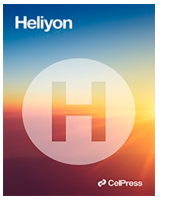
Heliyon 10 (2024) e35929

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

# A real-time vehicle safety system by concurrent object detection and head pose estimation via stereo vision

Julio C. Rodriguez-Quiñonez [a,*], Jonathan J. Sanchez-Castro [a,**], Oscar Real-Moreno [a], Guillermo Galaviz [a], Wendy Flores-Fuentes [a], Oleg Sergiyenko [b], Moises J. Castro-Toscano [a], Daniel Hernandez-Balbuena [a]

[a] Universidad Autónoma de Baja California, Facultad de Ingeniería, Blvd. Benito Juárez S/N, 21280, Mexicali, Baja California, Mexico
[b] Instituto de Ingeniería, Universidad Autónoma de Baja California, Calle de la Normal S/N y Blvd. Benito Juárez, Col. Insurgentes Este, 21280, Mexicali, Baja California, Mexico

ARTICLE INFO

*Keywords:*
Head pose estimation
Object detection
Driver pose classification
Stereo vision
Landmark detection

ABSTRACT

A considerable number of vehicular accidents occur in low-millage zones like school streets, neighborhoods, and parking lots, among others. Therefore, the proposed work aims to provide a novel ADAS system to warn about dangerous scenarios by analyzing the driver's attention and the corresponding distances between the vehicle and the detected object on the road. This approach is made possible by concurrent Head Pose Estimation (HPE) and Object/Pedestrian Detection. Both approaches have shown independently their viable application in the automotive industry to decrease the number of vehicle collisions. The proposed system takes advantage of stereo vision characteristics for HPE by enabling the computation of the Euler Angles with a low average error for classifying the driver's attention on the road using neural networks. For Object Detection, stereo vision is used to detect the distance between the vehicle and the approaching object; this is made with a state-of-the-art algorithm known as YOLO-R and a fast template matching technique known as SoRA that provides lower processing times. The result is an ADAS system designed to ensure adequate braking time, considering the driver's attention on the road and the distances to objects.

## 1. Introduction

In recent years, there has been an increment in research of multiple tasks of Advanced Driver Assistance Systems (ADAS) due to pedestrians having a significant risk of being in an accident at intersections, low mileage zones, on the roadway, etc. [1–3]. To address this problem, ADAS systems involve tasks like vehicle detection, driver gazes and pose estimation, and driver road attention to increment road safety, as attending to the drivers' attention is crucial, considering that a significant proportion of car accidents stem from drivers' lack of awareness [4,5].

As mentioned by Li et al. [6], an important aspect of implementing ADAS systems for preventing accidents is primarily the deaths and injuries that can occur, but in another perspective, accidents also leave enormous economic costs, and according to the National

* Corresponding author.
** Corresponding author.
*E-mail addresses:* julio.rodriguez81@uabc.edu.mx (J.C. Rodriguez-Quiñonez), jonathan.sanchez11@uabc.edu.mx (J.J. Sanchez-Castro).

https://doi.org/10.1016/j.heliyon.2024.e35929
Received 22 March 2024; Received in revised form 15 May 2024; Accepted 6 August 2024
Available online 7 August 2024

Highway Traffic Safety Administrator (NHTSA) in 2018 the showed a total costs of 242 billion dollars for crashes, and specifically for the case of confirmed distracted driver's the total cost is roughly sixteen percent of the total cost. In many instances, the lack of awareness or distraction can be deduced from the driver's posture, which can impede their ability to make critical decisions aimed at preventing collisions, like in the work of C. Addanki et al. [7], where they analyze the head pose but also include the consideration of the eyelid movement for drowsiness that is directly related for knowing the driver fatigue. Also, the work of S. Jha et al. [8] presents an approach to estimate the driver's gaze through the head orientation of the driver. Therefore, pose estimation techniques, such as Head Pose Estimation (HPE) [9–11], play a considerable role in ADAS application tasks. HPE is an ongoing field with a wide range of applications like facial tracking, facial expression, CGI, and driver awareness and fatigue, among others [12–14]. HPE estimation is a computer vision application that can be estimated through different techniques, which are image-based, depth-based, and mixed methods. In Abate et al. [15] a complete structure of HPE methods is depicted and also mentions the most used techniques for common applications in driver assistance, which are possible via the capture of body and facial images, which are processed to classify the driver's activity. However, several research studies focus on the driver's pose without considering the detection of road individuals within the scene who may be injured due to driver distraction. Therefore, concurrent systems that combine head pose estimation and road object detection can provide enhanced information to alert the driver when a dangerous scenario happens [7,16,17]. The proposed work presents a driver assistance technique designed to aid drivers in assessing potential hazards. This research focuses on enhancing safety by integrating two key components. First, the approach involves estimating the driver's head pose to classify the area of attention. Second, it includes the detection of objects external to the vehicle, enabling the measurement of the distance between the vehicle and these objects. Both aspects of this study leverage triangulation and stereo-vision methods for accurate depth estimation, thereby enhancing the overall scenario assessment.

The proposed system of this work continuously computing the HPE of the driver through constant solving of rotational matrices at 17 frames per second, concurrently an external stereo vision system and in combination with an object detection algorithm that computes the distance of the object, which allows the classification of different alarming scenarios for the driver. The novelty of this work is divided into the following points:

- The continuous HPE computation is done by using 3D points recovered by the triangulation method, which works at 17 FPS with a MAE of 0.87°.
- The incorporation of an object detection phase based on the YOLOv4 network that also extracts depth information via stereo vision, in tandem with the HPE of the driver, for assessment of dangerous scenarios.
- An intuitive driver alarm is designed to alert drivers in dangerous scenarios, particularly when they are not attentive to objects at intersections and low-speed areas, which pose a high risk of accidents involving pedestrians, bicycles, and other hazards.

Considering that low speed zones (40mph and under) have a considerable amount of accidents, for example, for single person injuries in accidents below 30(mph) are 14.5 %, and combined with the accidents at 40 mph, the percentage increases to 42.5 %; as the U.S. Department of transportation mentions in the traffic safety facts [2], this research aims to contribute towards the development of more robust safety assistance systems.

The rest of the manuscript is organized as follows: Section 2: Related Works, where recent works are presented. Section 3: Methodology, where the basic principles and core functionality of the proposed system is presented. Section 4: Experiments and Discussion, where the results of our testing is shown and the viability of the present system is discussed. And finally, Section 5: Conclusion.

## 2. Related Works

Recently, image-based, depth-based, and mixed methods have incorporated Deep Learning (DL) [18,19] for pose estimation and HPE, which has lowered the management of extensive data sets for image applications. For example, S. Jha et al. [8] used the Gaussian Process Regression as their back-bone for estimating probabilistic regions of drivers' gaze and pose and combined it with neural networks, which delivered better results for the drivers' gaze estimation. Liu et al. [20] introduced an end-to-end network for human pose estimation, which also includes cues of the human head, whereby a CNN was trained with second-order statistics to describe the region features of the human joints in the image. Deep learning (DL) applications for ADAS systems are diverse, with each employing unique evaluation methods. Nevertheless, artificial intelligence is emerging as a critical aspect of vision-based applications, including HPE, as noted in the review of [21] i.e., the work of H. Liu et al. [22] is an approach of combining convolutional networks for HPE tasks and more specifically for infrared images, with the goal to provide an alternative for RGB cameras that are affected by illumination variations. In S. Biswas et al. [23], the authors implement a pre-trained VGG16 model to classify three zones, left, right, and straight, of the passengers gaze, by using multiple sensing devices (electroencephalographic, electrocardiographic, among others) applied to the passenger's and image captures from the passengers behavior, this with the premise of using the passengers to create an indirect judgment on the driver behavior. Lastly, in Zhang et al. [24], adding to the use of an NN, the information of distinctively head dues known as orientation tokens is presented, which serve as auxiliary data to the HPE estimation. An important aspect to highlight is that conventional HPE techniques rely on 2D imagery which suffer from changes in occlusion, blur, illumination, among other factor, being the illumination one of the most common and works like there are research made with infrared cameras. Ju et al. [25], used IR cameras for obtaining the HPE through a CNN trained with 2D gamma distribution labels of the drivers' head pose to address the changes in illumination show considerable progress in mitigating this problem but they still rely on one camera and the HPE is estimated by steps and not as a continuous value. But Lately, as 3D data sets or depth techniques and also depth sensors become more available, the use of

3D data has become desirable given that it allows to treat HPE as a Perspective-n- Point problem and also provides robustness against the mentioned problems in 2D image techniques also,like changes in illumination, occlusion, and blur, among other variables when capturing images [26,27],i.e., like in Akrout et al. [14], where the authors used a conventional algorithm of Viola and Jones to create 3D points for the inclusion of the head pose in their fatigue detection system. Also, in C. Bisogni et al. [28],the authors propose an end-to-end network that relies only upon depth data, with which through independent HP from fusion of fractal encoded features and Key-points detection of the facial depth data is estimated, and then combined to create a final result of HPE through a fusion technique known as Nelder-Mead Method (NMM), their result show promising advance in lowering processing times but still show levels of Mean Absolute Error of above 4°.

Koay et al. [21] noted that driver assistance systems differ based on sensor type, methodology, and applications. The different applications of ADAS have included the utilization of depth sensors for assessing driver road attention due to their capability to capture significant data regarding the person's pose, including body, arms, hands, head, and facial features. Moreover, depth sensors or cameras are desirable as they are non-intrusive devices that do not distract or disturb the driver. Therefore, in recent years, multi-camera concurrent systems have emerged as an efficient, non-invasive solution. For example, Khairdoost et al. [29], equipped a vehicle with multiple cameras to obtain distance measurements and an IR depth sensor for the driver pose, and all this visual information (vehicle distance, gaze, head pose, among other factors.) predicts drivers maneuvers for switching lanes. The work of Khairdoost et al. is also an example of how stereo vision helps with the decision-making of ADAS systems, as depth estimation is used to aid the driver maneuver with the gaze and pose estimation. This system creates a circular attention zone where the driver is paying attention on the road, and with the depth estimation, creates 3D points that help to see if the driver will make a right or left turn or remain straight. Another work incorporating stereo vision is from Leng et al. [30], who apply the disparity maps obtained with a Regional Convolutional Neural Network (RCNN) to detect and provide better spatial and geometrical information about the objects on the road. This research provided state-of-the-art performance against common Faster-RCNN models but still suffers with overlapping vehicles.

The distinctive works referenced before estimate the pose and sometimes the gaze of the driver, and commonly, AI techniques are used to solve these tasks as a classification problem. In many cases, these works are for one application i.e., driver monitoring or object detection. However this research, as discussed in the previous section, looks to provide a system that identifies the objects on the road and concurrently analyses the drivers' heads pose to identify its attention on the driving and all of this is capable thanks to the leverage of stereo vision which enables us to obtain depth at considerable real-time speeds with a pair of cameras inputs. And, as shown by Shi et al. [31] and Khairdoost et al. [29] that leverage stereo vision, depth information is desirable with ADAS systems given that it provides extra layers of information that help with the drivers' assistance tasks (HPE and object distance), and specifically in this research, provide an ADAS system that considers low mileage zones.

## 3. Methodology

As mentioned in the introduction, this paper proposes an approach for real-time HPE and concurrently detecting pedestrians, providing their depth estimation for the main objective of alarming the driver of dangerous scenarios when crossing a crosswalk or in a low mileage zone.

This work uses a mathematical approach (triangulation technique) to provide the 3D coordinates for the alarm system, which is different from previously cited works where they use AI to estimate the depth of information. In conjunction with landmark detection approaches, the triangulation technique enables extracting 3D coordinates. It can be used in various tasks, like structural deformation, HPE, or the medical sector [32,33]. Also, as in the presented research, the authors in Ref. [33] applied landmark detection to recover the angles through stereo vision. However, in this research, the translation of the recovered 3D points in two different timestamps is used to calculate the Euler angles for the HPE.

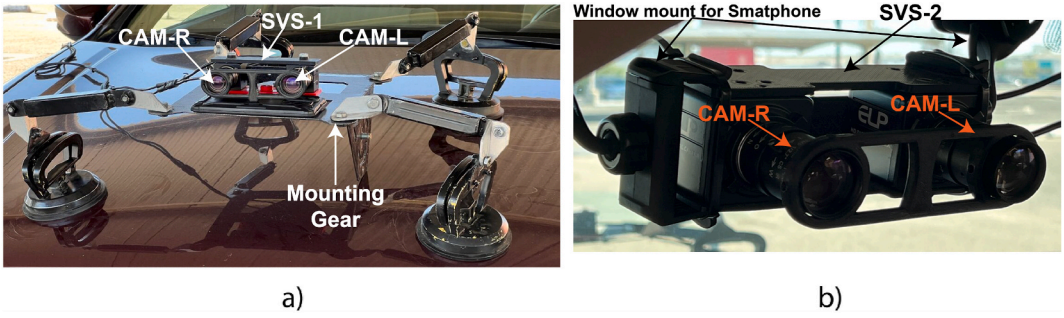



**Fig. 1.** Stereo Vision Systems: a) SVS-1 mounted on a base for object detection on the road, and b) SVS-2 mounted inside of the vehicle for the HPE of the driver.

### 3.1. Stereo vision system

For this research, we propose using two stereo vision systems. The first is presented in Fig. 1 a), where it is mounted on a base for adhering it to the car hood; this uses object detection to detect pedestrians. The second is, presented in Fig. 1 b), this stereo vision system inside the vehicle sees the face of the driver and performs HPE by 3D facial landmark detection.

To compute the 3D Coordinates, there are a series of prerequisites to verify first: The angles of view of each camera (Horizontal and vertical), the camera resolution, the compensation angle, and the Calibration setting [34]. Once these settings are introduced, depth information can be computed.

The depth information is computed by Eq. (1), Eq. (2) and Eq. (3):

$$X = a\left(\frac{\sin C * \sin B}{\sin(B + C)}\right) \tag{1}$$

$$Y = a\left(\frac{\cos B * \sin C}{\sin(B + C)} - \frac{1}{2}\right) \tag{2}$$

$$Z = a\left(\frac{\sin B * \sin C * tan\,\beta}{\sin(B + C)} - \frac{1}{2}\right) \tag{3}$$

where X, Y, and Z are spatial coordinates given in meters, and $a$ is the separation between the cameras, also known as the baseline. These equations can obtain the 3D facial points for HPE and also an object's distance from the vehicle with the mounted cameras.

As shown in Figs. 2 and 3, $a$ is the baseline of the cameras. $B$, $C$, and $\beta$ are the corresponding angles to the specific point (P) in space. The specific computation of $B$ and $C$, is shown in Eq. (4) and Eq. (5):

$$B = B_i + B_0 \tag{4}$$

$$C = C_i + C_0 \tag{5}$$

where $B_i$, $C_i$ and $B_0$, $C_0$ are the computed variables of the point (P) and constants, respectively. The constants $B_0$, and $C_0$ are the values that fall outside the horizontal angle of view of each camera. These constants are values estimated with Eq. (7) and Eq. (6).

$$B_0 = 90 - \frac{H}{2} \tag{6}$$

$$C_0 = B_0 \tag{7}$$

For the unknown variables $B_i$ and $C_i$, Eq. (8) and Eq. (9) are used for their computation.

$$B_i = H\frac{W - P_{xl}}{W} \tag{8}$$

$$C_i = H\frac{P_{xr}}{W} \tag{9}$$

where $H$ is the horizontal angle of view of the corresponding camera, $W$ is the width resolution of the corresponding image. $P_{xl}$ and $P_{xr}$ are the surface point pixel coordinates on the left and right images on the horizontal axis, respectively. $\beta$ is the corresponding angle for the vertical axis and is computed by Eq. (10).

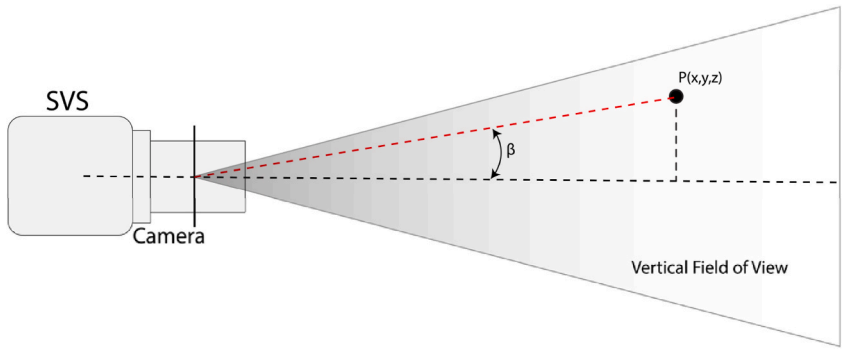



**Fig. 2.** Vertical field of view of the SVS.

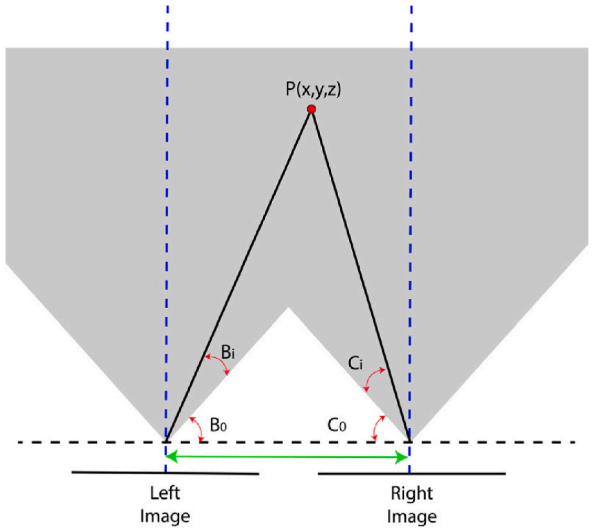



**Fig. 3.** Horizontal field of view of the SVS.

$$\beta = (V)\left(\frac{\frac{V_R}{2} - P_y}{V_R}\right) \tag{10}$$

where $P_y$ is the corresponding pixel value of the point (P) on the vertical axis, this value can either be from the right or left image. $V$ is the corresponding angle of view on the vertical axis of the cameras. $V_R$, is the height of the image in pixels for the corresponding camera, and finally, the $SP_{ylr}$ is the surface pixel point coordinate on the left image.

### 3.2. Head pose estimation

In this paper Eq. (1), Eq. (2), and Eq. (3) are used to provide 3D points to estimate the head pose of a person by solving rotational matrix equations to know the pose of a driver [35]. The rotation of spatial points can be made by applying rotational matrices like the ones seen in Equations (11)–(13).

$$\boldsymbol{R}_x(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} \tag{11}$$

$$\boldsymbol{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{12}$$

$$\boldsymbol{R}_z(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

where $\boldsymbol{R}_x$, $\boldsymbol{R}_y$, and $\boldsymbol{R}_z$ are the rotational matrices on each axes. $\psi$, $\theta$, and $\varphi$ are the corresponding angles for each rotation on each axis, respectively, which are also known as Yaw, Pitch, and Roll angles or Euler Angles.

The process of rotating a set of spatial points is not commutative, so it has to be made in a specific order, first by the $X$ axis ($\boldsymbol{R}_x$), then by the $Y$ axis ($\boldsymbol{R}_y$), and finally by the $Z$ axis ($\boldsymbol{R}_z$), as seen in equation (14).

$$\boldsymbol{R} = \boldsymbol{R}_x(\psi)\boldsymbol{R}_y(\theta)\boldsymbol{R}_z(\varphi) \tag{14}$$

The multiplication of the component in Eq. (14), provides an extensive and complicated matrix, and conventionally it is abbreviated to make it easy to manage. This matrix abbreviation is presented by equation (15).

$$\boldsymbol{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{15}$$

where $R_{i,j}$ are elements of the rotation matrix. These elements rely on the sine and cosine functions.

The process of rotating a set of spatial points involves a simple matrix multiplication and is shown by the following Eq. (16).

$$\boldsymbol{P}' = \mathbf{R} * \mathbf{P} \tag{16}$$

where ($\boldsymbol{P}'$) are the rotated points of the original ones (**P**). The only requirements are the known points and the specified angles for rotation.

Now, if there is a need to recover the rotation of a known set of 3D points, like in this work, as we investigate the rotation of a driver's head, an inverse process must be undertaken. First, the original 3D points (**P**) must be known as the rotated ones ($\boldsymbol{P}'$), and the only unknown variables left to compute are the $R_{i,j}$ elements of the rotation matrix, with which Eq. (17), Eq. (18), and Eq. (19) can be solved to find each Euler Angle.

$$\psi = atan\,2\left(\frac{R_{23}}{\cos\theta}, \frac{R_{33}}{\cos\theta}\right) \tag{17}$$

$$\theta = -\sin^{-1}(R_{31}) \tag{18}$$

$$\varphi = atan\,2\left(\frac{R_{21}}{\cos\theta}, \frac{R_{11}}{\cos\theta}\right) \tag{19}$$

$$\begin{bmatrix} x_1 * R_{11} + y_1 * R_{12} + z_1 * R_{13} & x_2 * R_{11} + y_2 * R_{12} + z_2 * R_{13} & x_3 * R_{11} + y_3 * R_{12} + z_3 * R_{13} \\ x_1 * R_{21} + y_1 * R_{22} + z_1 * R_{23} & x_2 * R_{21} + y_2 * R_{22} + z_2 * R_{23} & x_3 * R_{21} + y_3 * R_{22} + z_3 * R_{23} \\ x_1 * R_{31} + y_1 * R_{32} + z_1 * R_{33} & x_2 * R_{31} + y_2 * R_{32} + z_2 * R_{33} & x_3 * R_{31} + y_3 * R_{32} + z_3 * R_{33} \end{bmatrix} = \boldsymbol{P}' \tag{20}$$

These equations are known to recover each Euler angle and are derived by solving for each angle using the elements $R_{i,j}$ from Eq. (15). The next step is to find the elements $R_{1,1}$, $R_{2,1}$ , $R_{3,1}$, $R_{2,3}$ and $R_{3,3}$ this unknown elements are computed by solving the system of Eq. (20) for each of them.where $x_n$ , $y_n$, and $z_n$ are the spatial points used as templates to calculate the rotation angles. After solving for each matrix element needed for the angle recovery, the following equations are left to compute.

$$R_{11} = \frac{(x_2' y_1 - y_2 x_1')(-z_1 y_3 + z_3 y_1)}{(x_2 y_1 - x_1 y_2)(-z_1 y_3 + z_3 y_1)}$$

$$\frac{-(x_3' y_1 - y_3 x_1')(-z_1 y_2 + z_2 y_1)}{-(x_3 y_1 - x_1 y_3)(-z_1 y_2 + z_2 y_1)} \tag{21}$$

$$R_{21} = \frac{(y_2' y_1 - y_2 y_1')(-z_1 y_3 + z_3 y_1)}{(x_2 y_1 - x_1 y_2)(-z_1 y_3 + z_3 y_1)}$$

$$\frac{-(y_3' y_1 - y_3 y_1')(-z_1 y_2 + z_2 y_1)}{-(x_3 y_1 - x_1 y_3)(-z_1 y_2 + z_2 y_1)} \tag{22}$$

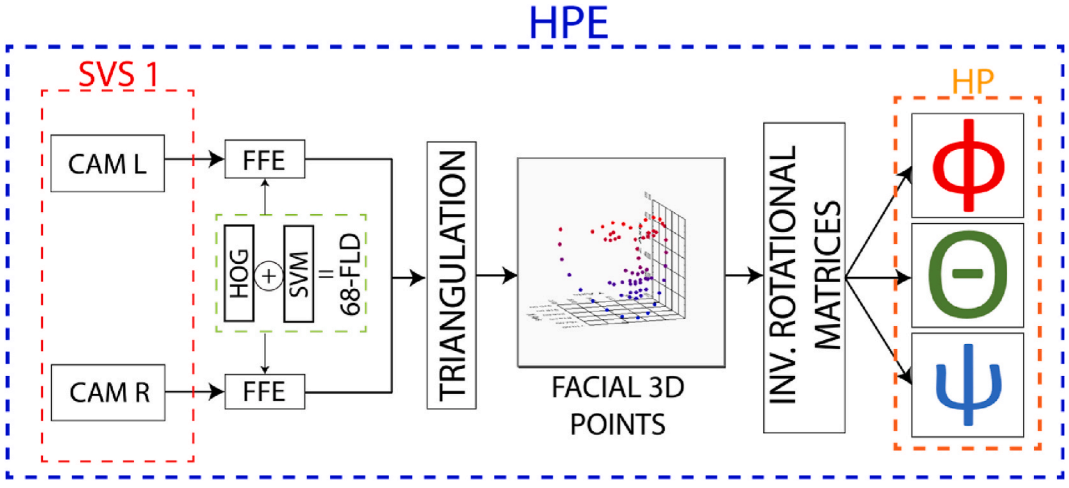



**Fig. 4.** Graphic description of each corresponding section of the HPE process, where the flow of the data is depicted and also an overview of the 68 facial landmark classifier is included. Here, both images taken pass thru the Dlibs landmark detector, for obtaining matching points for the next step, which is triangulation, then the points are used to obtain the HPE thru inverse rotational matrices.

$$R_{31} = \frac{(z'_2y_1 - y_2z'_1)(-z_1y_3 + z_3y_1)}{(x_3y_1 - x_1y_3)(-z_1y_2 + z_2y_1)}$$

$$\frac{-(z'_3y_1 - y_3z'_1)(-z_1y_2 + z_2y_1)}{-(x_3y_1 - x_1y_3)(-z_1y_2 + z_2y_1)} \tag{23}$$

$$R_{13} = \frac{(x'_2x_1 - x_2x'_1)(-y_1x_3 + y_3x_1)}{-((-z_1x_3 + z_3x_1)(-y_1x_2 + x_1y_2))}$$

$$\frac{-(x'_3x_1 - x_3x'_1)(-y_1x_2 + y_2x_1)}{+(-z_1x_2 + x_1z_2)(-y_1x_3 + y_3x_1)} \tag{24}$$

$$R_{33} = \frac{(z'_2x_1 - x_2z'_1)(-y_1x_3 + y_3x_1)}{-((z_1x_3 + z_3x_1)(-y_1x_2 + x_1y_2))}$$

$$\frac{-(z'_3x_1 - x_3z'_1)(-y_1x_2 + y_2x_1)}{+(-z_1x_2 + x_1z_2)(-y_1x_3 + y_3x_1)} \tag{25}$$

where $x'_n$ , $y'_n$, and $z'_n$ are the rotated points.

In this work, three spatial points are used for the head pose recovery, which are extracted by a facial feature extractor algorithm (FFE). Fig. 4 shows the flow of the HPE section of our scenario assessment system. First, the images from both cameras of the SVS 1 are sent to the FFE algorithm, which provides a set of 68 points representing the characteristics of a person's face, and by applying the same FFE algorithm to both images in the SVS system, a set of matching points can be obtained for the triangulation. From these 68 3D points computed by triangulation, three are used for the estimation of the Euler Angles, which are the tip of the nose and two points of the face contour, one on each side. After the facial 3D points are available, the next step is to use the solved Eq. (21), Eq. (22), Eq. (23), Eq. (24), and Eq. (25) for the recovery of the Euler Angles (see Fig. 4).

### 3.3. Stereo Object Detection

The proposed system requires the concurrent operation of two units; the first one shown in sec.3.2 is for HPE and the second one with Stereo Object Detection (SOD). This second unit of the proposed ADAS approach is graphically shown in Fig. 5, and is described as follows: First, the two available frames from each camera are captured, and using the object detection algorithm known as YOLO-R with the image from "CAM1″ multiple persons and vehicles can be detected within the frame. Next, the pattern matching technique from Ref. [36] is used to find the same object on the image from "CAM2″, to finally perform the triangulation from both images to obtain the objects' X, Y, and Z coordinates. With these coordinates, the object's position can be classified and then used to assess the situation.

The pattern matching presented by Ref. [36] where selected for de SOD unit as it consists of fast template matching with a linear complexity, which provides fast and efficient template matching when compared with pattern matching algorithms like Sum of Absolute Differences (SAD) [37,38], among others.

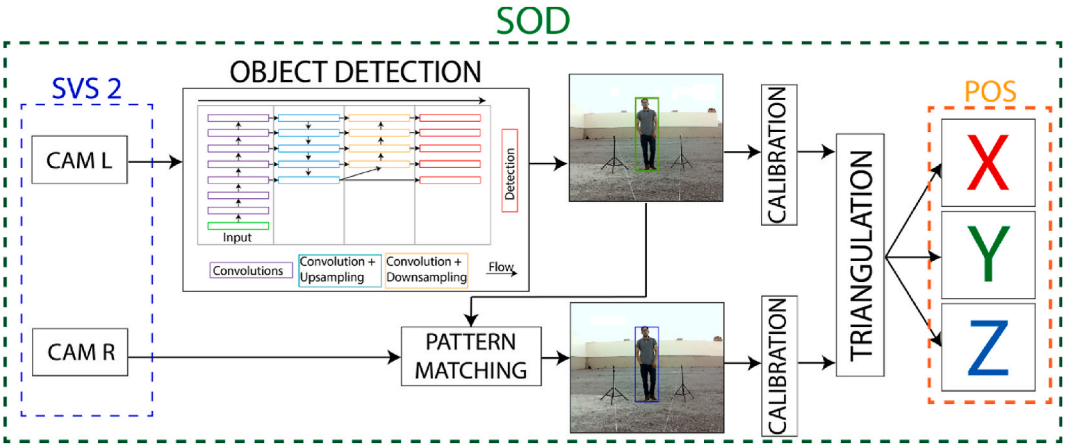



**Fig. 5.** Graphic description of the SOD section. The input images from the external SVS are taken, and only the left image is passed to the object detection network (based on YOLOv4 Network), from the objects detected, a pattern matching technique is used to find the correct template match on the right images to enable the triangulation process to be made.

The implemented pattern matching technique was developed to provide a fast and low computation technique to use with an object detection algorithm. The algorithm is known as Subtraction of Relationship Array (SoRA). This algorithm is based on SAD, but instead of matrices subtraction, SoRA implements the inclusion of relationship arrays and the characteristics of a SVS, which reduces the computational cost and provides a decreased processing time when compared to other known techniques, a general view of SoRA is shown in 1, for a comprehensive understanding of the SoRA algorithm, we recommend referring to the article by Ref. [36] et al. article.

### 3.4. Scenario assessment

The complete diagram of the proposed system is shown in Fig. 6, where both HPE and SOD units are shown and represent the data extraction portion for the scenario assessment, in said image, the proposed ADAS, takes into consideration the Euler Angles for classifying the position on to the driver is viewing, the five positions are frontal view (FV), left mirror (L), middle (M), Stereo or center console (S), right mirror (R), and finally if the driver looks down to its smartphone (T), as shown in Fig. 7. For the classifications of these six sectors a neural network (NN) is implemented, as shown in Fig. 6, where a demonstrative figure shows in which part of the process it is located, this NN has one input layer with four neurons, two hidden layers with 15 neurons each, and at the output layer there are six neurons corresponding to the six classification zones.

**Algorithm 1.** SoRA Pseudo Code

1: Read frames **L**(n × m) and **R**(n × m)
2: Set template size: **template**(n × n)
3: **Select the template in Left Image:**
4: Initialize template **T**
5: Compute row means of **T**
6: Compute template relationship array **Tr** from **T**
7: **Search for Similarity in Right Image:**
8: Initialize best match variable:*minDissimilav rity*
9: Select parallel rows Rc(n × m) on the **R**(n × m) 10: Compute column means of the matrix Rc*n* × m
11: Compute Relationship array **Rr** from **Rc**
12: **for** i = 0; length(**Rr**); i++ **do**
13: **Dissimilarity**[i] = **Tr**(i)-**Rr**(i)
14: **Best Match position:**
15: Row = get rows of **Rr**
16: Column = min(**Dissimilarity**) + *n*/2

The data used for the training of the NN are obtained after validating the performance of the HPE section, where it was found that the proposed approach could achieve levels of Mean Absolute Error (MAE) of 0.87° moving steps of 3°. The HPE method does not require a specific data set, the Dlibs FLD does the estimation of the facial features for each camera image and consequently creates a adequate match between the estimated points to enable the triangulation technique, then rotational matrices provide the Euler angles when solving Eq. (17), Eq. (18), and Eq. (19).

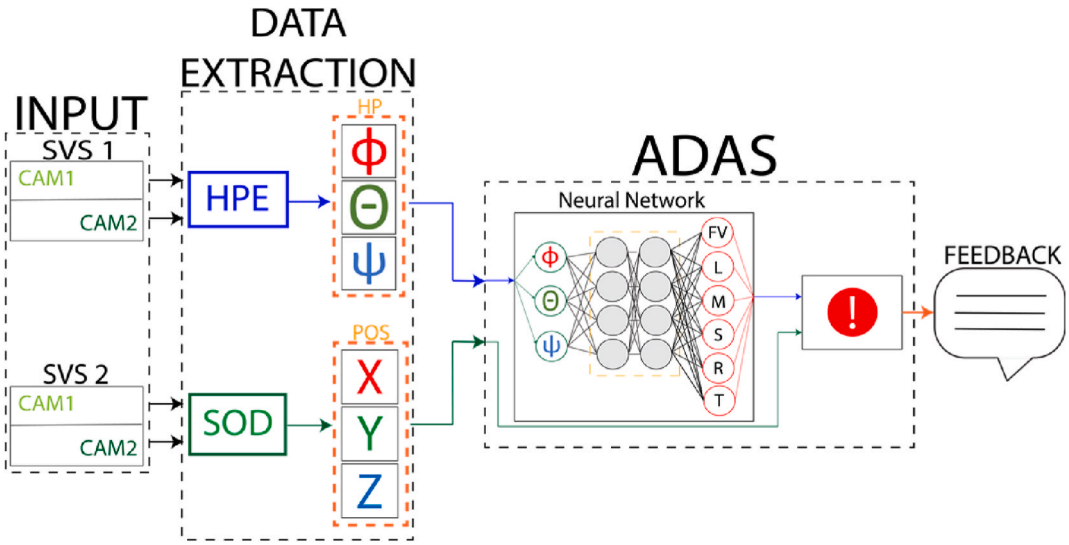



**Fig. 6.** Proposed ADAS system diagram, in this figure all the main units and process are shown HPE, SOD and the proposed NN which has only two hidden layers with 15 neurons each, the exclamation symbol is referring to a warning and this node provides with the classification of the driving scenario in question.

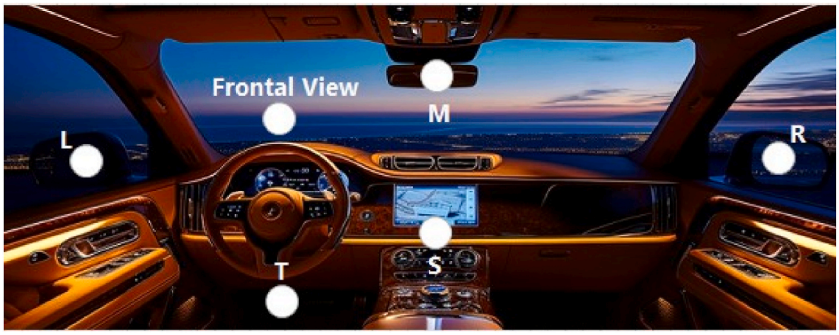



**Fig. 7.** Virtual visualization of the zone classification of the drivers pose.

For the driver assistance system, the head zone classification and the distance between the car and the object are used to provide a warning to the driver to pay attention to the zone that is not viewable to the driver and there might be a potential hazard. The considerations for providing feedback of a hazardous situation are established by the Boolean Eq. (26). This equation depends on the detection to which zone the driver is looking and the detection of objects near the vehicle.

$$Alarm = L(A+B) + M + S + R(C+B) + T \tag{26}$$

where A, B, and C are the corresponding sectors of the visible area on the stereo images. These sectors can be adjusted to the desired value, also, there is a depth parameter that defines the max distance to consider an object "close" to the vehicle for the alarm signal.

## 4. Experiment and analysis

The presented work is developed to provide information of SVS tackling an important application on the automotive industry: a driver safety system. This research implements fast template matching, HPE, and NN with the characteristics of a SVS to develop a system that can analyze the drivers head movement and the distance between the vehicle and a visible object. The software of the proposed system was developed with a combination of LabVIEW 2021 and Python language 3.9, with a resolution of 320 × 240 pixels. Our experiments were performed using two laptops one with an Intel core i5 7th Gen., the other one with an Intel core i7 12 Gen. and both with 16 GB of RAM. Two SVS, each SVS is equipped with two varifocal lens between 5 and 50 mm and equipped with a Sony IMX179 sensor. These cameras are mounted on a 3D printed base, which is separated with a baseline of 100 mm. A driving scenario was simulated on a parking lot to evaluate the proposed work. Inside the vehicle in front of the driver the first SVS was collocated. Outside, the second SVS was mounted on the hood of the vehicle, and from the center of the SVS two lines were attached to divide the three sectors previously discussed A, B, and C, these lines were attached to a camera mounting base and collocated at approximately 5 m of the car. In Figs. 9 and 10 it can be observed the collocation and distribution of said cameras and sectors of the image.

The NN used in this research, was trained using compiled information from a test regarding the five zones established. The cameras were mounted inside a vehicle, and a driver was instructed to look at each zone for a specific amount of time to compile enough data values for each angle. After this trial, the computed Euler Angles were analyzed and used to calculate synthetic data with the same standard deviation and mean to increase the data value points for the training of the NN. The point clouds of each zone can be seen in Fig. 8, where it is shown that Yaw and Pitch are the angles well separated and provide the best definition of each zone.

With the estimated point clouds, different clustering models were used but a NN proved a better choice for this task. After the

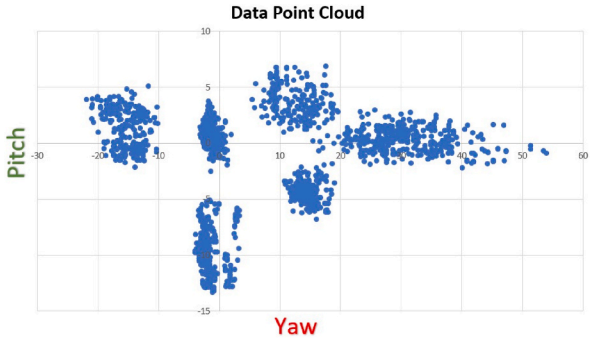



**Fig. 8.** Training data used on the NN for the zone classification, this data is obtained from running the mentioned HPE system, with which a test of the six positions analyzed in this work is made, to obtain well separated data for the training of NN with three layers and a soft-max activation function at the output layer.

**Fig. 9.** View of the SVS and the driver while testing for the HPE.

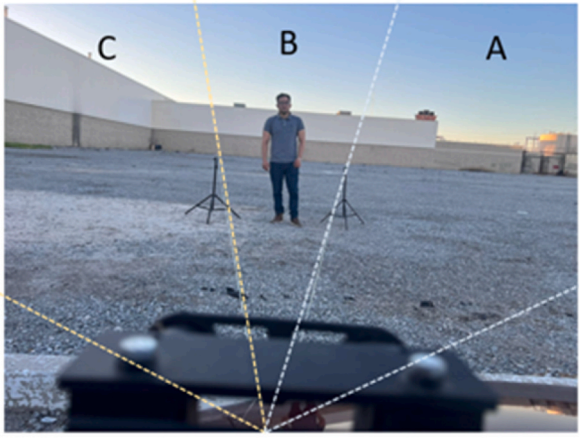



**Fig. 10.** Second SVS collocation and sectors A, B, and C delimitation for the experimentation. This delimitation is defined to divide the scene into sections for the implementation of the driver alarm when a possible object is located in a place where it is not seen.

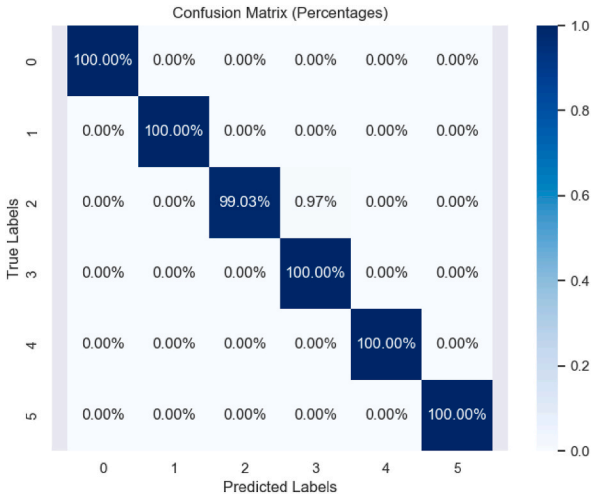



**Fig. 11.** Confusion matrix evaluation of the proposed NN with the provided data points, where: 0 is FV, 1 is L, 2 is S, 3 is R, 4 is M, 5 is T.

training of the NN, it was evaluated and with that data a confusion matrix was created. The confusion matrix is shown in Fig. 11, and it shows the capacity of the final NN trained with the data shown in Fig. 8. The general accuracy of the NN is of 99.83 %, the only label that it has trouble which is when the driver looks at the radio (zone: S), the advantage of implementing the NN instead of a KNN classifier was because of the intrinsic characteristics of the loss function, in this work it is desired a system with flexibility over the drivers pose and using a loss function like soft-max that learns probabilistic distribution of the data.

In Fig. 9, it can be observed the first SVS in front of the driver, the cameras don't obstruct the visibility of the driver and are capable of viewing the drivers face to provide the desired images for the HPE. In Fig. 10, the second SVS is shown and the aforementioned sectors, a subject is used for the evaluation and is also shown in said image.

For the experimentation, a driver is situated in the vehicle and a subject is outside at the same distance as the camera mounts. The evaluation starts with situating the outside subject first at sector A, then B, and finally C. When moving the outside subject onto each sector, the driver will also perform a parallel test by directing its attention to each previously defined classification zone (FV, L, M, S, R, and T). This, to analyze the processing time and the ability of the proposed NN and the overall system to provide feedback if there is a dangerous scenario up front or not.

The implemented object detection system, as said beforehand, consist of the application of the object detection algorithm YOLO-R, which proved to provide excellent results in object detection and classification, and the use of the pattern matching algorithm SoRA [36], due to the shorter processing times over others. The overall SOD unit results under a controlled environment is 0.4 m of Average Error, at distances from 5 to 15 m. When testing the system two camera bases were used to delimit each sector but also to have a distance mark, these bases were placed at 5 m from the car, and the subject was situated under that distance line. In outdoor testing the average RMSE is 0.73 m, which proved that still under outdoor lighting it can still provide relatable measurements for these applications.

When compared to similar depth estimations like in Yang et al. [39], the proposed system shows competitiveness denoted by a comparison of RMSE and the Squared Relative Error (SqRel). The work of Yang et al. [39] presents an approach of a deep network paired with a custom epipolar attention module for multi-view depth estimation (called MVS2D), which is for 3D scene reconstruction, where a considerable number of depth points are computed for short and medium distances., also they compared their findings with an existing data set and network both named DeMoN [40], also used for multi-view depth estimation. In contrast, the proposed system works for larger distances and is designed to work with a smaller number of objects, and for each object, the depth is computed in relation to its centroid on the captured stereo images, and also this is done with just a pair of frames instead of a consecutive stream of video or multiple camera images. The proposed system of MVS2D obtains depth estimation with a low RMSE, and when comparing it to our proposed work, their documented level of RMSE is 0.38 and for our proposed work, is 0.73, and lastly, for DeMoN is 2.27, which is the highest of the three. When comparing these three systems with their Relative Squared Error, MVS2D showed a lower level of SqRel with 0.1; ours came in second with 0.21, and DeMoN in third with 1.68. The MVS2D system showed the lowest level of error under these two-metric comparisons. Nevertheless, MVS2D is tailored for short distances, necessitating the capture of multiple images and/or videos for optimal functionality. Furthermore, its design is optimized for operation within laboratory conditions. Hence, we consider that the proposed approach is more convenient as it is suitable for medium to long-distance applications where processing time is critical, and the achieved RMSE is adequate for this application.

The proposed ADAS system performs well under adequate levels of illumination (with clouds or shade over the vehicle), with a 98 % accuracy rate. Applying the proposed system to different evaluations under different lighting scenarios provided the expected outcomes of the system, which were a decrease in accuracy, with a 96 % accuracy in the morning at 10 a.m. (with the sun over the vehicle) and 91 % accuracy in the evening at 6 p.m. (with the sun directly illuminating the face of the driver). Regarding the number of objects that the system could detect, this is bound to the YOLO network performance of detection, but for our application, the system proved capable of detecting multiple persons throughout the testing phase and would register the closest person to the vehicle. The proposed system runs on average at 17 frames per second, which is capable of being used in low-speed road sections, like: crossing sections, school zones, and others. As, stated by the American Association of State Highway and Transportation [41] a driver has a 1.5 s reaction time (RT) to a possible stopping situation, and as shown in Table 1 the RT at different speeds impacts heavily on the breaking distances (BD), given that as the vehicle speed increments the RT distance also increases. Table 1 shows an example of an object detected at a distance of 25m, and with the AASHTO guidelines on breaking, the accountability of the RT distance and the addition of the distance that an image frame represents, the range of speed on which our system can be applied is estimated (bellow 30 km/h), this table shows that at 20 km/h there is approximately 12 m to apply the brakes with sufficient time to come to a stop before impact (SbI), and as it increases to 30 km/h it is reduced to almost 2 m to brake. These numbers are estimated with a constant deceleration of 3.4 $m/s^2$. In a case of emergency braking the vehicle can decelerate at a much higher rate which widens the minimum breaking distance to much higher velocities.

Fig. 12 represents the example of an object at 25m, in this figure it can be graphically appreciated the amount of meters needed for a safe breaking situation, it shows that with a speed of 20 km/h there is sufficient distance for the driver to start the braking process or with an AESB. In this example an adequate supervision of the driver awareness and object detection can provide a quick feedback to the driver to apply the brakes before the minimum BD needed for a safe stopping, but when the speed of the vehicle increases this minimum BD will also increase, and according with Table 1, with a speed of 30 km/h there is only 1.71m left to start braking, which makes having an automatic system to alert the driver can help with the braking situation. This analysis shows that the proposed approach is applicable under 30 km/h given the USA guidelines, that manage a constant deceleration of the vehicle.

Overall, the proposed system can leverage stereo vision characteristics to provide stable and accurate depth estimation and HPE, which can be used in ADAS. Thanks to its level of depth perception, which is obtained through parallax, The proposed system proves capable of obtaining reliable depth measurements, either inside or outside of the vehicle, as noted in our results, but also, through

**Table 1**
Table of Breaking Distance for different vehicle velocities. For this research application, the response time of our system is estimated at 17 FPS, with which the calculations for the Reaction time (RT) and breaking distances (BD), taking into account the AASHTO guidelines, to estimate if the vehicle will stop before impact (SbI) are computed.

| Frame rate | Speed (Km/h) | Speed (m/s) | RT | RT (m) | FPS (m) | OD | BD (3.4m/s^2) | BD Window | Min BD | SbI |
|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 20 | 5.56 | 1.50 | 8.33 | 0.33 | 25 | 4.60 | 20.40 | 11.73 | Yes |
| 17 | 30 | 8.33 | 1.50 | 12.50 | 0.49 | 25 | 10.30 | 14.70 | 1.71 | Yes |
| 17 | 40 | 11.11 | 1.50 | 16.67 | 0.65 | 25 | 18.40 | 6.60 | −10.72 | No |
| 17 | 50 | 13.89 | 1.50 | 20.83 | 0.82 | 25 | 28.70 | −3.70 | −25.35 | No |

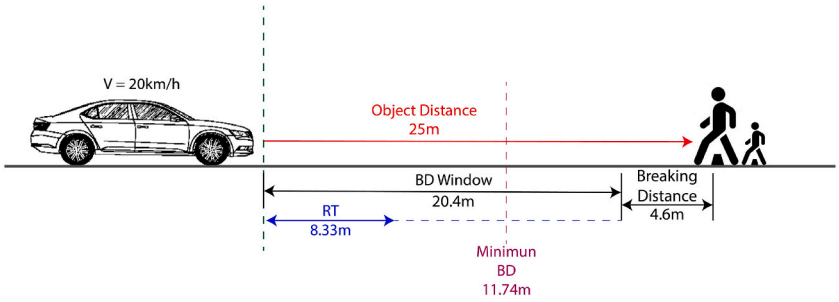



**Fig. 12.** Breaking distances of a vehicle traveling at 20 km/h. This figure shows a graphic demonstration of the vehicle breaking operation, where the driver has to take with precaution all the primary distances that have to be considered when breaking: Object distance, Braking Distance (BD), Reaction Time (RT) lost in meters and the minimum BD.

testing; a downside was encountered. When using a set of stereo cameras inside a vehicle, it is desirable that they have a wider field of view to accommodate the different possible collocations of the cameras and provide a complete view of the driver at all times to reduce the physical complexity of mounting the cameras on difficult sections of the vehicle, on the dashboard or the windshield. Also, the proposed SVS proved capable of working at real-time speeds, this was done by leveraging the implementation of working in parallel with two computers, one for the main system and the other one for the usage of a GPU for the DL algorithm. Finally, the proposed system also proved capable of working under different lighting conditions, providing promising results when working on indoor or outdoor scenarios, but for extreme scenarios like low and extreme lighting conditions, the system under performed, like in most camera-based systems, the lighting of a scene can over saturate the image and provide undesired results, and for this application, in a future iteration of the system the RGB cameras will be replaced by IR cameras, this to improve the performance under illumination variations.

## 5. Conclusion and future works

The proposed work showcases a suitable ADAS system that requires as input a pair of images to deliver a possible dangerous situation assessment in real-time, to decrease the percentage of street accidents. This assessment is possible by concurrently estimating the HPE estimation and classifying the zone on which the driver is viewing with a 99.77 % accuracy. The proposed system is able to classify the head movement through a continuous computation of the head pose with an average error of 0.87°, which is an important characteristic of our work; this characteristic provides an opening to future works where we can analyze the flow of the drivers' head to classify their behavior. Finally, the proposal to integrate the YOLO-R model with the template matching SoRA has enabled real-time object detection and distance determination processing, thus furnishing essential information to the alarm system. With a processing speed of 17 fps, this system can be effectively deployed in low mileage zones, as shown in Table 1.

In conclusion, our proposed work can provide driver feedback for a possible collision or can be paired with an AESB for automatic deceleration. For future research, the driver's emotion and behavior analysis will be considered to increase the robustness of the driver's road attention and automatic control configurations. Also, the RGB cameras used for HPE will be replaced by IR cameras to improve the system against illumination variation and to introduce low light application.

## CRediT authorship contribution statement

**Julio C. Rodriguez-Quiñonez:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Jonathan J. Sanchez-Castro:** Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization. **Oscar Real-Moreno:** Writing – review & editing, Software, Methodology. **Guillermo Galaviz:** Writing – review & editing, Supervision, Investigation. **Wendy Flores-Fuentes:** Supervision, Methodology. **Oleg Sergiyenko:** Supervision, Methodology. **Moises J. Castro-Toscano:** Supervision, Methodology. **Daniel Hernandez-Balbuena:** Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the Universidad Autónoma de Baja California, Universidad Autónoma de Sinaloa and CONAHCYT.

## References

[1] T.S. Combs, L.S. Sandt, M.P. Clamann, N.C. McDonald, Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection, Am. J. Prev. Med. 56 (2019) 1–7.
[2] N.H.T.S. Administration, et al., Motor Vehicle Crashes: Overview, Traffic Safety Facts: Research Note 2016, 2015, pp. 1–9, 2016.
[3] M. Yanagisawa, E.D. Swanson, W.G. Najm, et al., Target Crashes and Safety Benefits Estimation Methodology for Pedestrian Crash Avoidance/mitigation Systems, Department of Transportation. National Highway Traffic Safety, United States, 2014. Technical Report.
[4] S. Malla, C. Choi, I. Dwivedi, J.H. Choi, J. Li, Drama: joint risk localization and captioning in driving, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 1043–1052.
[5] B. Peng, D. Gao, M. Wang, Y. Zhang, 3d-stcnn: spatiotemporal convolutional neural network based on eeg 3d features for detecting driving fatigue, J. Data Sci. Intell. Sys. 2 (2024).
[6] W. Li, J. Huang, G. Xie, F. Karray, R. Li, A survey on vision-based driver distraction analysis, J. Syst. Architect. 121 (2021) 102319.
[7] S.C. Addanki, N. Jaswanth, R. Assfalg, H. Venkataraman, Analysis of traffic related factors and vehicle environment in monitoring driver's driveability, Int. J. Intell.Trans. Sys. Res. 18 (2020) 277–287.
[8] S. Jha, C. Busso, Estimation of driver's gaze region from head position and orientation using probabilistic confidence regions, IEEE Trans. Intell. Vhl. 8 (2022) 59–72.
[9] H. Liu, C. Zhang, Y. Deng, T. Liu, Z. Zhang, Y.-F. Li, Orientation cues-aware facial relationship representation for head pose estimation via transformer, IEEE Trans. Image Process. 32 (2023) 6289–6302.
[10] H. Liu, T. Liu, Z. Zhang, A.K. Sangaiah, B. Yang, Y. Li, Arhpe: asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction, IEEE Trans. Ind. Inf. 18 (2022) 7107–7117.
[11] H. Liu, S. Fang, Z. Zhang, D. Li, K. Lin, J. Wang, Mfdnet: collaborative poses perception and matrix Fisher distribution for head pose estimation, IEEE Trans. Multimed. 24 (2021) 2449–2460.
[12] G. Sikander, S. Anwar, A novel machine vision-based 3d facial action unit identification for fatigue detection, IEEE Trans. Intell. Transport. Syst. 22 (2020) 2730–2740.
[13] A. Asperti, D. Filippini, Deep learning for head pose estimation: a survey, SN Comput. Sci. 4 (2023) 349.
[14] B. Akrout, W. Mahdi, A novel approach for driver fatigue detection based on visual characteristics analysis, J. Ambient Intell. Hum. Comput. 14 (2023) 527–552.
[15] A.F. Abate, C. Bisogni, A. Castiglione, M. Nappi, Head pose estimation: an extensive survey on recent techniques and applications, Pattern Recogn. 127 (2022) 108591.
[16] W. Peng, H. Pan, H. Liu, Y. Sun, Ida-3d: instance-depth-aware 3d object detection from stereo vision for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13015–13024.
[17] M. Ding, Z. Zhang, X. Jiang, Y. Cao, Vision-based distance measurement in advanced driving assistance systems, Appl. Sci. 10 (2020) 7276.
[18] H. Liu, T. Liu, Y. Chen, Z. Zhang, Y.-F. Li, Ehpe: skeleton cues-based Gaussian coordinate encoding for efficient human pose estimation, IEEE Trans. Multimed. (2022) 1–12, https://doi.org/10.1109/TMM.2022.3197364.
[19] H. Liu, Y. Chen, W. Zhao, S. Zhang, Z. Zhang, Human pose recognition via adaptive distribution encoding for action perception in the selfregulated learning process, Infrared Phys. Technol. 114 (2021) 103660.
[20] T. Liu, H. Liu, B. Yang, Z. Zhang, Ldcnet: limb direction cues-aware network for flexible human pose estimation in industrial behavioral bio- metrics systems, IEEE Trans. Indus. Inf. 20 (2024) 8068–8078, https://doi.org/10.1109/TII.2023.3266366.
[21] H.V. Koay, J.H. Chuah, C.-O. Chow, Y.-L. Chang, Detecting and recognizing driver distraction through various data modality using ma- chine learning: a review, recent advances, simplified framework and open challenges (2014–2021), Eng. Appl. Artif. Intell. 115 (2022) 105309.
[22] H. Liu, X. Wang, W. Zhang, Z. Zhang, Y.-F. Li, Infrared head pose estimation with multi-scales feature fusion on the irhp database for human attention recognition, Neurocomputing 411 (2020) 510–520.
[23] S. Biswas, D. Chambers, W.D. Hairston, S. Bhattacharya, Head pose classification for passenger with cnn, Transport Eng. 11 (2023) 100157.
[24] C. Zhang, H. Liu, Y. Deng, B. Xie, Y. Li, Tokenhpe: learning orientation tokens for efficient head pose estimation via transformers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pat- Tern Recognition, 2023, pp. 8897–8906.
[25] J. Ju, H. Zheng, C. Li, X. Li, H. Liu, T. Liu, Agcnns: attention- guided convolutional neural networks for infrared head pose estimation in assisted driving system, Infrared Phys. Technol. 123 (2022) 104146.
[26] T. Hu, S. Jha, C. Busso, Temporal head pose estimation from point cloud in naturalistic driving conditions, IEEE Trans. Intell. Transport. Syst. 23 (2021) 8063–8076.
[27] J.D. Ortega, N. Kose, P. Cañas, M.-A. Chao, A. Unnervik, M. Nieto, O. Otaegui, L. Salgado, Dmd: a large-scale multi-modal driver monitoring dataset for attention and alertness analysis, in: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16, Springer, 2020, pp. 387–405.
[28] C. Bisogni, L. Cascone, M. Nappi, C. Pero, Iot-enabled biometric security: enhancing smart car safety with depth-based head pose estimation, ACM Trans. Multimed. Comput. Commun. Appl. 20 (2024) 1–24.
[29] N. Khairdoost, M. Shirpour, M.A. Bauer, S.S. Beauchemin, Real-time driver maneuver prediction using lstm, IEEE Trans. Intell. Vhl. 5 (2020) 714–724.
[30] J. Leng, Y. Liu, D. Du, T. Zhang, P. Quan, Robust obstacle detection and recognition for driver assistance systems, IEEE Trans. Intell. Transport. Syst. 21 (2019) 1560–1571.
[31] Y. Shi, Y. Guo, Z. Mi, X. Li, Stereo centernet-based 3d object detection for autonomous driving, Neurocomputing 471 (2022) 219–229.
[32] M.J. Castro-Toscano, J.C. Rodríguez-Quiñonez, O. Sergiyenko, W. Flores-Fuentes, L.R. Ramirez-Hernandez, D. Hernández-Balbuena, L. Lindner, R. Rasc'on, Novel sensing approaches for structural deformation monitoring and 3d measurements, IEEE Sensor. J. 21 (2020) 11318–11328.
[33] J.C. Rodríguez-Quiñonez, G. Trujillo-Hernández, W. Flores-Fuentes, O. Sergiyenko, J.E. Miranda-Vega, J.J. Sanchez-Castro, M.J. Castro-Toscano, O. Real-Moreno, Anthropometric stereo vision system for measuring foot arches angles in three dimensions, IEEE Trans. Instrum. Meas. 73 (2023) 1–11.
[34] O. Real-Moreno, J.C. Rodríguez-Quiñonez, W. Flores-Fuentes, O. Sergiyenko, J.E. Miranda-Vega, G. Trujillo-Hernández, D. Hern'andez-Balbuena, Camera calibration method through multivariate quadratic regression for depth estimation on a stereo vision system, Opt Laser. Eng. 174 (2024) 107932.
[35] G.G. Slabaugh, Computing Euler Angles from a Rotation Matrix, Re- Trieved on August 6, 1999, pp. 39–63.
[36] O. Real-Moreno, J.C. Rodríguez-Quiñonez, O. Sergiyenko, W. Flores- Fuentes, P. Mercorelli, J.A. Valdez-Rodríguez, G. Trujillo-Hernández, J.E. Miranda-Vega, Fast template match algorithm for spatial object detection using a stereo vision system for autonomous navigation, Measurement 220 (2023) 113299.
[37] E. Dall'Asta, R. Roncella, A comparison of semiglobal and local dense matching algorithms for surface reconstruction, the international archives of the photogrammetry, Rem. Sens. Spatial Inf. Sci. 40 (2014) 187–194.

[38] S. Oron, T. Dekel, T. Xue, W.T. Freeman, S. Avidan, Best-buddies similarity—robust template matching using mutual nearest neighbors, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2017) 1799–1813.
[39] Z. Yang, Z. Ren, Q. Shan, Q. Huang, Mvs2d: efficient multi-view stereo via attention-driven 2d convolutions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8574–8584.
[40] B. Ummenhofer, H. Zhou, J. Uhrig, N. Mayer, E. Ilg, A. Dosovitskiy, T. Brox, Demon: depth and motion network for learning monocular stereo, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5038–5047.
[41] T. Officials, A Policy on Geometric Design of Highways and Streets, AASHTO, 2011, 2011.