



**Data Analytics and Artificial Intelligence**  
**Vol: 1(3), 2021**  
**REST Publisher, ISBN: 978-81-948459-4-2**  
**Website: <http://restpublisher.com/book-series/daai/>**

# **Exploring the Frontiers of Computer Vision: A Journey through Image Processing, Object Recognition, and Machine Learning**

**P. Jayanthi**

St. Joseph's College of Arts and Science for Women, Hosur, Tamil Nadu, India.

Corresponding Author Email: [jayanthi79p@gmail.com](mailto:jayanthi79p@gmail.com)

## **Abstract**

Computer vision has been studied by many perspectives. It expands from raw data recording into techniques and ideas combining digital image processing, pattern recognition, machine learning and computer graphics. The extensive usage has attracted many scholars to integrate with many disciplines and fields. This paper provides a survey of the recent technologies and theoretical concept explaining the development of computer vision mainly related to image processing using different areas of their field application. Computer vision helps scholars to analyse images and video to obtain necessary information, understand information on events or descriptions, and beautiful pattern. It used the method of multi-range application domain with massive data analysis. This paper contributes to recent development on reviews related to computer vision, image processing, and their related studies. We categorized the computer vision mainstream into four group, e.g., image processing, object recognition, and machine learning. We also provide brief explanation on the up-to-date information about the techniques and their performance.

**Keywords** – Computer vision, Image processing, Digital image

## **1. Introduction**

Computer vision has been expanded into the vast area of field ranging from recording raw data into the extraction of image pattern and information interpretation [1]. It has a combination of concepts, techniques, and ideas from digital image processing, pattern recognition, artificial intelligence and computer graphics [2]. Most of the tasks in computer vision are related to the process of obtaining information on events or descriptions, from input scenes (digital images) and feature extraction. The methods used to solve problems in computer vision depend on the application domain and the nature of the data being analyzed. Computer vision is a combination of image processing and pattern cognition [2],[3]. The output of the Computer Vision process is image understanding. Development of this field is done by adapting the ability of human vision in taking information. Computer Vision is the discipline of extracting information from images, as opposed to Computer Graphics [4]. The development of computer vision depends on the computer technology system, whether about image quality improvement or image recognition. There is an overlap with Image Processing on basic techniques, and some authors use both terms interchangeably [4],[5]. The primary purpose of Computer Vision is to create models and data extracts and information from images, while Image Processing is about implementing computational transformations for images, such as sharpening, contrast, among others [4]. It also has similar meaning and sometimes overlapping with In Human and Computer Interaction (HCI)[6]. HCI coverage focus on full design, interface and all aspects of technologies related to the interaction between human and computer. HCI is then developed as a separate discipline (which is the field of interdisciplinary science) which discusses the interrelationships between human-computer mediated by technology development including human aspects. Functionally, computer vision and human vision are the same [7], with the aim of interpreting spatial data, i.e., data indexed by more than one dimension. However, computer vision cannot be expected to replicate just like the human eye [8]. This is due to computer vision system has limited performance and function compared to the human eye. Even though many scholars have proposed broad area of computer vision techniques to replicate human eye, however, in many cases, there is any limitations of the performance of computer vision system [9]. One of the significant challenges in their technique is the sensitivity of the parameters, the strength of the algorithm, and the accuracy of the results. It impact on the complexity of performance evaluation of computer vision systems. Generally, the performance evaluation involves measuring some of the basic behaviors of an algorithm to achieve accuracy, strength, or extensibility to control and monitor system performance. As the performance of computer vision system depends

on the application system design, there is comprehensive effort proposed by many scholars to expand and categorized computer vision into many areas and specific applications such as automation on the assembly line, remote sensing[11], robotics[10], computer and human communications[12], tools for the visually impaired, and others.

## 2. Literature Review

Computer vision works by using an algorithm and optical sensors to stimulate human visualization to automatically extract valuable information from an object[13]. Compared to conventional methods that take a long time and require sophisticated laboratory analysis, computer vision has been expanded into a branch of artificial intelligence (artificial intelligence) and simulated human visualization. It also combined with lighting systems to facilitate image acquisition continued with image analysis. In more detail, the stages of image analysis are: 1) image formation, in which image of object is captured and stored in computer; 2) image preprocessing, whereby quality of image is improved to enhance the image detail; 3) image segmentation, in which the object image is identified and separated from the background, 4) image measurement, where several significant features are quantized, and 5) image interpretation, where the extracted images are then interpreted[14]. The recent development of image processing technology has provided the possibility to create a system to recognize a digital image. The construction is summarized in table 1.

**TABLE 1.** Development of research related to computer vision and digital image processing

No	Researchers	Usefulness	Definition
1	Savioja et al., [15], Fadeyev, V., & Haber, C. [16]	image processing applied in the digital sound system	Audio information stored in the flowing threads in the media such as phonograph recording can be reconstructed, with no minimal contact, by measuring the shape of the groove using precision metrology and digital image processing methods.
2	Various, N. A., & Sun, D. W. [17] (2012) Langton et al. [18], Li et al. [20], Jackman et al. [19], Du and Sun [21], Jusoh et al. [22], Pallottino et al. [23]	image processing used in food analysis	It has three main steps, e.g., feature selection and extraction, and classification. Feature selection and retrieval, and classification. This approach will be reviewed from aspects of role portrayal and its impact on the food and beverage industry.
3	Canny [24]; Vacchetti et al., [25] Wang [26], Zheng et al., [27]	Convolutional Neural Networks (CNN) is used for object detection.	Use of ANN in in-depth learning to solve edge detection problems in image processing areas.

Digital image processing or other image processing tend to evolve into the most extensive mainstream with support from other theoretical fields supported by the rapid development of specific disciplines such as mathematics, Linear algebra, statistics, Soft Computing, and Computational neurosciences the supported regulations supporting the development of digital image processing are given in Table 2 Pattern Recognition as a branch of computer vision focused on the process of object identification through image transformation to get a better image quality and image interpretation. This process aims to extract information to make decisions based on images obtained from sensors [5]. In other words, computer vision seeks to build an intelligent machine to "see." Common frameworks used in computer vision are image acquisition, pre-processing, feature extraction, detection/segmentation, high-level processing, and decision-making [5], [6]. The computer vision frameworks consisted two main groups, e.g., 3D morphological analysis and pixel optimization. The 3D morphological review has been a standard theory for computer image processing and pattern recognition, whereas pixel optimization is related to characterization of pixel morphology, including structural analysis and internal components for a better understanding of vector function [32]. Also, the approach should be performed on relatively large data sets covering many layers of geometrical composition. Therefore, efficient and accurate computing algorithms to extract the relevant quantitative information are important to understand the complex color clusters as a whole. The integration of morphological analysis with some artificial intelligence methods can result in better

performance through computing algorithms. The computing algorithm is fuzzy logic[33], artificial neural networks[34], and genetic algorithms[35]. They can be combined to completely complete complex tasks.

**TABLE 2.** Related research on computer vision

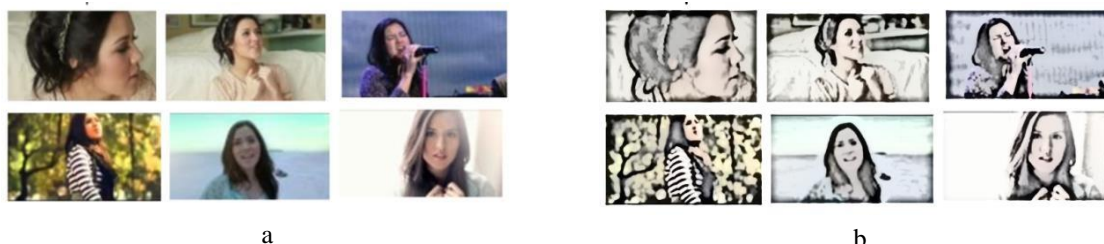
No	Name	Goals	Method	Results
1	Su, et al.,[28]	Estimation of solid depth map of a single monocular natural image	Bayesian methods to restore 3D scene structures	Superior performance after minimization of pixel-wise depth errors with training dataset but lower performance for objects taken from natural image.
1	Su, et al.,[28]	Estimation of solid depth map of a single monocular natural image	Bayesian methods to restore 3D scene structures	Superior performance after minimization of pixel-wise depth errors with training dataset but lower performance for objects taken from natural image.
2	Liu, et al., [29]	Comparison of image quality with different retargeting methods quickly and reliably	Reverse order (top-down)	Good consistency using certain metrics
3	Forder et al.,[30]	Presentation of a series of psychophysical experiments to determine the simultaneous dynamic sequence of human visual systems under the full adaptation of background lighting.	Psychophysical experiments in the dynamic display. Dynamic device design (HDR)	The performance of pixel difference in the range of 3.7log units under certain observational conditions.
4	Yhann et al.,[54]	Development of mentoring system that notifies the driver if the speed is insufficient according to the visibility conditions	Sensitivity technique based on the Contrast Function (CSF)	An approach to solve and validate algorithms from the psychophysical point of view is proposed. This algorithm can be used to develop sophisticated driver visibility descriptors.
5	Anbarjafari, et al., [31]	Image-based Illumination Inspired by Using Decomposition of Local Singular Value and Discrete Wavelet Transformation	Image illumination enhancement on color pixel correction	This technique is compared to the general standard histogram distribution (GHE) and local histogram equalization (LHE).

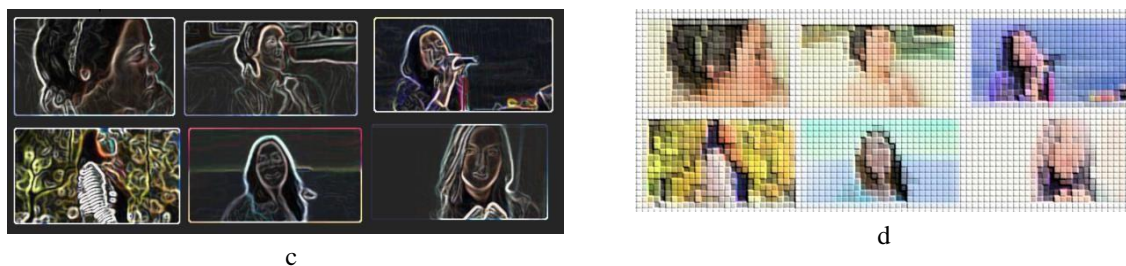
There are two approaches to the segmentation and retrieval of image data. Segmentation is basically to divide an image into areas that are not overlapping (overlapping) [36] through specific algorithms to estimate an area of the image. The city is a collection of pixels that have the same unique characteristics as color, gray level, texture, and others. The area is retrieval region of images, e.g., search engine parts, human search, and similar imagesearch. In general, the image segmentation approach that is often used is the method of intensity, color approach and shape approach [37]. In most computer vision applications, edge/border detection and image segmentation are a very important in the object recognition and interpretation. In image analysis literature, the performance of segmentation is shown in a minimal sample image. However, in large-scale image database annotations require parameter settings. Segmentation is obtained by gradient texture and feature space [38], [39], or by unsupervised clustering [40],[41],[42] or by texture classification [43]. Segmentation of labeling is important in localization performance and boundary localization [44]. It uses grouping and segmentation as an initial estimate of objects in the image by setting the threshold on the feature grouping algorithm especially in estimating the number of areas [45]. Fundamentally, segmentation has four main stages as below (Fig.1).Input image Segmented map before integration Edge map before integration Segmented map and edge map after combination Pixel clustering Segmentation has a primary goal to create resemblance map which derived from a prominent object detection model or hierarchical segmentation of the input image. The plan is an aggregation model tries to form a more accurate salience map. It needs components of pixel salience value  $x$  toward  $i$ -salience map cell location. In Borji et al.[46], it proposed a model of the standard saliency method of aggregation. The image is segmented into saliency score for  $n$ -total pixels and  $n$ -segments index which labeled as a prominent cluster. As the groups are an aggregation model [46], it adopted pixel-wise aggregation asset of model parameters. It has a weakness that such direct incorporation is ignorance of interaction between neighboring pixels.

Original images				
Segmented map before integration				
Edge map before integration				
Edge map with minimum noise				
Pixel clustering				

FIGURE 1. Segmentation stages

Therefore, CRF is proposed by Khan [47] to combine caliber maps of several methods and capture values of neighboring pixels. CRF aggregation model parameters are considered better to optimize training data since the reliability of each pixel has a higher probability of prominent when it is trained with CRF. Whereas data extraction requires photographed objects from camera, sensor, or satellite devices in the form of single images or image sequences. This extraction aims to separate background objects with foreground objects. It has three combinations, e.g., (a) the objectives are still the same color as the original, (b) the object changes color to black and white, or (c) the purpose becomes transparent.





**FIGURE 2.** extraction requires

Fig.2(a) object instance is changed to black and white; (b) objects have a different size than their original size after scale factor; (c) purposes being transparent or color combinations; (d) objects with BW color after scale arrangements and changing positions differently from their real opinions. Pixel is determinant of object sharpness in the image. Pixel optimization is useful for object detection, segmentation, and recognition. In boundary-based techniques, an edge detector is used to locate the boundary of an object. This method is based on the fact that the intensity of pixels will change rapidly on the perimeter of two regions. For color segmentation, the edge detection is performed on each RGB color channel. It results in edges which can be combined to get final edge image. In local-based techniques, pixels are grouped according to uniformity criteria. Examples of these methods are regional growth and separation techniques and split and merge. In regional growth techniques, pixels are arranged into larger areas based on predetermined criteria [48]. This method started from a set of core points and grew into more extensive regions by adding each neighboring pixel from the above starting points that have similar features, such as gray or color values. The technique of separation and merging the pixels begin by dividing an image into several regions and then combining the areas to meet predetermined criteria [48]. This region-based technique has two significant weaknesses [5]. Firstly, methods of separation and incorporation of regional growth depend primarily on predetermined global criteria at the outset. Secondly, the process of area growth depends on the initial segment and original pixels which impact the object detection performance. Object detection is useful for searching an object or display by querying data from recorded and real-time datasets.



**FIGURE 3.** Example of image retrievals using query image from big datasets of PASCAL MTH, MSD, SLAR, CDH, and RADAR with various RGB pixel values

Object detection often has error rate because the object or object does not include the pattern specified by the algorithm and usually has to be supplemented with an additional algorithm. The algorithms are usually used for detecting smaller parts to obtain more detailed images. For example, in face processing, the algorithm is used to identify the element of head and face which has a lower resolution [49],[50]. The algorithm also improves the machine performance to determine the eyes, eyebrows, and mouth [51],[52],[53]. Whereas, the final parts such as ears and neck are rarely studied in face processing. A bitmap is an image stored as a set of pixels associated with a computer screen. Bitmap images are often referred to as raster images. A bitmap image is an image formed from a pixel, with each pixel having a specific color. If the bitmap image is enlarged, for example, to be four times, then the vision will be blurred because the pixel also increases to 4 times which impact the image quality. Bitmap image formats are often used in photos and images. Two terms that need to be understood when working with bitmap images are resolution and color depth. Bitmap images are usually obtained by scanners, digital cameras, video capture, and others. As digital photos are susceptible to various types of noise. It needs a bitmap template. Bitmap templates are considered as standard images that have been readable by the computer. While the bitmap image is a raw image that has not been detected by a computer. It can contain an error in the image acquisition process that results in an unstructured pixel value that does not reflect the correct intensity of the actual scene. There are several ways that noise can be introduced into an image, depending on how the image is made. If a picture is scanned from a photo made in a movie, the film grain is the source of the noisy pixels. The noisy pixel can also be due to damage to the film or introduced by the scanner itself. If images are obtained directly in digital format, data collection mechanisms (such as CCD detectors) can cause noisy. Electronic data transmission of data also can

introduce noise effect. Research on image processing has been directed to meet machine learning and computing processes that can recognize patterns of increasingly diverse objects. Machine learning is closely related to computational statistics which consisted of spam filtering, optical character recognition, search engines and computer vision. There are extensive arrays of studies observing algorithms to reduce noise effect such as linear filtering of the Gaussian-based algorithm. The algorithms can eliminate certain types of grain noise from a photo. Because each pixel is set to average values in its environment, the local variations caused by the grain are reduced.

### 3. Conclusion

Computer vision has been related to image processing and machine learning. Computer vision as a field of a wide array of discipline has been linked closely to image processing discipline. The Image processing, itself has brought benefits in different areas of technology especially to analyze images to obtain the necessary information. As technological areas to be developed with computer vision, it has been expanded to other engineering fields such as geographical remote sensing, robotics, computer and human communication, healthcare, and satellite communication. Researchers who interested in computer vision can use the knowledge to predict individual events by analyzing images and videos and extracting their features. As the developments in the field of computer vision are related closely to image processing and machine learning, it can be used to more extensive areas of studies to predict or detect object behavior and characteristics including human activities and natural events. Figure: (a) the object instance is changed to black and white; (b) Examples of objects of different size than their original size because of a scale factor; (c) Examples of objects being transparent or color combinations; (d) Examples of objects being black and white with scale arrangements and also changing positions differently from their real opinions.

### References

1. Patel, Krishna Kumar, A. Kar, S. N. Jha, and M. A. Khan. "Machine vision system: a tool for quality inspection of food and agricultural products." *Journal of food science and technology* 49, no. 2 (2012): 123-141. doi: 10.1007/s13197-011-0321-4
2. Cosido, Oscar, Andres Iglesias, Akemi Galvez, Raffaele Catuogno, Massimiliano Campi, Leticia Terán, and Esteban Sainz. "Hybridization of Convergent Photogrammetry, Computer Vision, and Artificial Intelligence for Digital Documentation of Cultural Heritage-A Case Study: The Magdalena Palace." In *Cyberworlds (CW)*, 2014 International Conference on, pp. 369-376. IEEE, 2014. DOI: 10.1109/CW.2014.58