



A two-level computer vision-based information processing method for improving the performance of human–machine interaction-aided applications

Osama Alfarraj¹ · Amr Tolba^{1,2}

Received: 14 July 2020 / Accepted: 23 September 2020 / Published online: 8 October 2020
© The Author(s) 2020

Abstract

The computer vision (CV) paradigm is introduced to improve the computational and processing system efficiencies through visual inputs. These visual inputs are processed using sophisticated techniques for improving the reliability of human–machine interactions (HMIs). The processing of visual inputs requires multi-level data computations for achieving application-specific reliability. Therefore, in this paper, a two-level visual information processing (2LVIP) method is introduced to meet the reliability requirements of HMI applications. The 2LVIP method is used for handling both structured and unstructured data through classification learning to extract the maximum gain from the inputs. The introduced method identifies the gain-related features on its first level and optimizes the features to improve information gain. In the second level, the error is reduced through a regression process to stabilize the precision to meet the HMI application demands. The two levels are interoperable and fully connected to achieve better gain and precision through the reduction in information processing errors. The analysis results show that the proposed method achieves 9.42% higher information gain and a 6.51% smaller error under different classification instances compared with conventional methods.

Keywords Computer vision · Classification learning · Human–machine interaction · Information processing · Regression analysis

Introduction

Computer vision (CV) is an artificial intelligence-aided paradigm that is used to detect the digital world via cameras, videos, and deep-learning methods [1]. It classifies objects in an appropriate manner and responds to them automatically. CV focuses on the 3D modeling of objects, multiple-camera geometric analyses, cloud processing, and inferences based on the motion [2]. It acquires input from the machine and processes the output based on specific knowledge, namely, object labels and synchronization. It also includes developing particular technologies, such as image recognition, visual recognition, and facial recognition [3]. Mostly, CV is used to achieve high-level understanding of digital processing.

The task of CV is to capture, examine, and recognize digital objects and extracts in higher dimensions [4]. This task allows it to provide the scientific or representative information needed for the transformation of images used in geometry, physics, statistics, and learning theory [5].

Human–machine interactions (HMIs) involve communicating and cooperating with the machine through a user interface. These interactions are carried out between the user and the machine to control the machine's intuitive behaviors [6]. In recent years, specific sensors have been used to capture the normal, abnormal, and neural postures necessary for controlling the machine [7]. In HMIs, CV is used to acquire high-level photos and videos used in the understanding of these postures/images. CV is related to HMI in that they both detect and monitor objects as well as control them by determining the conditions under which they operate [8]. One of CV's upgraded applications is hand-gesture recognition, which monitors the postures used in HMIs and interacts with them. The goal of this application is to obtain vigorous non-specific vision [9]. It is used in automatic parking, controlling music with a gesture, eye tracking with multi-

✉ Osama Alfarraj
oalfarraj@ksu.edu.sa

¹ Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia

² Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin-El-Kom 32511, Egypt

ple touches, and so forth. CV and HMI are commonly used together in virtual reality applications. In comparison to other baseline technologies, CV combined with HMI produces the smallest errors [10].

The data processing in CV is done by acquiring an image, then extracting the information that needs to be processed. CV obtains the input images using a high-level approach; based on this approach, classifications are attained that render the appropriate results [11]. Thus, recognition methods are developed to detect the images. The obtained images interact with the machine through HMI applications, such as assistance devices relying on voice and hearing [12], developed for specific platforms. Many prediction-based methodologies and machine-learning algorithms have been developed for CV detection that are related to HMI [13, 14]. The current work addresses the errors that occur when combining structured and unstructured data. It uses the tree method for classification, which allows it to attain maximum gain, and regression is used to reduce the error in combining the structured and unstructured data, maximizing the precision. The main contributions of this paper are as follows:

- We maximize the structured and unstructured data classification accuracy by applying the classification and regression approach.
- We reduce the error rate using two-level visual information processing (2LVIP) techniques, which help to minimize the misclassification rate.
- Furthermore, we obtain the maximum information gain value for both the structured and unstructured data.
- Finally, we improve the precision by classifying the structured and unstructured data and reduce the error by combining the classification and regression methods.

The rest of the paper is arranged as follows. “[Related works](#)” describes the various research options in terms of vision-based information processing, “[Proposed 2LVIP method](#)” explains the 2LVIP working process, “[Results and discussion](#)” evaluates the efficiency of the 2LVIP system, and conclusions are made in “[Conclusion](#)”.

Related works

Monocular robot vision (MRV) was proposed by Chan and Riek [15] for unseen salient object detection done in parallel with discovery prediction. Unsupervised foraging of objects (UFO) is the fastest and most accurate method for notable object discovery. It is done via the real-world perceptions of robots. The main intention is to improve autonomy and resolve robotic challenges while engaging in a salient object discovery process. The embedded computer-vision system in traffic surveillance was introduced by Mhalla et al. [16]

for multi-object detection. This method, which is used for detecting traffic objects in traffic scenarios, consists of a robust detector that makes use of a generic deep detector and enhances detection accuracy.

Wang et al. [17] developed a scale-aware rotating-object detection system at low-level high resolution for obtaining high-level semantic information with aerial imagery. Intersection-over-union (IoU) loss coupled with scale diversity detects orientation. The proposed method is used to improve the accuracy of the rotating bounding box. Kulik et al. [18] addressed CV for intelligent robots by proposing a convolutional neural network for object detection that detects flags indicating unsatisfactory results for different objects. Training and the testing of objects is maintained throughout.

Shin et al. [19] equipped unmanned surface vehicles with object detection and tracking abilities to improve accuracy. The proposed system contributes to an extensive baseline stereo vision system designed to enhance sea surface estimation. It is applicable for long-range object detection, and the semantic segmentation required for detecting objects is done via the oblique convolution designed by Lin et al. [20]. The artificial intelligence developed for the CV uses pixel classification. The hourglass network analyzes the local extremes.

Maggipinto et al. [21] modeled two-dimensional data for CV virtual metrology (VM) using deep learning. The model is used for automatic feature extraction to improve the accuracy and scalability of the VM. The modeling is done using both spatial and time evolutions on real industrial data, including data semiconductor manufacturing. Luo et al. [22] introduced a vision-based detection system for a dynamic workspace involving workers on foot. Multiple detections are used for object tracking and action recognition, and the system determines two types of action data, such as classes and locations. A density-based spatial clustering algorithm is then used to analyze the dynamic workspace.

Jiang et al. [23] proposed fusing spatiotemporal data for hydrological modeling with vision-based data. Three steps are used in this model. The first step involves the fusion of multi-source spatiotemporal for the incorporation of big data. The second step is associated with short- and long-term forecasting, whereas the third step models the streamflow. A multi-object detection (MOD) method is introduced for autonomous vehicle applications that fuses three-dimensional light detection and ranging (LIDAR) and camera data. Zhao et al. [24] provided solutions for recognizing objects by identifying regions of interest (ROI) in the initial processing stage. Later, a convolutional neural network (CNN) is adopted for the recognition of objects. Sliding windows are used for the candidate object region detection of real-time autonomous vehicles. The introduced system maximizes the object detection region and minimizes the misclassification error rate.

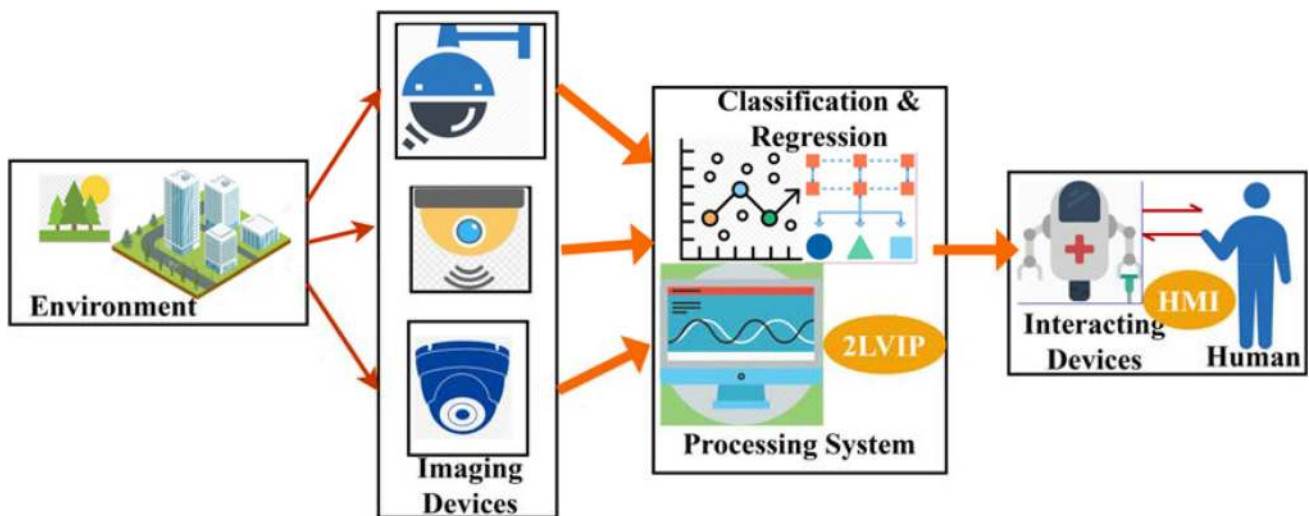


Fig. 1 HMI using 2LVIP

Liu et al. [25] implemented a reference frame Kanade–Lucas–Tomasi (RF-KLT) algorithm for extracting the features in fixed regions. The dimensions of the features are reduced to detect the class boundaries. This work was done in a real-time environment, allowing the efficiency of the system to be evaluated with the help of an augmented and actual video dataset. The system was able to classify anomalies successfully in a robust and cost-effective manner. Fang et al. [26] proposed integrating CV with an ontology that can be used to identify the hazards on construction sites using a knowledge graph. Shu et al. [27] introduced a human–computer interaction mode through the interactive design of intelligent machine vision. Their objective was to improve the accuracy of the functioning algorithm, and the point-and-click results are based on Fitts’ Law.

In this work, two-level visual information processing is considered for performing semantic object detection. The efficiency of the system is evaluated against the methods used in [15, 24], and [25] because these methods work perfectly while analyzing objects. In other words, they are effective in detecting objects and regions in a robust and cost-effective manner. In addition, these methods can be used to evaluate real-time datasets and improve the object and region recognition process.

Proposed 2LVIP method

CV is used for learning the instances of semantic objects in an automatic detection manner. This paper’s objective is to improve the precision of classifying structured and unstructured data together, thus reducing the error introduced by combining these classification methods. It uses regression methods to do so; that is, the proposed method uses two-

level visual information processing (2LVIP) to obtain the maximum gain from the inputs. Figure 1 depicts the proposed model.

Figure 1 clearly depicts the overall architecture of the HMI system with 2LVIP. The imaging devices are used to gather information from the environment. The collected information is then processed by the 2LVIP. The regression and classification techniques are incorporated into the 2LVIP to maximize unstructured and structured data processing. The proposed processing method identifies the gain-related feature information in the data in the first level and optimizes it to maximize the gain. In the second level, the errors in the classifications are determined, and they are removed via the regression process, thus stabilizing the precision to meet the HMI application’s demands. The following equation represents the structured and unstructured images used as the input for the CV system:

$$l_0 = \sum_{t_a}^{d_t} \left(1 + \frac{i_0}{ni_0} \right) * c_0 + \sqrt{\left(\frac{s_r + u_r}{ni_0} \right)}. \tag{1}$$

Equation (1) indicates the analysis l_0 of structured s_r and unstructured u_r images, where the structured image includes the size of the image and presentation, whereas the unstructured image includes variations in size, frame, and patterns. The structured image remains the same for the single input image when a number of images $\{i_0^1, i_0^2, \dots, ni_0\}$ are used. In this case, ni_0 represents the number of images that are captured at the appropriate time t_a . The time the image is captured is denoted as $d_t(c_0)$. After the classification of the structured and unstructured data is done using the tree model. Equation (2a) below indicates the grouping of the trees:

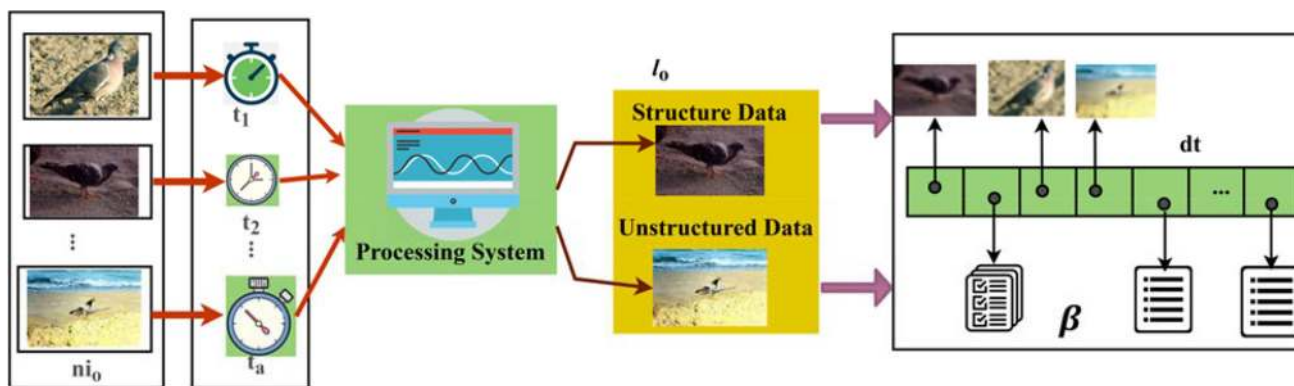


Fig. 2 Input data processing

$$\theta = \left. \begin{aligned} & i_0 + \prod_{c_0}^{ni_0} \left(1 + \frac{i_0}{ni_0} \right) * s_r \forall \text{ structured data} \\ & (c_0 + ni_0) * \sqrt{\frac{i_0}{\sum_{t_a} c_0}} + u_r \forall \text{ unstructured data} \end{aligned} \right\}. \quad (2a)$$

Object detection occurs for both structured and unstructured data, where $1 + \frac{i_0}{ni_0} * s_r$ represents the structured image data. The analysis of the unstructured data is denoted as $\sqrt{\frac{i_0}{\sum_{t_a} l_0}} + u_r$. The combination of these two parts contains the necessary information. Equation (2b) is used to obtain useful information:

$$\beta = \left[l_0(i_0) + \prod_{f'}^{c_0} ni_0 + \frac{ni_0}{\sum s_r} * (d_t - t_a) \right] + \left(1 + \frac{c_0 + (u_r - s_r)}{ni_0} \right). \quad (2b)$$

The extraction of useful information β from the data is the initial step for classification here; the analysis is done by evaluating Eq. (2b). The data processing is illustrated in Fig. 2.

In Fig. 2, the processing is carried out for the structured and unstructured data via $\frac{ni_0}{\sum s_r} * (d_t - t_a)$. Thus, it considers the two types of data with based on the classification based on the mathematical computation $1 + \frac{c_0 + (u_r - s_r)}{ni_0}$ and examines the data that are necessary for automatic detection in a real-time environment.

Classification via the discrete finite value

The discrete finite value is used to derive the classifications for structured and unstructured data. It is based on data that are acquired at a particular time and indicates two sets, such as the finite ∂_0 and infinite ∂' . Thus, by combining Eqs. (2a) and (2b), Eq. (3) below is derived for identifying the data in f' :

$$\alpha(\beta) = \left\{ \begin{aligned} & \left(1 + \frac{i_0}{ni_0} \right) * s_r + i_0 \\ & \text{such that } s_r \in f' + c_0 \\ & l_0(i_0) + \left(1 + \frac{c_0 + (u_r - s_r)}{ni_0} \right) * (d_t - t_a) \\ & \text{such that } u_r \in \left(1 + \frac{l_a}{ni_0} \right) \end{aligned} \right\}. \quad (3)$$

The useful information is extracted for the classification $\alpha(\beta)$ is formulated above. The structured data $s_r \in f' + c_0$ is identified when the image is captured. The unstructured $u_r \in \left(1 + \frac{l_a}{ni_0} \right)$ is used in the analysis of a number of input images. The information is extracted in this step, and a discrete set of values is identified. Equation (4a) below is then used to determine the classification process for this discrete set of values:

$$\alpha(l_0) = \prod_{t_a}^{d_t} l_0 + \beta * (\partial_0 + \partial') \left[\left(1 + \frac{c_0 + (u_r - s_r)}{ni_0} \right) + \left(\sqrt{\frac{i_0}{\sum_{t_a} l_0}} + ni_0 \right) * (c_0 + i_0) + \left(\sqrt{\frac{s_r + u_r}{ni_0}} + (s_r + u_r) \right) \right] + \left(1 + \frac{l_a}{ni_0} \right) * (t_a - d_t). \quad (4a)$$

Datasets are either finite or infinite based on this classification scheme. The data belonging to finite sets are distinguished as $\partial_0(i_0 + l_0)$; these are the data that are acquired at a particular time. If data are not gathered on time, then it is considered to be unstructured data. Equation (4b) is then used to represents the finite and infinite classifications:

$$\alpha = \left\{ \begin{aligned} & \left. \begin{aligned} & \frac{\sum_{l_0}^{c_0} \left(l_0 + \frac{f'}{ni_0} \right) * \frac{e'}{s_r}}{\frac{i_0}{\sum_{t_a} l_0} * \prod_{t_a} (c_0 + e')} \right\}, \partial_0 \\ & \left. \begin{aligned} & \prod_{d_t}^{c_0} \beta + l_0 * \frac{u_r}{ni_0} \\ & \left(1 + \frac{l_a}{ni_0} \right) + (c_0 + i_0) + \prod_{d_t}^{i_0} u_r + e' \end{aligned} \right\}, \partial' \end{aligned} \right\}. \quad (4b)$$

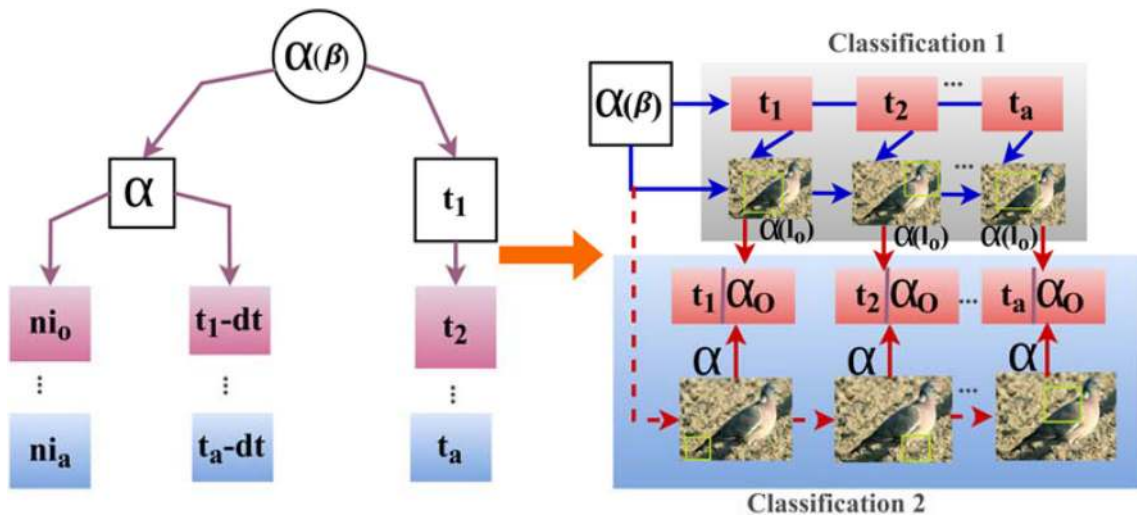


Fig. 3 Classification process

For discrete values e' , the classification is carried out by observing if the data are processed at fixed times with the structure $\frac{e'}{s_r} + (l_0 * f') + (d_t - t_a)$. The unstructured data that are not processed at the specific time are represented by $(l_0 * \frac{u_r}{ni_0}) + (u_r + e') - t_a$. The classification is then carried out and attains maximum gain. The maximum gain is obtained by combining the structured data s_r and unstructured data u_r . The error that occurs during processing is $l_0 + (s_r + u_r)$. The error is due to the misclassification of discrete values (α_0), which is computed using Eq. (4c):

$$\alpha_0 = \left[\left[\left(\left(l_0 + \frac{f'}{ni_0} \right) * (e' + \beta) + \frac{i_0}{\sum_{t_a} l_0} \right) + d_t / f' \right] * s_r + \left[(t_a * c_0) * \left(l_0 * \frac{f'}{ni_0} \right) + (l_0 + \beta) \right] * u_r \right] - (d_t - f') \tag{4c}$$

A misclassification leads to a finite value e' (∂_0) that is not discrete, as observed in Eq. (4c). In these cases, $(l_0 + \frac{f'}{ni_0}) * (e' + \beta)$ represents the discrete value that extracts useful information, although the identification is not made on time. The classification process is illustrated in Fig. 3.

The computation of $(l_0 + \beta) * u_r$ results in an infinite set of values. Following this process, the accurate classification of the finite set is examined, and the error is reduced by deriving the regression algorithm.

Minimizing the errors with regression

The regression method has been used to predict the structured and unstructured data to reduce

the error caused at the time of classification ∂' . This second-level extracts better precision from the HMI application. The process obtained in the first level extracts the maximum gain by evaluating both s_r and u_r . During this process, some errors occur. A prediction is made by observing the training set that is processed in the preceding process. Equation (5a) below is used to evaluate the regression through prediction:

$$q_0 = \begin{cases} \sum_{l_0}^{c_0} \frac{f'}{ni_0} + (\beta + \alpha) * t_a \\ \left(\alpha + \sqrt{\frac{l_0}{\prod_{c_0} l_0}} \right) * (s_r + u_r) + \sum_{t_a}^{l_0} (f' + \beta). \end{cases} \tag{5a}$$

The regression allows the prediction of ∂_0 and ∂' . The dependent value is used to find the prediction q_0 used in $\left[\frac{f'}{ni_0} + (\beta + \alpha) \right]$ to identify the unstructured data. Here, $\sum_{t_a}^{l_0} (f' + \beta)$ represents the useful information that is extracted for further processing; the independent value is derived via Eq. (5b) below:

$$p' = \prod_{\alpha}^{l_0} \left(1 + \frac{s_r + u_r}{ni_0} \right) * \sum_{\beta=1}^{\alpha_0} (l_0 + f') \tag{5b}$$

Here, $1 + \frac{s_r + u_r}{ni_0}$ denotes the structured and unstructured data that are combined to attain the maximum gain. The prediction method is formulated in Eq. (6), as follows:

$$\gamma = o'(g_0) - t_a * \left\{ \left[\prod_{q_0}^{p'} \left(\frac{\beta(i_0)}{\alpha} * c_0 \right) + \sum_{s_r}^{u_r} (t_a - \beta) \right] + \left[(\partial_0 + \partial') + (l_0 / f') - (d_t - e') + \prod_{\beta} c_0 + i_0 \right] \right\} + (g_0 - h_0). \tag{6}$$

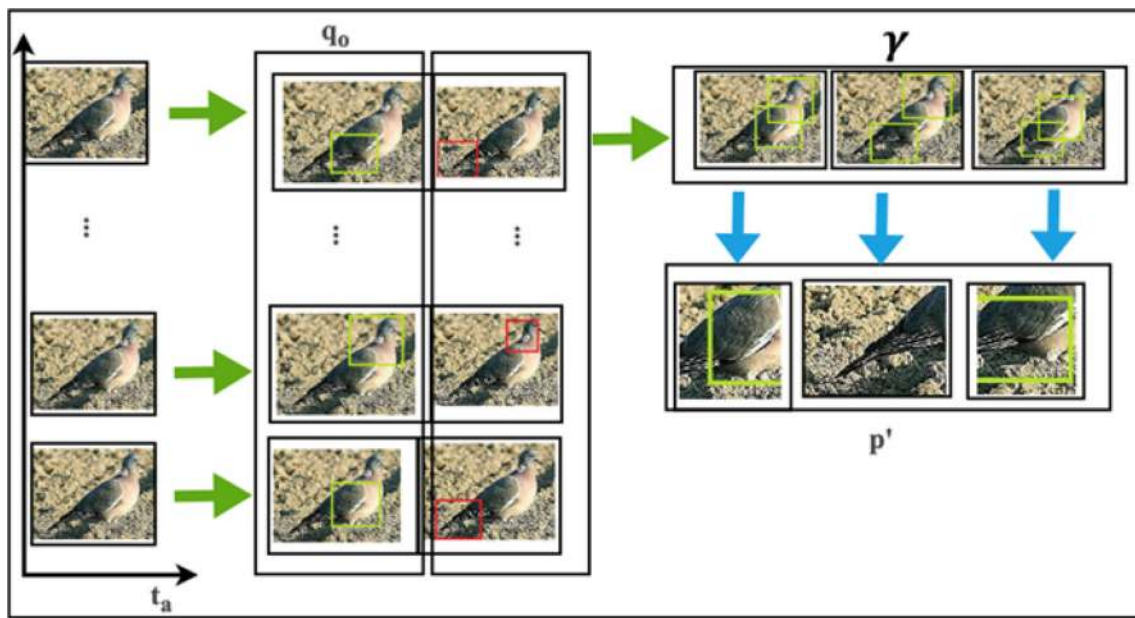


Fig. 4 Illustration of the regression process

Here, $\frac{\beta(i_0)}{\alpha} * c_0$ is used to gain information for the input image. Figure 4 illustrates the regression process.

It determines the $(t_a - \beta)$ targeted for extraction from the obtained gain at a particular time. The error o' is used to examine the foregone and forthcoming data for the HMI applications g_0 and h_0 . In this manner, the error of the foregone data is compared with that for the forthcoming data to provide an optimal result. The objective of the second level is to reduce the errors made in information processing. Thus, the training set is used to compare and improve processing further, see Eq. (7).

$$\varnothing = \begin{cases} 1 & \text{if } \sum_{q_0}^{p'} \|(l_0 + f') * i_0\| + \left(\frac{\beta + e'}{i_0}\right) + (g_0 - h_0) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The training data \varnothing are used to formulate the 'if and otherwise' conditions for $\|(l_0 + f') * i_0\|$, thus $\frac{\beta + e'}{i_0}$ is used for identifying the discrete data. The foregone and forthcoming data are used to evaluate this prediction-based method. The foregone data is again trained and used for the forthcoming data based on the 'if' condition. The training calculation in Eq. (7) is used to make the comparison needed for the HMI application. The cost function calculated for the regression method minimizes the error between the predicted and actual values used for processing the data. Equation (8) below is used to obtain the cost function:

$$\Delta = \frac{1}{j'} \prod_{\alpha}^{j'} (\gamma - h_0)^2 + \left(\beta + \frac{p' + q_0}{i_0}\right). \quad (8)$$

The cost function Δ is used to attain the data points j' used to find the prediction value, then forthcoming data is subtracted, along with the prediction value. The $\left(\beta + \frac{p' + q_0}{i_0}\right)$ is processed to obtain the actual value. The cost function is used for obtaining better results for the classification and regression method. Next, the prediction-based regression method is used to reduce the error with Eq. (9), as follows:

$$\rho = \begin{cases} \left(\Delta + \frac{\beta + e'}{i_0}\right) * (g_0 - h_0) + (\alpha + \beta) = 1 \\ \prod_{\alpha}^{l_0} (f' + i_0) * \left(\frac{\beta + \gamma}{\varnothing} + \Delta\right) (g_0 - h_0) \neq 1 \end{cases} \quad (9)$$

The reduction in the error ρ made in Eq. (9) is used to evaluate the error. The cost function $\Delta + \frac{\beta + e'}{i_0}$ is used in the analysis of the regression method. The training set is used to determine the data from the maximum gain $\left(\frac{\beta + \gamma}{\varnothing} + \Delta\right) (g_0 - h_0)$. It derives the lesser error. Afterwards, a final regression method is used to reduce the error from the classification of the structured and unstructured data, see Eq. (10):

$$\delta(i_0) = \begin{cases} \left[\left(\beta + \frac{p' + q_0}{i_0}\right) + (l_0 + f') * \prod_{\beta}^{e'} (\alpha + \gamma) \right] - d_t < o' \\ \sum_{l_0}^{\alpha} (\beta + \gamma) + (l_0 + f') + \left(1 + \frac{l_a}{n_0}\right) * (t_a - d_t) > o' \end{cases} \quad (10)$$

In this regression, $(\alpha + \gamma) - d_t$ has more errors at the time of computing. Figure 5 illustrates this regression-based detection method.

The second condition $(\beta + \gamma) * (t_a - d_t)$ has a lesser error. In this equation, the first condition does not satisfy the objective,

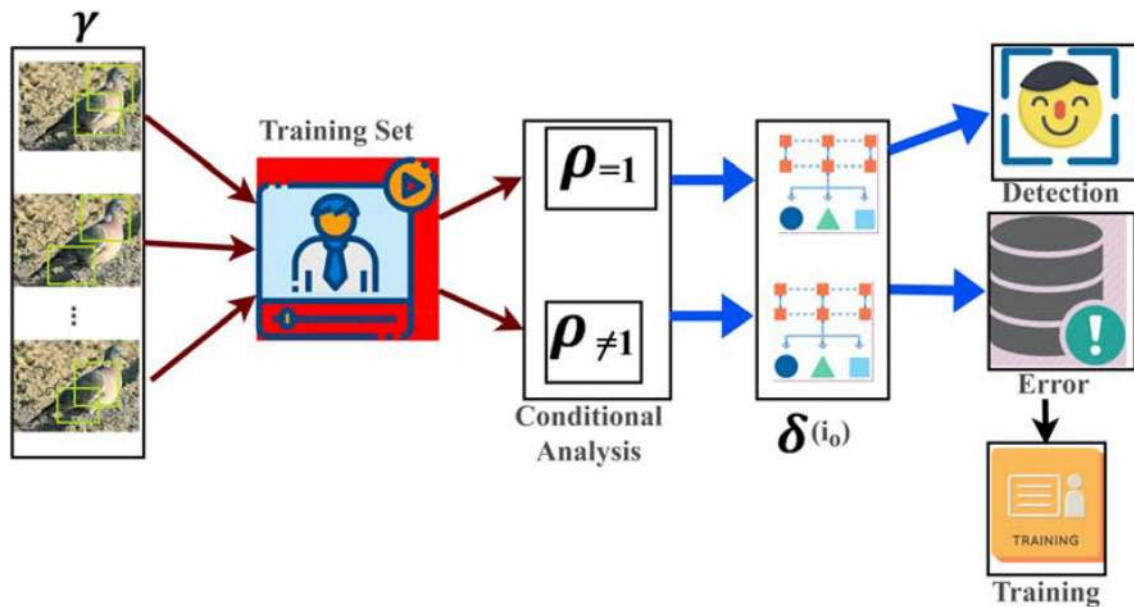


Fig. 5 Regression-based detection process

whereas the second condition satisfies the proposed method (2LVIP). If an object is not classified correctly, the derivative $\frac{f'+\beta}{ni_0} * l_0$ is applied. Thus, the process is updated for better classification. As a result, the precision is enhanced, and error is reduced in the HMI application through the evaluation of $(\alpha + \gamma) * \delta(i_0)$.

Results and discussion

This section discusses the performance evaluation of the proposed method that is done through a comparative study of different metrics. In this assessment, the metrics considered are precision, gain ratio, error, and processing time. The image input from [28] is used in this analysis. It consists of 102 sets of images, along with their annotations. A region’s image with 45 training data points is verified in the analysis, and the results are discussed. The size of the training dataset is 149 mb, and the annotations fill nearly 7 mb. The number of classification and regression instances considered are 20 and 40, respectively. The RF-KLT [25], MOD-3D LIDAR [24], and UFO [15] methods are used in the comparison.

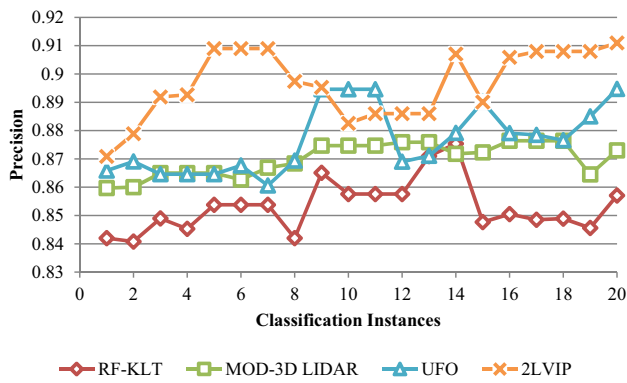
Precision analysis

In Fig. 6a and b, the precisions of the classification instances and information gain ratios are compared. The precision for the proposed 2LVIP is higher due to the evaluation of $\frac{i_0}{\sum_{i_a} l_0}$, i.e., the data obtained from the classification tree. The analysis of the unstructured data $(1 + \frac{i_0}{ni_0}) * s_r + u_r$ deter-

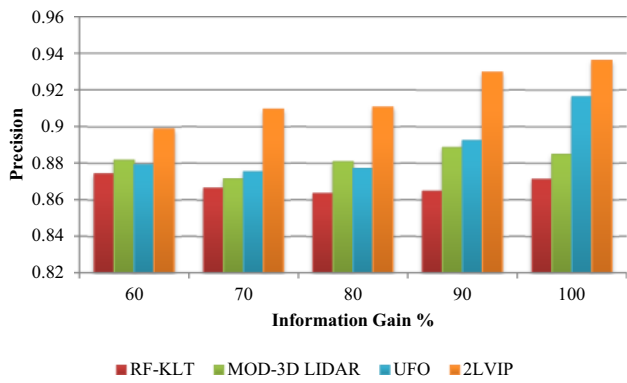
mines the resultant data $(c_0 + ni_0)$. The analysis proceeds by introducing the regression methods to acquire the number of classification instances based on time $\partial_0 + \frac{s_r+u_r}{i_0}$. The data so attained have higher classifications; thus, the gain also increases. The gain is determined by deriving $(\alpha + \frac{f'}{i_0}) + \frac{\beta}{e'}$, which represents an actual stage in the processing. The data processing in this 2LVIP method is analyzed by evaluating $\emptyset + i_o * (g_0 + h_0)$. The foregone and forthcoming data are used in the analysis of the precision. By comparing the classification of the proposed method with those of the other three methods, it can be seen that a smaller precision value is obtained when the information gain decreases. Thus, if the classification is improved, then the gain also increases. Thus, the error reduces $(i_0 + \frac{\alpha+f'}{\beta}) * t_a - d_t$ and vice versa. In Fig. 6, a higher precision is obtained by computing $t_a + f' + \beta$.

Information gain analysis

The gain for the proposed method is shown in Fig. 7, which indicates that as the classification increases, the gain increases. When comparing the 2LVIP method with the other three methods, namely, RF-KLT, MOD-3D LIDAR, UFO, the 2LVIP method has the highest gain. The gain performance is found through $1 + \frac{c_0+(u_r-s_r)}{ni_0}$, which acquires the image and classifies the data. Improvements in the gain can be analyzed by determining $s_r \in f' + c_0$ and $u_r \in (1 + \frac{l_a}{ni_0})$. The maximum gain is obtained when the classification of the structured and unstructured data is done correctly. The classification methods are derived in Eq. (4a), where $\beta * (\partial_0 + \partial')$ represents the extraction of the useful information.



(a) Precision Versus Classification Instances



(b) Precision Versus Information Gain %

Fig. 6 a Precision versus classification instances. b Precision versus information gain %

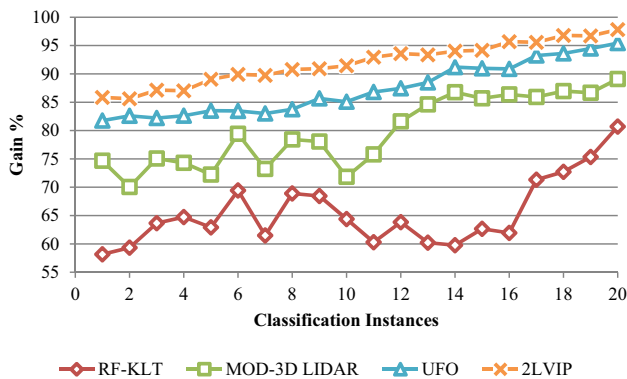
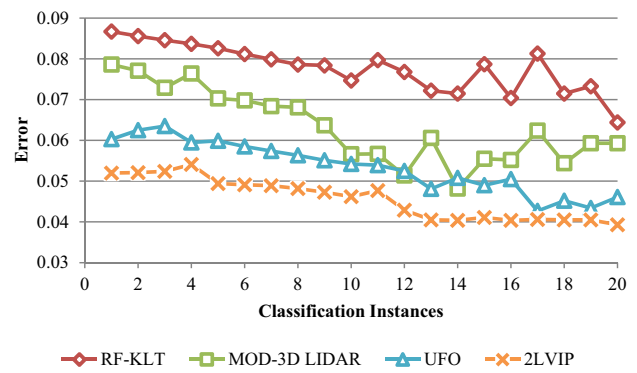
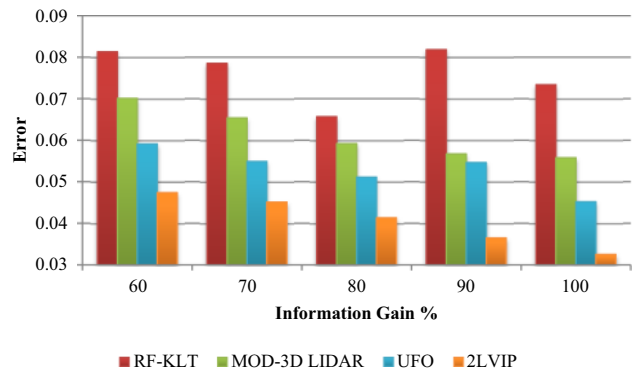


Fig. 7 Gain % versus classification instances

The data belonging to finite sets are distinguished as $\partial_0(i_0 + l_0)$. If the data are not collected promptly, then they are represented as unstructured data, namely, by $(l_0 + \frac{f'}{ni_0}) * \frac{e'}{sr}$. These data are not processed at the specific time $(l_0 * \frac{ur}{ni_0}) + (ur + e') - ta$.



(a) Error Versus the Classification Instances

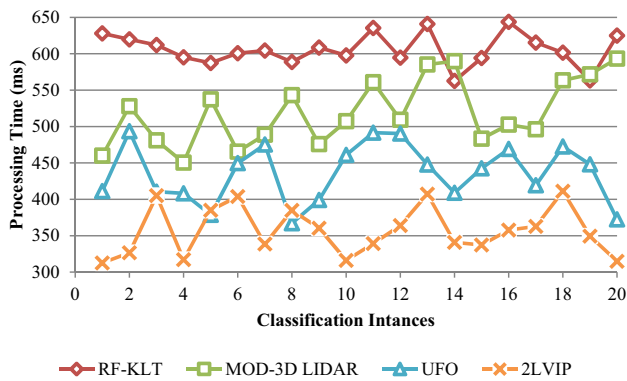


(b) Error Versus Information Gain %

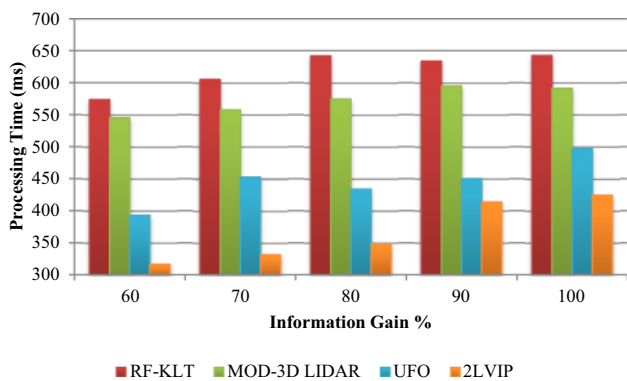
Fig. 8 a Error versus the classification instances. b Error versus information gain %

Error analysis

In the 2LVIP method, the error is reduced by increasing the number of correct classifications at the specific time $(l_0 * \frac{ur}{ni_0}) + (ur + e') - ta$. Thus, these classifications signify the useful information obtained and also the error that occurs during processing, namely, $l_0 + (sr + ur)$. The error is due to the misclassification of discrete values. The expression $1 + \frac{sr+ur}{ni_0}$ denotes the structured and unstructured data that are combined to attain the maximum gain. Next, the prediction $\gamma + \beta(i_0) * \alpha$ is determined to reduce the error in the forthcoming data. The analysis is done by computing the process at a specified time. Thus, the prediction is used to obtain errorless unstructured and structured data. Equation (9) is used to reduce the error, and $(\Delta + \frac{\beta+e'}{i_0}) * f' + i_0$ is derived. Thus, information gain obtained from the classification increases, so that the error is reduced. The error is smaller for the 2LVIP method than the errors for the existing three methods, as seen by evaluating the $o' + f'(i_0)$ (refer to Fig. 8a, b).



(a) Processing Time Versus Classification Instances



(b) Processing Time Versus Information Gain %

Fig. 9 a Processing time versus classification instances. b Processing time versus information gain %

Analysis of the processing time

Although there is a high amount of data for classification when using the 2LVIP method, it has the smallest processing time, as shown in Fig. 9a. As the number of classifications increases, the analysis necessary increases. It is found by determining $(t_a * c_0) * \left(l_0 * \frac{f'}{n_{i_0}} \right)$, which evaluates the

specific time needed for processing. Using Eq. (4a), the appropriate classification is carried out for the acquired data. If $\left(l_0 + \frac{f'}{n_{i_0}} \right) * (e' + \beta)$ represents the discrete value that extracts useful information and identification is not made on time, it must be improved. The data obtained from the classification reduces the error at the time of classification ∂_0 , thus providing a better processing time for the HMI application. The classification analysis uses $\alpha(l_0) * (t_a - d_t)$, which extracts the data at the appropriate time. The data belonging to finite sets are distinguished as $\partial_0(i_0 + l_0)$. If the information gain increases, then the processing time for the proposed method, formulated with $\gamma + \beta(l_0)$, holds constant (refer to Fig. 9a, b). The results of the above comparison are tabulated in Tables 1 and 2, respectively.

The cumulative outcome of the proposed method shows that it improves precision and gain% by 3.61% and 9.42%, respectively. It also reduces the error and processing time by 6.51% and 21.75%, respectively.

In terms of the information gain, the proposed method achieves 4.54% higher precision, 7.65% less error, and 26.48% less processing time.

Analysis of the instances

In Fig. 10, the regression instances for the classified data increase, so the error decreases through the derivation of $\frac{\beta(i_0)}{\alpha} * c_0$. It determines $(t_a - \beta)$ for extraction at a particular time from the obtained gain. The error σ' is used to examine the foregone and forthcoming data for the HMI applications g_0 and h_0 . Here, $\|(l_0 + f') * i_0\|$ is used to identify the analysis of the input data. Thus, the $\left(\beta + \frac{p'+q_0}{i_0} \right)$ are computed for better precision. The maximum gain $\left(\frac{\beta+\gamma}{\sigma} + \Delta \right) (g_0 - h_0)$ is used to reduce the error.

The maximum gain is increased by finding $\left(l_0 / f' + \beta \right) - (d_t - e')$, which is used to decrease the error. In Fig. 11, if the number of classifications the proposed method makes

Table 1 Metrics for the different classification instances

Metrics	RF-KLT	MOD-3D LIDAR	UFO	2LVIP
Precision	0.8571	0.873	0.8947	0.911
Gain %	80.7	89.1	95.43	97.83
Error	0.0644	0.593	0.0461	0.0393
Processing time (ms)	624.97	593.45	372.42	414.94

Table 2 Metrics for the different gain %

Metrics	RF-KLT	MOD-3D LIDAR	UFO	2LVIP
Precision	0.871	0.8847	0.9161	0.936
Error	0.0734	0.0558	0.0452	0.0325
Processing time (ms)	642.55	591.2	496.8	424.11

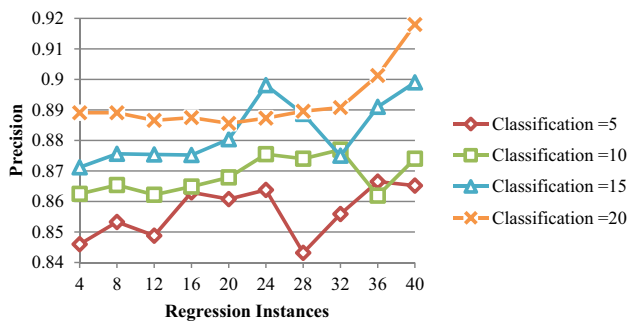


Fig. 10 Precisions of the different regression and classification instances

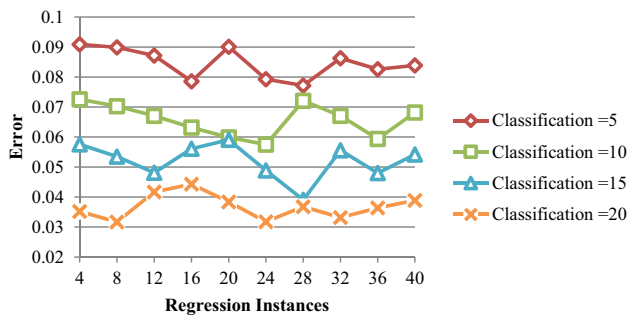


Fig. 11 Errors for the different regression and classification instances

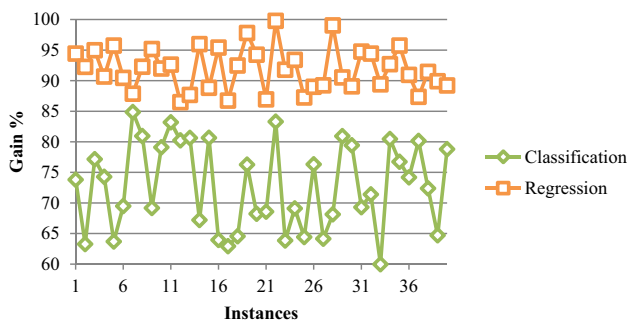


Fig. 12 Gain % for the different classification and regression instances

increases, then its error is reduced. For the 2LVIP method, the $(g_0 - h_0) + (\alpha + \beta)$ obtained with the final regression method is used to reduce the error in the classification of structured and unstructured data. The reduction in the error ρ is found with Eq. (9), then used to evaluate the error.

Given the number of instances, the regression produces more gain than the classification; hence, the error is reduced. The classification of the structured and unstructured data is evaluated with the regression equation $\delta(i_0)$. Equation (10) is used to reduce the error and determines the regression process based on the prediction $\gamma + \frac{f'+e'}{i_0}$ (refer to Fig. 12).

Conclusion

In this paper, a two-level visual information processing method is discussed for improving the precision of human—machine interaction systems. The input obtained from the imaging devices is classified into structured and unstructured data based on the available information. The maximum information gain is then extracted, and regression is used to mitigate the errors in the information gained. The regression process uses the training set data to reduce the error and processing time via recursive discrete and finite value estimations. The regression analysis is recursively handled using predictive cost estimation. Training is performed to reduce the processing time used in extracting useful information from the visual input. By attuning the classification and regression processes, the proposed method is found to maximize the precision and information gain in detecting the objects and reduce the error and processing time. In the future, an optimization-based regression approach will be applied to classify the objects.

Acknowledgements The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through Research Group no. RG-1438-070.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Guo L, Zhou D, Zhou J, Kimura S, Goto S (2018) Lossy compression for embedded computer vision systems. *IEEE Access* 6:39385–39397
- Liu Y, Liu J, Ke Y (2020) A detection and recognition system of pointer meters in substations based on computer vision. *Measurement* 152:107333
- Manogaran G, Shakeel PM, Fouad H, Nam Y, Baskar S, Chilamkurti N, Sundarasekar R (2019) Wearable IoT smart-log patch: an edge computing-based Bayesian deep learning network system for multi access physical monitoring system. *Sensors* 19(13):3030

4. Georgiou T, Liu Y, Chen W, Lew M (2019) A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. *Int J Multimed Inf Retr*. <https://doi.org/10.1007/s13735-019-00183-w>
5. Lopez-Fuentes L, van de Weijer J, González-Hidalgo M, Skinemoen H, Bagdanov AD (2018) Review on computer vision techniques in emergency situations. *Multimed Tools Appl* 77(13):17069–17107
6. Fouad H, Mahmoud NM, El Issawi MS, Al-Feel H (2020) Distributed and scalable computing framework for improving request processing of wearable IoT assisted medical sensors on pervasive computing system. *Comput Commun* 151:257–265
7. Shi Y, Zhang Z, Huang K, Ma W, Tu S (2019) Human-computer interaction based on face feature localization. *J Vis Commun Image Represent*. <https://doi.org/10.1016/j.jvcir.2019.102740>
8. Sultani W, Mokhtari S, Yun HB (2017) Automatic pavement object detection using superpixel segmentation combined with conditional random field. *IEEE Trans Intell Transp Syst* 19(7):2076–2085
9. Gardecki A, Podpora M, Kawala-Janik A (2018) Innovative internet of things-reinforced human recognition for human-machine interaction purposes. *IFAC Pap Online* 51(6):138–143
10. Fan S, Li J, Zhang Y, Tian X, Wang Q, He X, Zhang C, Huang W (2020) On line detection of defective apples using computer vision system combined with deep learning methods. *J Food Eng*. <https://doi.org/10.1016/j.jfoodeng.2020.110102>
11. Asadi P, Gindy M, Alvarez M, Asadi A (2020) A computer vision based rebar detection chain for automatic processing of concrete bridge deck GPR data. *Autom Constr* 112:103106
12. Yin Y, Li H, Fu W (2020) Faster-YOLO: an accurate and faster object detection method. *Digit Signal Process*. <https://doi.org/10.1016/j.dsp.2020.102756>
13. Shin J, Kim M, Paek Y, Ko K (2018) Developing a custom DSP for vision based human computer interaction applications. *Multimed Tools Appl* 77(22):30051–30065
14. Ren X, Silpasuwanchai C, Cahill J (2019) Human-engaged computing: the future of human–computer interaction. *CCF Trans Pervasive Comput Interact* 1(1):47–68
15. Chan DM, Riek LD (2020) Unseen salient object discovery for monocular robot vision. *IEEE Robot Autom Lett* 5(2):1484–1491
16. Mhalla A, Chateau T, Gazzah S, Amara NEB (2019) An embedded computer-vision system for multi-object detection in traffic surveillance. *IEEE Trans Intell Transp Syst* 20(11):4006–4018
17. Wang Y, Zhang Y, Zhang Y, Zhao L, Sun X, Guo Z (2019) SARD: towards scale-aware rotated object detection in aerial imagery. *IEEE Access* 7:173855–173865
18. Kulik S, Shtanko A (2020) Using convolutional neural networks for recognition of objects varied in appearance in computer vision for intellectual robots. *Proc Comput Sci* 169:164–167
19. Shin B-S, Mou X, Mou W, Wang H (2017) Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities. *Mach Vis Appl* 29(1):95–112
20. Lin Y, Sun X, Xie Z, Yi J, Zhong Y (2020) Semantic segmentation with oblique convolution for object detection. *IEEE Access* 8:25326–25334
21. Maggipinto M, Terzi M, Masiero C, Beghi A, Susto GA (2018) A computer vision-inspired deep learning architecture for virtual metrology modeling with 2-dimensional data. *IEEE Trans Semicond Manuf* 31(3):376–384
22. Luo X, Li H, Wang H, Wu Z, Dai F, Cao D (2019) Vision-based detection and visualization of dynamic workspaces. *Autom Constr* 104:1–13
23. Jiang S, Zheng Y, Babovic V, Tian Y, Han F (2018) A computer vision-based approach to fusing spatiotemporal data for hydrological modeling. *J Hydrol* 567:25–40
24. Zhao X, Sun P, Xu Z, Min H, Yu H (2020) Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications. *IEEE Sens J* 20(9):4901–4913
25. Liu Y, Yu H, Gong C, Chen Y (2020) A real time expert system for anomaly detection of aerators based on computer vision and surveillance cameras. *J Vis Commun Image Represent* 68:102767
26. Fang W, Ma L, Love PE, Luo H, Ding L, Zhou A (2020) Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Autom Constr* 119:103310
27. Shu Y, Xiong C, Fan S (2020) Interactive design of intelligent machine vision based on human–computer interaction mode. *Microprocess Microsyst* 75:103059
28. Fei-Fei L, Fergus R, Perona P (2006) One-shot learning of object categories. *IEEE Trans Pattern Anal Mach Intell* 28(4):594–611

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.