for

the comparative basis, the works carried out in the field of data analytics is also surveyed briefly for the betterment of this study.

3.2. State of the art in data analytics

A generic system of a framework to decompose Big Data platforms were classified into four components, i.e., data analysis, data acquisition, data storage, and data generation by Hu et al. (2014) [5]. The survey was conducted in their research for every stage and they made efforts for its further development. For architectural design, engineering insights were also provided in this research. The tools like Hadoop were used to find solutions for specific analysis in the Bigdata platforms along with standard results.

Acharjya and Ahmed (2016) [6] proposed a system for Big Data which is necessary for describing the upheaval of data. They felt that it may be a strong motivation for the scholars of the future generation. However, the basic questions aroused among the scholars who are dealing with Big Data are to describe the vital features in a quantitative way. The authors explained about Fourth 'V' in their work which refers to veracity comprising accountability and convenience.

Big Data utilized in different ways like data, problem, methodology, and technology were proposed by Hong et al. (2016) [7]. Few of the ways that they proposed in relation to Big Data are:

i) At least from anyone of the following i.e. Variety, Volume or Velocity, the data analysis may be involved in a big way. The utilization of Big Data includes the industries related to weather data analysis, smart beat data, social media data and phasor quantity unit data.
ii) The problem of Big Data is also analyzed in the areas of data compression and data security.
iii) For individual system modules, a customised modeling method is more required for the Big Data. However, all its individual components need a systematic understanding.
iv) Forgetting quicker and speeded response, Big Data can be widely used almost in all image processing techniques. Besides, for an existing descriptive model, scholars like Himmi et al. (2017) [8] added the value of fifth "V", in which the Value 'V' has defined in two meanings. Literally, the basic meaning stands for the data that creating value for the companies. On the other hand, the depth of the utilization of data can be replicated through a company's value. In such a way, many of the prevailing works had tried to find the value of the first meaning. Therefore, to find the second meaning value, the upcoming scholars are encouraged to explore the data in their future research.

The above discussed are some of the scholarly works done and published by the experts in the field of data analytics. However, this study shows some eagerness to survey on the works done using the Big Data analytics in the field of image processing and its related areas. Nowadays most of the industries are largely showing their interest in using the Big Data Analysis. Most particularly many of the Government agencies prepared their plans for their developing research fields using the Big Data analytics. Thus, the concept of Big Data helps us to understand the abstraction and in-depth understanding of various hide values. The impact of Big Data on the present day organizations and almost any of the existing models its outcome can be impregnable.

3.3. State of the art in big data analytics from image data – a review

There are some expert attempts to find the utilization of Big Data analytics in the field of image processing. However, very few of its kind were traced out through various sources and they are gathered and studied carefully. Some of the scholarly works which are related to this field are surveyed and given as follows.

Image mining as proposed by Sudhir (2011)[9] is a dynamic field of research to explore further information from the image. One of the major phases in image mining technology is Image segmentation. In the field of image processing, Image mining is a part which dealt with the mining of hidden data. It is also associated with the image data and its supplementary configurations that are accrued indistinctly in those images. Image mining is often a complicated process because it involves incoming of various related techniques including data mining, image retrieval, indexing schemes, pattern recognition, etc... In addition, the process of image mining is expected to be an excellently developed and qualitative system for providing the image storage area to the users through a convenient access and also it should distinguish the system of data configurations and makes information lesser than the representation of the given image. In this system, one may attempt to bring together the functions like feature extraction, storage of images, processing of images, image indexing, pattern knowledge, and discovery, and image retrieval.

Mohan and Porwal (2015) [10] suggested that the image analysis methods using Multispectral are mainly based on the availability of spatial aspects of the particular objects of interest shape, texture and spectral properties with few wavelength bands. By using the multispectral techniques one may acquire the required images. But, it is challenging for using some of the data to distinguish between vegetation categories, differentiation of lakes, identifying the combinations of rocks and soil.

In the Fuzzy classification method, image data are classified with highly complex relations for the creation of feature vector from images. Kannvdiya et al. (2016) [11] have broadly discussed the features of images that need to be considered for image mining. Here the ambiguous nature of the members of the feature vector is combined with different types of relations. Thus, the classification of a relation based image is highly liable on the image type format and on the threshold selection.

Goutam et al. (2016) [12] attempted to study Big Data and its technologies in the field of map reduce and its role. According to them, the function of Mapper deserves an important role in creating key and value. In this both Key and values combine with the help of referring various kinds of data viz., structured data, unstructured data and semi-structured data. Regarding unstructured data like image, the images may get erroneous because of noise or any other related features. For augmenting the quality of a tainted image, the mapper function could accomplish acceptably and precisely in creating the data related to key and value pair. This program again used as a reducer function for the auxiliary dispensation of major prominence. In this research, the author discusses one of the techniques of the histogram for equalization which is available in image improvement. This technique should be practical in combination with map reduce function for its effectual and precise process.

In the present context, as suggested by Pandian and Balasubramanian (2017) [13] the field of digital image gathering have been amplified due to the fast development. In this development, Images are generated in gigabytes levels which are used most particularly generated by using the apparatus of both civilian and military. Further, it is difficult for someone who accesses or willing to utilize data if it is organized as well as for allowing the well-organized, surfing, examining and recovering. The present research areas in the field of image processing are retrieving the images. In this process, computer vision and database management are considered as the two main field of research. Of which, the field of image retrieval is entirely text-based and visual based. Moreover, image retrieval based on content has been recommended to overwhelm those problems. Similarly, there exists the digital image searching difficulty in an enormous databank from the content- based image retrieval program. Accordingly, some instant images, as well as the text-based keywords are manually annotated with the help of visual content viz. shape, color, and texture.

The image processing software using MATLAB to perform detection, classification and recognition were reviewed by Paulchamy et al. (2017) [14]. It is suggested that this existing method involves Region of Interest (ROI) and pixel classification technique which requires a large amount of database. Though this system involves only road sign recognition, it does not intimate the driver about the

road sign detection. The author suggests about using MATLAB platform with the help of the camera. Here the detection and classification modules are performed by image processing software MATLAB. It is used here for traffic sign recognition purpose and also it uses huge data to recognize and classify the traffic sign. By using an ultrasonic sensor, one can intimate the driver about the obstacles on roads that detected by the sensor. In addition to this, here a motor is connected with Raspberry-pi. When an interrupt occurs in Raspberry-pi, then the speed of the motor will be reduced. Among many ways to control the speed, one of the easiest ways is pulse width modulation (PWM) technique. The authors also suggested about the E-speak and Monitor through which one can detect the output which is indicated through E-speak (audio output) and monitor on Raspberry-pi. By using this techniques certain road signs such as speed break, school zone speed of the vehicle will be reduced. Based on the above discussions on the study of image processing and data analytics, information is gathered and tabulated for the easy understanding of the above concept.

Table 1: Summary of the Survey on Big Data Analytics for Features Extracted From Image Processing

Refer-ence	Algorithm/ Techniques	Dataset	Results / Observation	Proposed Future Enhancements
Wang et al. (2008)	Median filtering and Wavelet transform	Real images	Achieves good effect	A better technique for de-noising image to be developed
Du et al. (2009)	Segmenta-tion and Vocal-based method	Medical image registra-tion	Low computation complexity and high speed	Optimized algorithm should be developed
Elakkia and Narendran (2016)	Noise is modelled as Additive White Gaussian Noise	Noisy image	Noise is reduced	Better performance should be found.
Yao (2017)	Sampling -based matting algorithm	Massive semi-transparent pixels	Large size trimap is achieved.	Accurate performance of image matting should be developed.
Hu et al. (2014)	Apache Hadoop tool is used	Grid Mix and Pig Mix develop-ment	Built- in the Testing framework are developed	Precocious Hadoop tools should be developed.
Paul-chamy et al. (2017)	Road sign recognition using MATLAB without indicator and speed control techniques.	Detecting traffic signal obstacles	Signal, automated tool is created and Speed is controlled	Recognizing the content of each traffic sign should be developed.
Acha-rjya and Ahmed (2016)	Apache Hadoop & Map reduce	Semi structured or unstructured data.	Used in fault tolerant storage, high output and data processing.	A better machine learning noise reduction to be developed.
Hong et al. (2016)	Discussing smart grid applications	Used for customer analytics	May be used in the industrial field.	Data compression and security issues are to be developed.
Himmi et al. (2017)	Restructure sales business model	Value added activities	Leverage Profit Revenue Optimi-zation for airlines are sufficiently restructured.	More operations should be introduced
Sudhir (2011)	Indexing methods, Image storages, retrieval and mining system.	Image	Image mining used for image storages and retrieval	Medical image of anatomical structure to be developed
Pandian and Balsubra-manian (2017)	Curvelet transform, Contourlet Transform algorithm and Local ternary pattern algorithm	3-D Biological data analysis, seismic data analysis	Contourlet transform succeeds.	The retrieval time should be improved.
Mohan and Porwal (2015)	Imaging spectroscopy	Hyperspectral image	Dimensionality reduction method achieved	Better spatial resolution and spectral resolution
Kannvdiya et al. (2016)	Discuss K nearest algorithm, support vector machine.	Algorithmic development	K nearest algorithm executed.	Algorithm to be developed
Goutam et al. (2016)	Histogram equalization technique	Generate key, value pair of map function	Mapper function can perform adequately	Efficient and accurate Algorithm to be developed

3.4. Discussion

Table 1 depicts the literature survey which is made to analyze the study of image processing and data analytics. It is found that the available sources in relation to image processing and data analytics are hardly done by the earlier scholars. As discussed above, Elakkia and Narendran (2016), Yao (2017) studied about various image processing techniques for image matting and image segmentation, remove the Gaussian Noise, control the automatic traffic signal and to maintain the healthcare systems. Himmi et al. (2017), Paulchamy et al. (2017), Hu et al. (2014) discussed highly using of Big Data Analytics and its tools. Most of the authors are supported the Hadoop and MapReduce tools. These tools helped the business people consider as well as to create automated tools for changing business models. Authors like Sudhir (2011), Kannvdiya et al. (2016) discussed image mining techniques. These two areas are most helpful for present-day computer world. However, most of the researchers are not attempted to analyse about these data analytics with the combination of image processing. Therefore there is a lag in this field that leads to a research gap are combining image processing techniques and data analytics.

This study reveals that some of the following gaps are found in the above-discussed survey and those gaps should be filled with by identified or finding the suitable solutions with the help of the existing experts of this field for the future prospects. So far in this survey, the above mentioned scholars have to find.

i) A new method for the noise reduction adopting new technics for the images for giving better performance.
ii) More accurate and good performance of image matting should be developed.
iii) In future precocious Hadoop tools should be developed. Even though adaptive and accurate mechanisms have been applied to extract and recognize the content of traffic sign on the above studies, it should be further developed for its betterment.
iv) Comparative analyses with various parameters of performances are more essential to understand the most efficient techniques. Therefore, further research for mitigating the various issues related to that domain should be conducted.
v) Besides, the medical image of the anatomical structure cannot be used so far in the medical field. Therefore some possible future studies will be made to find the better computational time.
vi) Apart from the theoretical part of the Big Data, the researchers should also focus on the algorithm to develop the characteristics to reduce time and space complexity.
vii) A better machine learning concept should be developed and implemented for reducing noisy, imbalance data and missing values in the forthcoming days.

4. Conclusion

The above survey reveals that most of the authors discussed various tools and techniques of two broad areas i.e. image processing and data analytics. This survey comprised of three broad sections in which the state of the art in image processing was dealt in the

first section, and the state of the art in Big Data analytics was dealt in the second section and finally in the third section, the combination of the two broad fields of image processing and data analytics from the features or data extracted from these images into a single efficient system for better applications and decision making was discussed extensively.

Thus, the future work may combine the image and Big Data to develop an advanced hybrid system. In future, research may be done in the fields related to detecting satellite images, medical or health care problems like brain tumor, cancer or for detecting some other diseases. We may also this application in future for identifying the unidentifiable diseases in vegetables, fruits which may help the farmers and agriculturalists. However, so far no attempt has been made to analyze the features extracted from the image for gaining better knowledge from the image data in the review so far. There are other more applications of Big Data for the betterment of humankind. Hence Big Data analytics is critical for mankind.

Acknowledgement

I would like to render my gratitude to my research supervisor Dr. Sujatha Srinivasan, HOD, Dept. of IT, VISTAS, for her effective and patient full guidance to complete this paper. I am also thankful to the VISTAS for giving me opportunity to carry my research. I further extend my thanks to friends and well-wishers who encouraged me both directly and indirectly to complete this work.

References

[1] Y. Wang, J. Zheng, H. Zhou, L. Shen, "Medical Image Processing by Denoising and Contour Extraction," International Conference on IEEE, 2008.

[2] Y. Du, A. Song, L. Zhu, W. Zhang, "A Mixed-Type Registration Approach in Medical Image Processing," In: International Conference on Biomedical Engineering and Informatics, E, pp 1–4, 2009.

[3] K. Elakkia and P. Narendran, "Survey of Medical Image Segmentation Using Removal of Gaussian Noise in Medical Image," vol. 6, no. 6, pp.7593–7595, 2016.

[4] G. Yao, "A Survey on Pre-Processing in Image Matting," vol. 32, no. 15, pp. 122–138, 2017.

[5] H. A. N. Hu, Y. Wen, S. Member, and T. Chua, "Toward Scalable Systems for Big Data Analytics : A Technology Tutorial," vol. 2, 2014.

[6] D. P. Acharjya, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools," vol. 7, no. 2, 2016.

[7] T. Hong, C.Chen, J. Huang, N. Lu, L. Xie, H. Zareipour, "Big Data Analytics for Grid Modernization", IEEE Trans Smart Grid vol.7, pp. 2395–2396, doi: 10.1109/TSG.2016.2593358, 2016.

[8] K. Himmi, J. Arcondara, P. Guan, and W. Zhou, "Value Oriented Big Data Strategy: Analysis & Case Study," pp. 1053–1062, 2017.

[9] R. Sudhir, "A Survey on Image Mining Techniques: Theory and Applications," vol. 2, no. 6, pp. 44–53, 2011.

[10] B. Krishna Mohan and A. Porwal, "Hyperspectral image processing and analysis," Curr. Sci., vol. 108, no. 5, pp. 833–841, 2015.

[11] M. Kannvdiya, K. Patidar, R.S. Kushwaha, "A Survey on: Different Techniques and Features of Data Classification," Int J Res Comput Appl Robot vol.4, pp.1–6, 2016.

[12] D. Goutam, S. Gupta, S. Deepti, "Big Data Analytics: Image Enhancement Based Approach," Int J Adv Res Comput Sci Softw Eng vol. 5, pp. 570–573, 2016.

[13] A. Anbarasa Pandian and R. Balasubramanian, "Performance analysis of texture image retrieval in curvelet, contourlet, and local ternary pattern using DNN and ELM classifiers for MRI brain tumor images," Adv. Intell. Syst. Comput., vol. 459 AISC, no. July, pp. 239–248, 2017.

[14] B. P. K. Archana, A. D. Bruno, and I. D. M. Gopala, "A Novel Approach Towards Road Safety Based On Investigational Survey Using Image Processing And User Interface System," pp. 105–108, 2017.