

Detecting Driver Drowsiness by Combining Viola-Jones Image Processing Method and Image Landmarks Detection

Sanaz Eskandari¹, Amir Masoud Rahimi^{1*}, Ehsan Ramezani-khansari²

¹ Department of Civil Engineering, Faculty of Engineering, University of Zanjan, Zanjan, Iran.

² Department of Transportation, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran.

Review History:

Received: May, 05, 2024

Revised: Dec. 14, 2024

Accepted: Dec. 20, 2024

Available Online: Apr. 06, 2025

Keywords:

Car Accidents

Fatigue

Viola-Jones

Drowsiness

Facial Tracking

ABSTRACT: Fatigue is one of the main causes of traffic crashes which leads to the death of many people on the roads every year. Several methods have been developed to detect the level of fatigue, one of which is intelligent fatigue detection based on driving history. This advantage leads to the detection of fatigue in the early stages and notifying before the crash. In this paper, a new method is applied to detect drowsiness from video frames using the YawDD database. The operation of the proposed system is based on the tracking of facial landmarks. The purpose of this paper is to detect drowsiness from facial expressions. By extracting and tracking key points, facial expressions such as normal or sleepy can be detected. First, the initial position of the face is determined and then the key points are extracted using Viola-Jones and SIFT algorithms, respectively. The results showed that the proposed method is capable of detecting the driver's drowsiness with good accuracy. The proposed method executed 47 frames per second in real-time with an accuracy of 94% and an error rate of 6%.

1- Introduction

Drowsiness while driving is a problem in all countries, especially industrialized ones, leading to significant consequences, fatalities, and economic costs. The causes of driving drowsiness are complex and may be related to chronic insufficient sleep, driving at inappropriate times in the circadian cycle, untreated sleep disorders, medical and psychiatric conditions, and the use of medications and sedatives [1]. Identifying and addressing these individual issues has shown improvements in driving performance and reductions in accidents in many cases. Driver fatigue causes hundreds of road accidents annually, and these accidents are associated with nearly a 50 percent higher likelihood of death or injury [2]. The diagnosis of fatigue is very complex, and it can be said that image processing is one of the best methods for its detection. Therefore, the main objective of this study is to utilize image processing techniques to detect fatigue levels in a driving simulator.

Jasina et al. studied and analyzed several techniques to obtain the best technique with the highest accuracy for detecting driver drowsiness. They proposed a method by utilizing Python, dlib, and OpenCV to create a real-time framework that employed a computer camera to monitor and process the driver's eyes and yawning. A camera was used to focus on the driver's eyes and yawning [3]. Thiruvallar

and Vimal examined two different algorithms based on Convolutional Neural Networks (CNN) for detecting drowsiness, and the results were compared accordingly. Thus, drowsiness detection based on the Viola-Jones and PERCLOS algorithms was tested using CNN, yielding satisfactory results. The Viola-Jones algorithm was superior to the other algorithm in terms of accuracy. However, both algorithms have their own specific importance in various aspects [4]. Given the significance of drowsiness and fatigue in safety discussions, many researchers have attempted to analyze the visual features of drivers using image processing to identify this phenomenon [5], [6].

2- Methodology

One of the methods for detecting drowsiness is the use of landmarks or key points on the face. By extracting and tracking these points, it is possible to identify the facial state of a person, such as normal or drowsy.

The proposed procedure is as follows: 1- After receiving the video frames, the face detection process in the first frame is initially performed using the Viola-Jones algorithm. 2- After face detection, key points are extracted from the facial area using the SIFT algorithm. 3- Considering that the detection and tracking of points in subsequent stages require the alignment of facial images and the transformation of images

*Corresponding author's email: amrahimi@znu.ac.ir

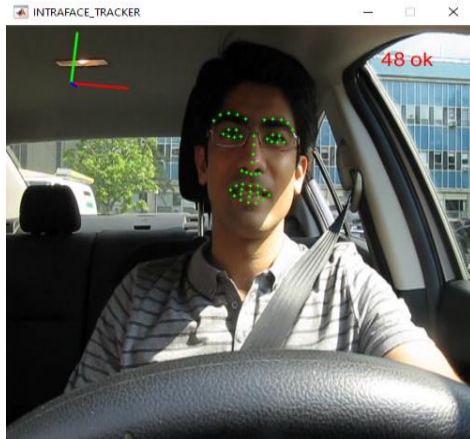


Fig. 1. Increasing the processing speed of the proposed algorithm

to a frontal view, the extracted key points are used for aligning the facial images. At this stage, the extracted key points are mapped to points using the Affine transformation so that if the person's head is angled, their image is transformed to a frontal view. 4- Another feature of the Viola-Jones algorithm is the extraction of regions related to the lips, mouth, and eyes. After detecting the area related to the person's mouth, to increase processing speed, only the key points related to the upper and lower lips are tracked. 5- To detect drowsiness, the Euclidean distance between the key points of the upper and lower lips is calculated. 6- If the calculated Euclidean distance in the previous step is less than the defined threshold, the driver's state is considered normal and the tracking of key points continues; otherwise, it indicates driver drowsiness and warning signals are activated.

3- Results and Discussion

Over time, the execution speed of the proposed algorithm (averaging 48 frames per second in the sample video) increased, indicating the real-time execution of the proposed method. Considering that in the subsequent frames, the process of face detection and key point extraction has been removed and only the tracking of facial landmarks was performed, this led to an increase in processing speed (Fig 1).

To demonstrate the accuracy of tracking and the performance of the tracking algorithm in consecutive frames, the mouth position was extracted using the Viola-Jones algorithm. The positions obtained by the Viola-Jones algorithm are considered as a reference. Subsequently, the trajectory obtained from tracking the mouth position in consecutive frames was compared with the actual output. As can be seen, the tracking algorithm can perform tracking operations with good accuracy (Fig2).

The results of comparing the proposed method of this research with the established methods for detecting drowsiness that have utilized the YawDD database are presented in Table 1.

The errors that occur are due to the consideration of a

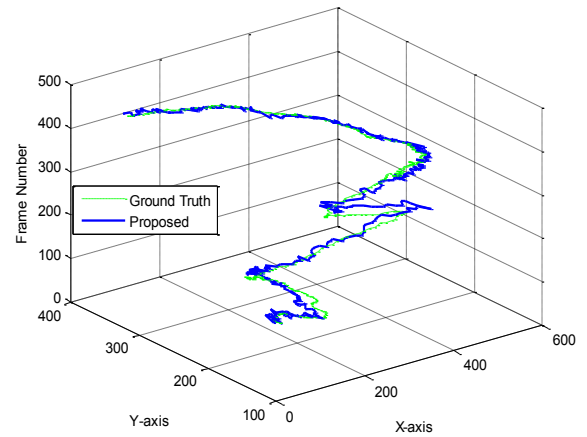


Fig. 2. The trajectory of the proposed algorithm is presented in comparison with the actual output.

Table 1. Result comparison

Algorithm	Detection accuracy	Error
Diagnosis of yawning through descriptive analysis of changes [7]	83	17
Monitoring of driver drowsiness under control of the Descent Method [8]	86	14
Detection of driver drowsiness using a hierarchical temporal deep learning network [9]	92	8
Detection of driver drowsiness by a set of convolutional neural [10] networks	92	8
The proposed method of this research	94	6

manual threshold, “the distance between the upper and lower lips,” during the detection of drowsiness. A person may open their mouth more than usual while speaking or may open their mouth less than usual while yawning, given that the distance between the lips is used to detect drowsiness and fatigue. The threshold considered for detecting drowsiness may encounter errors. It is suggested that the threshold be determined by an intelligent system to reduce the detection error of drowsiness.

4- Conclusion

Considering that the facial area in the image must first be identified to detect a person's drowsiness, the Viola-Jones algorithm is utilized for face detection in this study. Subsequently, facial landmarks will be extracted using the SIFT algorithm. To enhance processing speed, instead of

tracking the entire face, tracking of extracted key points was employed. Finally, the Euclidean distance between the key points of the upper and lower lips was used to detect driver drowsiness. The total algorithms used in this research indicate that the proposed method is capable of accurately detecting driver drowsiness with a higher processing speed.

References

- [1] R. W. Wohleber, G. Matthews, G. J. Funke, and J. Lin, "Considerations in Physiological Metric Selection for Online Detection of Operator State: A Case Study," in *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*, vol. 9743, D. D. Schmorrow and C. M. Fidopiastis, Eds., in *Lecture Notes in Computer Science*, vol. 9743, Cham: Springer International Publishing, 2016, pp. 428–439. doi: 10.1007/978-3-319-39955-3_40.
- [2] R. E. P. Soares, "Driver monitoring systems of fatigue based on eye tracking," PhD Thesis, 2017. Accessed: Apr. 24, 2024. [Online]. Available: <https://repositorium.sdum.uminho.pt/handle/1822/54750>
- [3] J. Jose, J. S. Vimali, P. Ajitha, S. Gowri, A. Sivasangari, and B. Jinila, "Drowsiness Detection System for Drivers Using Image Processing Technique," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Jun. 2021, pp. 1527–1530. doi: 10.1109/ICOEI51242.2021.9452864.
- [4] V. Naren Thiruvalar and E. Vimal, "A comparative analysis on driver drowsiness detection using CNN," *Int. J. Nonlinear Anal. Appl.*, Dec. 2021, doi: 10.22075/ijnaa.2021.5894.
- [5] Y. Albadawi, A. AlRedhaei, and M. Takruri, "Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features," *J. Imaging*, vol. 9, no. 5, Art. no. 5, Apr. 2023, doi: 10.3390/jimaging9050091.
- [6] K. Balasubramanian, "Drowsiness detection and safety monitoring using image processing," *Int. J. Appl. Sci. Eng.*, vol. 19, no. 3, pp. 1–9, 2022.
- [7] B. Akrouit and W. Mahdi, "Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness," in *2016 International Image Processing, Applications, and Systems (IPAS), IEEE, 2016*, pp. 1–5. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7880127/>
- [8] N. Neshov and A. Manolova, "Drowsiness monitoring in real-time based on supervised descent method," in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IEEE, 2017*, pp. 660–663. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8095173/>
- [9] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network," in *Computer Vision – ACCV 2016 Workshops*, vol. 10118, C.-S. Chen, J. Lu, and K.-K. Ma, Eds., in *Lecture Notes in Computer Science*, vol. 10118, Cham: Springer International Publishing, 2017, pp. 117–133. doi: 10.1007/978-3-319-54526-4_9.
- [10] R. M. Salman, M. Rashid, R. Roy, M. M. Ahsan, and Z. Siddique, "Driver Drowsiness Detection Using Ensemble Convolutional Neural Networks on YawDD," Dec. 19, 2021, arXiv: arXiv:2112.10298. Accessed: Apr. 24, 2024. [Online]. Available: <http://arxiv.org/abs/2112.10298>.



تشخیص خواب آلودگی راننده با استفاده از تلفیق روش پردازش تصویر ویولا-جونز و آشکارسازی لندمارک های تصویر

ساناز اسکندری^۱، امیرمسعود رحیمی^{۱*}، احسان رضانی خوانساری^۲

۱- گروه عمران، دانشکده مهندسی، دانشگاه زنجان، زنجان، ایران.
۲- دانشکده فنی و مهندسی، دانشگاه بین المللی امام خمینی، قزوین، ایران.

تاریخچه داوری:

دریافت: ۱۴۰۳/۰۲/۱۶
بازنگری: ۱۴۰۳/۰۹/۲۴
پذیرش: ۱۴۰۳/۰۹/۳۰
ارائه آنلاین: ۱۴۰۴/۰۱/۱۷

کلمات کلیدی:

خستگی
ردیابی چهره
ویولا جونز
خواب آلودگی
پردازش تصویر

خلاصه: خستگی یکی از دلایل اصلی تصادفات رانندگی است که سالانه منجر به مرگ و میر بسیاری از مردم در جاده ها می شود. روش های متعددی برای تشخیص میزان خستگی ایجاد شده است. یکی از روش ها، تشخیص خستگی هوشمند و استفاده از پارامترهای مختلف (ابزارها) در زمان طولانی و با استفاده از سابقه رانندگی است. این مزیت منجر به تشخیص خستگی در مراحل اولیه و فعال سازی زنگ هشدار قبل از وقوع تصادف رانندگی می شود. هدف این تحقیق تشخیص خواب آلودگی از روی حالت چهره می باشد. با استخراج و ردیابی نقاط کلیدی می توان حالت چهره شخص مانند حالت عادی یا خواب آلوده را تشخیص داد. در این مقاله روش جدید برای تشخیص خواب آلودگی از روی فریم های ویدیویی با استفاده از پایگاه داده YawDD می باشد. نحوه عملکرد سیستم پیشنهادی براساس ردیابی لندمارک های چهره استخراج شده از فریم های ویدیویی است. در ابتدا موقعیت اولیه چهره با استفاده از الگوریتم ویولا جونز مشخص می شود، سپس نقاط کلیدی توسط الگوریتم سیفت استخراج می شوند. این روش قادر است با دقت خوبی عملیات تشخیص خواب آلودگی راننده را انجام دهد. روش پیشنهادی قادر است با دقت خوبی عملیات تشخیص خواب آلودگی راننده را انجام دهد. از دیگر مزایای روش پیشنهادی می توان به اجرای بالاترین ۴۷ فریم بر ثانیه و دقت ۹۴ درصد و خطای ۶٪ بر روی پایگاه داده YawDD اشاره کرد.

۱- مقدمه

خواب آلودگی در هنگام رانندگی یک مشکل در همه کشورها بویژه کشورهای صنعتی است و منجر به عوارض، مرگ و میر و هزینه های اقتصادی قابل توجهی می شود. علل خواب آلودگی رانندگی پیچیده است و ممکن است به خواب ناکافی مزمن، رانندگی در زمان های نامناسب در چرخه شبانه روزی، اختلالات خواب درمان نشده، اختلالات پزشکی و روانپزشکی و استفاده از داروها و مواد آرام بخش مرتبط باشد [۱]. شناسایی و درمان این مشکلات فردی در بسیاری از موارد بهبود عملکرد رانندگی و کاهش تصادفات را نشان داده است [۲]. خستگی راننده سالانه باعث ایجاد صدها تصادف جاده ای می شود و این تصادفات تقریباً ۵۰ درصد بیشتر احتمال مرگ یا جراحت را به همراه دارند [۳].

در هنگام خستگی راننده، پلیس هیچ نشانه ای از تمرزگیری یا عکس العمل راننده نمی یابد، بنابراین احتمال برخورد سریعتر خودرو وجود

دارد. از جمله: کمبود خواب، ریتم شبانه روزی و تأثیر مستقیم خستگی کار بر عوامل شناختی راننده [۴]. خستگی راننده ناشی از خواب یا اختلالات که در حدود ۲۰ درصد از تصادفات رخ می دهد، یکی از عوامل اصلی آسیب های جاده ای است [۳]. از این رو محققان مختلفی به بررسی خستگی راننده و چگونگی تشخیص آن پرداخته اند.

ردیابی چشم یک روش اساسی برای جمع آوری اطلاعات در مورد خستگی رانندگی است. محققان این سیستم را برای ارزیابی خستگی راننده از طریق ردیابی رفتار ارائه کرده اند [۵] و [۶]. توزیع اندازه دامنه حرکت مردمک تحت شرایط عادی رانندگی و خستگی متفاوت است [۶].

تشخیص زود هنگام شروع خستگی را می توان با تجزیه و تحلیل فرکانس طیفی الکتروکاردیوگراف (ECG) [۷]، الکترومیوگرافی (Electromyography) [۸] و [۹] و تجزیه و تحلیل مدت زمان تثبیت چشم انجام داد. سابقه خواب راننده که با روشهای موشن گرافی ارزیابی می شود نیز ممکن است با توسعه الگوریتم ها به تشخیص خستگی کمک

* نویسنده عهده دار مکاتبات: amrahimi@znu.ac.ir



کند [۱۰].

را تعیین کند. از این رو با استفاده از ترکیب چند روش پردازش تصویر و در نظر گرفتن معیارهای مختلف به نتایج بهتری دست یافته شد که در ادامه بیان شده است.

۲- روش تحقیق

در این تحقیق تشخیص خواب‌آلودگی از روی حالات چهره صورت گرفته است. یکی از روش‌های تشخیص خواب‌آلودگی استفاده از لندمارک‌ها یا نقاط کلیدی چهره می‌باشد. در واقع با استخراج و ردیابی این نقاط می‌توان حالت چهره شخص مانند حالت عادی یا خواب‌آلوده را تشخیص داد.

با توجه به اینکه در فریم‌های ویدئویی امکان تغییر موقعیت سر از طریق چرخش وجود دارد، پس از یافتن نقاط کلیدی چهره ابتدا باید همترازی و اصلاح چرخش سر را انجام داد. سپس عملیات یافتن موقعیت دقیق لندمارک‌های چهره و ردیابی این نقاط انجام می‌شود. برای این منظور، در ابتدا موقعیت اولیه چهره با استفاده از الگوریتم ویولا-جونز (Viola Jones) مشخص می‌شود، سپس نقاط کلیدی توسط الگوریتم سیفت (SIFT) استخراج می‌شوند. نقاط کلیدی استخراج شده توسط الگوریتم سیفت به صورت دستی برای مشخص کردن نقاط کلیدی چهره مورد استفاده قرار می‌گیرد. در واقع از نقاط کلیدی استخراج شده توسط الگوریتم سیفت برای آموزش الگوریتم استفاده می‌شود. با توجه به اینکه امکان چرخش سر وجود دارد، از نقاط کلیدی سیفت برای همترازی تصاویر چهره و اصلاح چرخش سر استفاده می‌شود.

در این تحقیق روش جدیدی برای تشخیص خواب‌آلودگی از روی فریم‌های ویدئویی به دست آمده از دوربین نصب شده مقابل راننده ارائه شده است. با توجه به اهمیت سرعت پردازش، به جای ردیابی تمامی نقاط مربوط به صورت، تنها لندمارک‌های چهره که برای پردازش‌های بعدی مورد نیاز است، برای ردیابی استفاده می‌شود. به همین دلیل سرعت پردازش الگوریتم پیشنهادی مناسب می‌باشد.

۲-۱- آشکارسازی چهره با استفاده از الگوریتم ویولا-جونز

روش ویولا-جونز براساس جستجوی تصویر ورودی به وسیله پنجره‌های کوچک‌تر قادر به شناسایی ویژگی‌ها می‌باشد. به منظور آشکارسازی چهره در مقیاس‌های مختلف این پنجره‌ها مقیاس‌بندی می‌شوند. ویولا جونز یک آشکارساز مقاوم در برابر مقیاس ارائه داده‌اند که چندین بار روی تصویر ورودی اعمال می‌شود و هر بار اندازه پنجره جستجو تغییر می‌کند. به منظور مقاوم بودن در برابر مقیاس، آشکارساز نیاز به تعدادی محاسبات یکسان

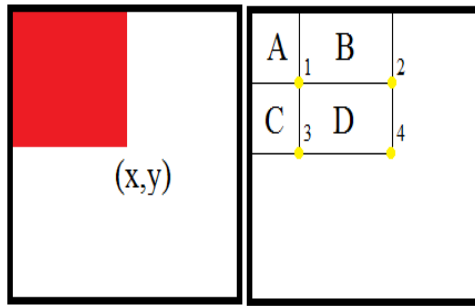
آنانثی^۱ و همکاران روش شناسایی نسبت ابعاد چشم (I-EAR) را برای شناسایی بسته شدن چشم راننده وسیله نقلیه پیشنهاد کردند. اگر چشم راننده برای حداقل ۴۰ فریم بسته باشد یا نسبت پلک زدن چشم کمتر از ۸ در دقیقه باشد، سیستم پیشنهادی مشخص می‌کند که راننده وسیله نقلیه خواب‌آلود است. در نتیجه، شروع به هشدار صدا می‌کند که راننده وسیله نقلیه خواب‌آلود است [۱۱]. لیتون حسین و همکاران مطالعه‌ای مشابه را در کشور بنگلادش انجام دادند که هدف آن کشف خواب‌آلودگی با استفاده از نسبت ابعاد چشم بود [۱۲].

جسینا و همکاران چندین تکنیک را مطالعه و تجزیه و تحلیل کردند تا بهترین تکنیک با بالاترین دقت برای تشخیص خواب‌آلودگی راننده به دست آید. روش پیشنهادی از Python، dlib و OpenCV برای ساخت یک چارچوب بلادرنگ استفاده می‌کند که از یک دوربین کامپیوتری برای نظارت و پردازش چشم و خمیازه راننده استفاده می‌کند. از یک دوربین استفاده می‌شد تا روی چشم راننده و خمیازه کشیدن متمرکز شود [۱۳]. تیروولار و ویمال^۲ دو الگوریتم مختلف مبتنی بر شبکه عصبی کانولوشن (CNN) را برای تشخیص خواب‌آلودگی بررسی کردند و نتایج به ترتیب با هم مقایسه شدند. بدین ترتیب تشخیص خواب‌آلودگی بر اساس الگوریتم‌های Viola-Jones و PERCLOS با استفاده از CNN مورد آزمایش قرار گرفت و نتایج رضایت بخشی به دست آمد. الگوریتم ویولا جونز بر اساس جنبه دقت، از الگوریتم دیگر بیشتر بود. اما، هر دو الگوریتم در جنبه‌های مختلف اهمیت خاص خود را دارند. [۱۴]. با توجه به اهمیت موضوع خواب‌آلودگی و خستگی در بحث ایمنی محققین بسیاری تلاش کرده‌اند ویژگی‌های بصری راننده را با استفاده از پردازش تصویر تحلیل کرده و این پدیده را شناسایی کنند [۱۵] و [۱۶].

با این حال، تشخیص خستگی بسیار پیچیده بوده و می‌توان گفت که پردازش تصویر یکی از بهترین روش‌های تشخیص آن می‌باشد. از این رو، هدف اصلی از این مطالعه استفاده از تکنیک‌های پردازش تصویر برای تشخیص سطوح خستگی در شبیه‌ساز رانندگی است.

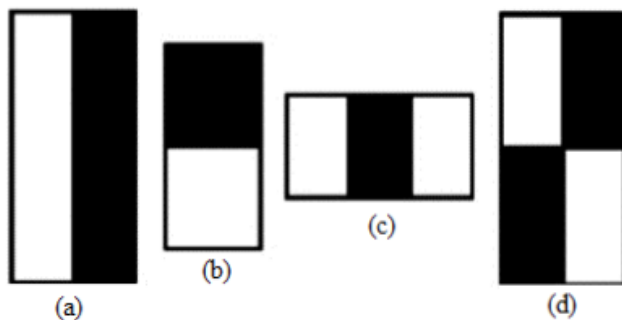
این تحقیق به منظور دستیابی به این هدف، از پارامترهای شناختی متنوعی استفاده می‌کند که شامل فاصله بین لب‌های بالا و پایین، فاصله بین پلک‌ها، حرکات سر و فاصله بین ابروها و چشم‌ها می‌باشند. با تحلیل این پارامترها، تلاش می‌شود تا شاخص‌های خواب‌آلودگی و خستگی رانندگان

1. Ananthi et. al.
2. Thiruvalar and Vimal



شکل ۱. تشکیل انتگرال تصویر در الگوریتم Viola – Jones

Fig. 1. Integral Image Formation In Viola-Jones Algorithm



شکل ۲. مستطیل‌های ویژگی مورد استفاده برای آشکارسازی [۱۸]

Fig. 2. Feature rectangles used for detection

$$(2+3) - 1 + 4 \text{ محاسبه شود [۱۸].}$$

در روش ویولا-جونز برای به دست آوردن ویژگی‌ها از اعمال الگوریتم در پنجره‌های کوچک (sub-window) به منظور آشکارسازی چهره استفاده می‌شود. این ویژگی‌ها شامل دو یا تعداد بیش‌تری مستطیل می‌باشد. هر ویژگی از طریق اعمال تفاضل روی جمع مستطیل‌های سفید از جمع مستطیل‌های سیاه به دست می‌آید. مستطیل‌های مورد استفاده برای ویژگی‌های گوناگون در شکل ۲ نشان داده شده است.

با استفاده از انتگرال تصویر می‌توان ویژگی نشان داده شده در شکل ۳ را تنها با هفت عمل جمع و تفریق محاسبه کرد. مقدار این ویژگی برابر است با $S_1 - S_2$ که S_1 و S_2 با استفاده از معادله ۱ محاسبه می‌شوند.

صرف نظر از اندازه تصویر دارد. ساختار الگوریتم ویولا-جونز مبتنی بر اتصال آبشاری^۱ یک مجموعه از آشکارسازها می‌باشد. گام اول از تعدادی آشکارساز ساده تشکیل شده است که تنها پنجره‌هایی را که شامل چهره نمی‌باشند، حذف می‌کنند. در گام‌های بعدی پیچیدگی آشکارسازها به منظور تجزیه و تحلیل ویژگی‌ها با جزئیات بیش‌تر افزایش می‌یابد. یک چهره تنها در صورتی که توسط کل مجموعه آبشاری مشاهده شود، آشکارسازی خواهد شد. این آشکارسازها از انتگرال تصویر و ویژگی‌های^۲ تشکیل شده‌اند [۱۷].

الگوریتم ویولا-جونز (Viola-Jones) به دلیل سرعت بالا و ساختار ساده خود برای تشخیص چهره در زمان واقعی، گزینه‌ای بسیار مناسب برای این پروژه بوده است. یکی از ویژگی‌های اصلی این الگوریتم، سرعت بسیار بالای آن در پردازش تصاویر است که این امکان را فراهم می‌کند تا سیستم بتواند در شرایط بلادرنگ (real-time) به درستی کار کند. علاوه بر این، ویولا-جونز از ویژگی‌های Haar-like برای تشخیص چهره استفاده می‌کند که ساده و در عین حال مؤثر هستند. این الگوریتم به دلیل ساختار آبشاری خود، می‌تواند به طور پیوسته چهره را تشخیص دهد. اگرچه الگوریتم‌های دیگری مانند HOG و CNN نیز برای تشخیص چهره استفاده می‌شوند، HOG سرعت کمتری دارد و شبکه‌های عصبی کانولوشنی (CNN) نیاز به منابع محاسباتی بیشتری دارند. در مقایسه با سایر الگوریتم‌ها مانند Dlib، که دقت بالاتری در تشخیص نقاط کلیدی چهره دارد، در نهایت الگوریتم ویولا-جونز به دلیل سادگی در پیاده‌سازی و مقاومت در برابر تغییرات شرایط برای این تحقیق انتخاب شد [۲۵]. از آنجا که تحقیق حاضر هم نیاز به تشخیص چهره راننده دارد و براساس جمع‌بندی نویسندگان از مزایا و معایب الگوریتم‌های مختلف، الگوریتم ویولا-جونز انتخاب گردید.

۲-۲- انتگرال تصویر

برای محاسبه ویژگی‌ها در زمان ثابت از انتگرال تصویر استفاده می‌شود. گام اول در این الگوریتم تبدیل تصویر ورودی به انتگرال تصویر می‌باشد. انتگرال تصویر برابر جمع تمام پیکسل‌های بالا و سمت چپ پیکسل مورد نظر می‌باشد. با استفاده از این روش، جمع تمامی پیکسل‌های درون هر مستطیل داده شده فقط توسط چهار مقدار محاسبه می‌شود. در شکل ۱ مقدار انتگرال تصویر در موقعیت ۱ برابر جمع تمام پیکسل‌های درون مستطیل A ، در موقعیت ۲ برابر $A + B$ ، در موقعیت ۳ برابر $A + C$ و در موقعیت ۴ برابر $A + B + C + D$ می‌باشد. بنابراین جمع درون D می‌تواند به صورت

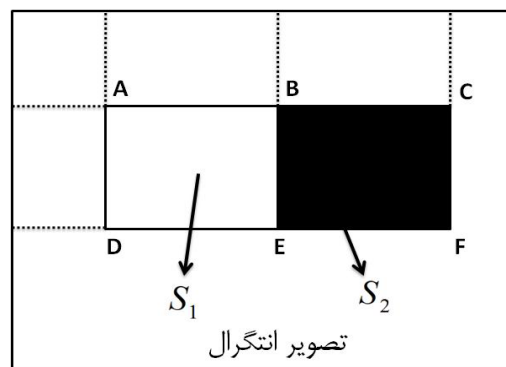
1. cascade
2. Haar features

۲-۳- طبقه بند مبتنی بر بوستینگ

دسته‌بندی که در ویولا-جونز استفاده شده است از الگوریتم Ada-Boost برای انتخاب ویژگی استفاده می‌کند که از نظر محاسباتی کارآمد می‌باشد. AdaBoost یک الگوریتم تقویتی در زمینه یادگیری ماشین می‌باشد که یک دسته‌بند قوی را از طریق ترکیب وزن‌داری از دسته‌بندهای ضعیف ایجاد می‌کند.

الگوریتم آشکارسازی چهره ویولا-جونز چندین بار با پنجره‌هایی با ابعاد مختلف روی تصویر ورودی اعمال می‌شود. آشکارساز، ناحیه‌هایی در تصویر را که شامل چهره نمی‌باشد شناسایی کرده و آن‌ها را از فرآیند آشکارسازی حذف می‌کند. برای حذف ناحیه‌هایی که شامل چهره نمی‌باشد، الگوریتم ویولا-جونز از خاصیت آبشاری استفاده می‌کند. در این روش پنجره‌های کوچکی که درصدی از چهره را دارا می‌باشند به مرحله بعدی ارسال می‌شوند و پنجره‌هایی که شامل چهره نباشند حذف خواهد شد. گام نهایی در فرآیند آبشاری باید درصد بالایی از وجود اشیاء چهره را شامل شود. فرآیند آبشاری مورد استفاده در الگوریتم ویولا-جونز برای پذیرش یک زیر پنجره به عنوان ناحیه‌ای از چهره به صورت شکل ۴ می‌باشد.

در فرآیند آموزش، این الگوریتم به صورت باناظر (Supervised) عمل می‌کند. یعنی برچسب تصاویر برای آموزش مشخص می‌باشد (تصاویر چهره و غیر چهره). در فاز آموزش در هر مرحله تعدادی ویژگی مستطیلی هار انتخاب شده و روی تصویر اعمال می‌شود، بنا به تشخیص درست یا غلط



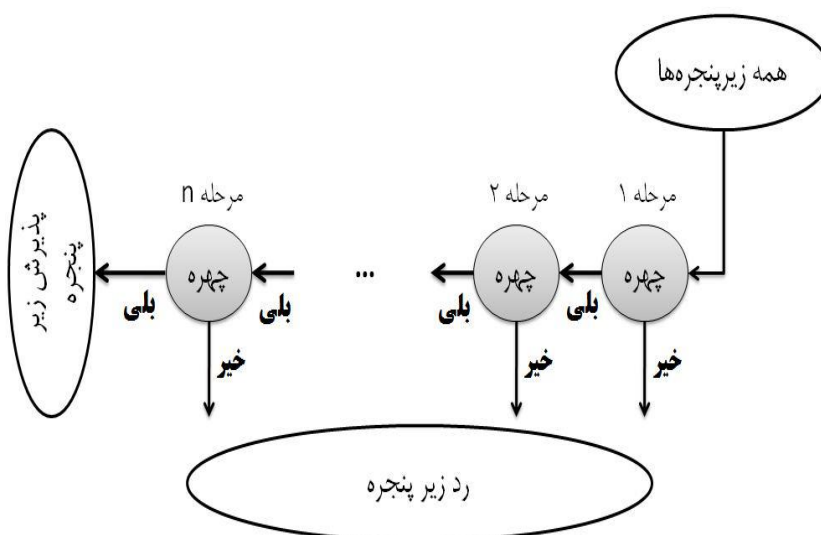
شکل ۳. محاسبه یک ویژگی مستطیلی با استفاده از انتگرال تصویر [۱۸]

Fig. 3. A Rectangular Feature Calculation Using the Integral Image

$$\begin{cases} S_1 = E - B - D + A \\ S_2 = F - C - E + B \end{cases} \quad (1)$$

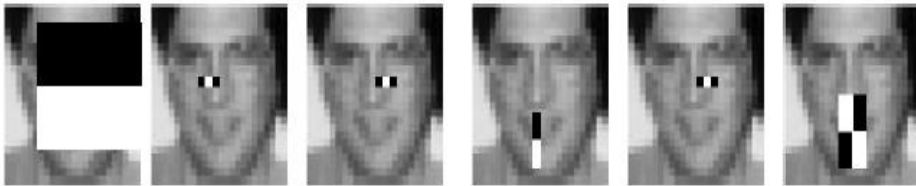
که انتگرال تصویر نیز طبق معادله ۲ قابل محاسبه است:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2)$$



شکل ۴. طبقه بند آبشاری

Fig. 4. Cascade Classifier

گام اول		۳ ویژگی مستطیلی
گام دوم		۱۰ ویژگی مستطیلی
گام ۲۱		۲۰۶ ویژگی مستطیلی

شکل ۵. اعمال فیلتر با ابعاد مختلف روی تصاویر در فرآیند آبشاری

Fig. 5. Applying Filters with Various Sizes to Images in Cascade Process

۲-۴- آشکارسازی نقاط کلیدی با استفاده از الگوریتم SIFT

با استفاده از مفاهیم هرم تصویری (Image Pyramid) و فضای مقیاس، آشکار کردن نقاط کمینه و بیشینه در فضای مقیاس انجام می‌پذیرد. به‌وسیله هرم تصویری یک تصویر رگومی در قدرت تفکیک‌های متفاوت نمایش داده می‌شود. با توجه به مفهوم فضای مقیاس که در آن تصویر در مقیاس‌های مختلف ساخته می‌شود، می‌توان مطابق شکل ۶ ابتدا تصویر را در چند مقیاس مختلف تشکیل داد و سپس برای هر مقیاس تصویری، پنج تصویر با انحراف معیار مختلف به‌وجود آورد که در نتیجه به تعداد مقیاس‌های مختلف دسته‌های پنج تصویری ایجاد می‌شود. در هر اکتاو اندازه تصویر نصف می‌شود. با توجه به این که از خصوصیات اصلی تصاویر هرمی کاهش نویز تصویر در اثر نرم کردن در قدرت تفکیک‌های پایین‌تر می‌باشد، لذا در این روش از فیلتر گوسین (Gaussian) استفاده گردیده است و دسته‌های پنج‌تایی از طریق ضرب کرنل گوسین در تصویر اصلی با انحراف معیارهای متفاوت به‌وجود آمده است. سپس همان‌گونه هر دو تصویر مجاور در داخل دسته‌های پنج‌تایی با اختلاف تفاضلی بر اساس جبر ماتریسی از هم کم می‌شود و تصاویر جدید بر اساس فیلتر اختلاف تفاضلی ایجاد می‌گردد. این مشتق‌گیری ثنائیه در واقع جهت پیدا کردن نقاط کمینه

هر کدام از دسته‌بندی‌های تشکیل دهنده این فرآیند آبشاری یک وزن به آنها تعلق خواهد گرفت.

در فاز تست در هر گام از فرآیند آبشاری تعدادی مستطیل ویژگی انتخاب شده و با زیر پنجره‌های استخراج شده از تصویر اصلی کانالو (Convolution) می‌شود. در صورتیکه در هر مرحله از این فرآیند مشخص شود که زیر پنجره استخراج شده مربوط به چهره نمی‌باشد از فرآیند آزمایش خارج خواهد شد، در غیر این صورت به مرحله بعدی ارسال می‌شود تا با ویژگی‌های بیشتری مورد بررسی قرار بگیرد. طی این فرآیند زیر پنجره‌ای به عنوان ناحیه چهره مورد پذیرش واقع خواهد شد که توسط تمامی مراحل آبشاری مورد تایید باشد. نمونه‌ای از گام‌های مختلف در فرآیند آبشاری برای شناسایی زیر پنجره‌های مربوط به چهره در شکل ۵ نشان داده شده است [۱۷].

در فاز تست بسیاری از زیر پنجره‌ها رد می‌شوند چون بیشتر ناحیه تصویر مربوط به پس زمینه خواهد بود. در صورتی که یک الگو از تصویر در مراحل ابتدایی به عنوان چهره شناسایی شود، به مراحل بعدی ارسال می‌شود تا پردازش‌های بیشتری روی آن انجام شود. همانطور که در شکل بالا مشخص شده است. در گام‌های بالاتر تعداد ویژگی‌ها بیشتر شده و محاسبات پیچیده‌تر خواهد شد [۱۷].

الگوریتم‌های پیچیده‌تر مانند CNN به منابع پردازشی بالاتری نیاز دارند. این در حالی است که SIFT تعادل مناسبی بین دقت و سرعت ایجاد می‌کند و بدون نیاز به منابع پردازشی سنگین، برای این پروژه بهینه است.

۲-۵- انتخاب نقاط کلیدی از بین نقاط کمینه و بیشینه اصلی

در این مرحله مقادیر شدت روشنایی سطوح خاکستری هر پیکسل با ۸ پیکسل مجاور آن مقایسه می‌گردد و همچنین با ۹ پیکسل در تصاویر مجاور بالایی و پایینی (که از لحاظ σ با هم اختلاف داشتند و از تابع اختلاف تفاضلی گوسین حاصل شده بود) مقایسه می‌گردد، اگر مقدار این پیکسل از تمام ۲۶ پیکسل همسایه بیش‌تر یا کم‌تر بود این نقطه یک عارضه کاندیدا بوده و به عنوان نقطه کلیدی انتخاب می‌شود.

۲-۶- تعیین موقعیت مکانی هر نقطه و توجیه آن

در این مرحله به طور دقیق موقعیت هر نقطه اصلی از لحاظ مختصات تعیین می‌گردد و نقاطی که دارای کنتراست کم می‌باشند حذف می‌شوند، سپس مقدار شیب (ضریب زاویه‌ای) و همچنین زاویه شیب هر نقطه اصلی به صورت معادلات ۶ و ۷ تعیین می‌گردد.

$$m(x, y) = \sqrt{\frac{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}{2}} \quad (6)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right) \quad (7)$$

در نتیجه برای تمام نقاط اصلی ماتریس $n \times 4$ ، عبارت ۸ تشکیل می‌شود. معادله ۸ موقعیت نقاط کلیدی و مقدار و زاویه‌ی شیب را برای هر نقطه مشخص می‌کند. در مرحله بعدی برای هر نقطه کلیدی مشخص شده توصیف‌گری متناظر با آن به‌دست می‌آید.

$$\begin{bmatrix} x_1 & y_1 & m_1 & \theta_1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & m_n & \theta_n \end{bmatrix}_{n \times 4} \quad (8)$$

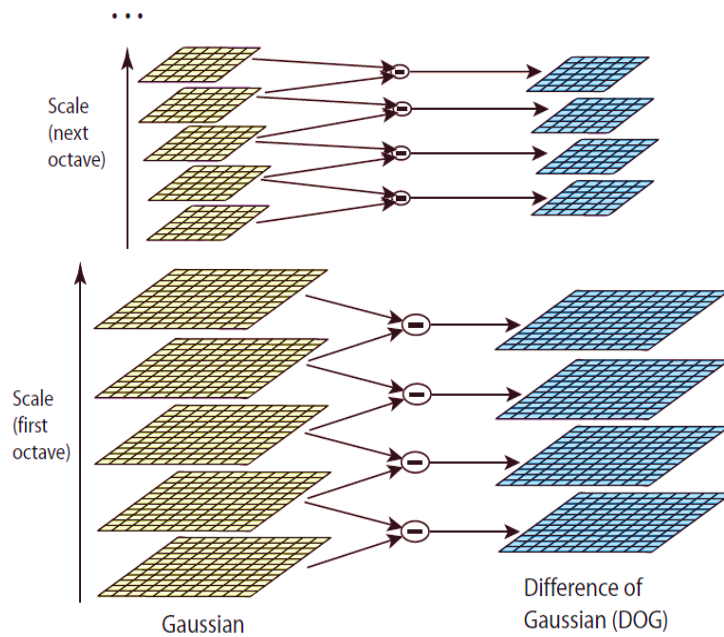
و بیشینه در فضای مقیاس انجام می‌شود. در نتیجه مقادیر کمینه و بیشینه در تصاویر به طور خاص در مقیاس‌های مختلف ایجاد شده و در این مرحله عارضه‌ها مستقل از مقیاس می‌گردند. در معادلات ۳، ۴ و ۵ نحوه انجام آورده شده است، که در آن I تصویر اصلی می‌باشد و G تابع کرنل گوسی می‌باشد که با کانولوشن، تصویر نرم شده L را تولید می‌نماید و سپس با کم کردن دو تصویر حاصل شده L در هر دسته پنج‌تایی با تصویر مجاورش، تصویر D بر اساس اختلاف تفاضلی گوسی‌ها ایجاد می‌گردد. به تابع $L(x, y, \sigma)$ فضای مقیاس یک تصویر نیز می‌گویند که از طریق کانولوشن تصویر با یک فیلتر گوسی به‌دست می‌آید [۱۹]

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

$$L(x, y, k\sigma) = G(x, y, \sigma) * I(x, y) \quad (4)$$

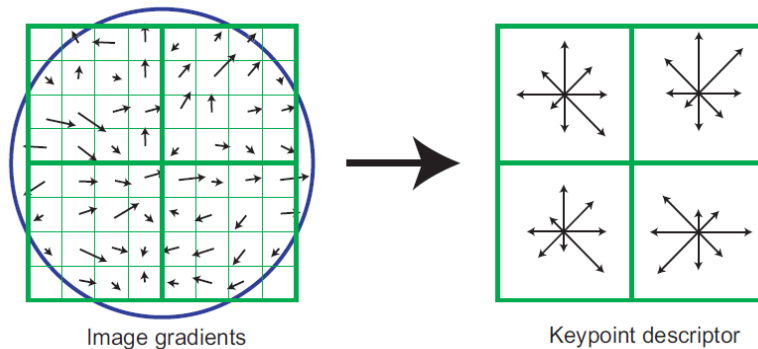
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{x^2 + y^2}{2\sigma^2} \right) \quad (5)$$

الگوریتم SIFT به دلیل توانایی منحصر به فرد خود در تشخیص و دنبال کردن ویژگی‌های محلی در تصاویر دو بعدی، انتخاب مناسبی برای پروژه‌هایی نظیر تشخیص خمیازه از روی لب‌های راننده است. دلیل اصلی این انتخاب این است که هرچند خمیازه و حرکات لب به طور واقعی سه بعدی هستند، داده‌های ورودی به سیستم به صورت تصاویر دو بعدی ثبت می‌شوند. در چنین شرایطی، نیازی به پردازش سه بعدی نیست و الگوریتم SIFT می‌تواند به خوبی ویژگی‌های مربوط به لب را تشخیص دهد. یکی دیگر از مزایای SIFT، انعطاف‌پذیری آن در برابر تغییرات مقیاس، چرخش و نور است. این الگوریتم می‌تواند به طور پایدار ویژگی‌های مختلف در تصاویر دو بعدی را استخراج کند و نسبت به تغییرات شرایط نوری و زاویه دید نیز مقاوم است. همچنین، به دلیل سرعت و کارایی بالا در محیط‌های بلادرنگ، به خصوص در پروژه‌هایی مانند تشخیص خواب‌آلودگی راننده، SIFT انتخاب مناسبی است. در مقایسه با سایر الگوریتم‌های مشابه، مثل SURF یا HOG، اگرچه این الگوریتم‌ها مزایای خاص خود را دارند، اما SURF ممکن است در شرایط پیچیده نوری دقت کمتری داشته باشد و



شکل ۶. هرم تصویری با تابع گوسین [۱۹]

Fig. 6. Image Pyramid with Gaussian Function



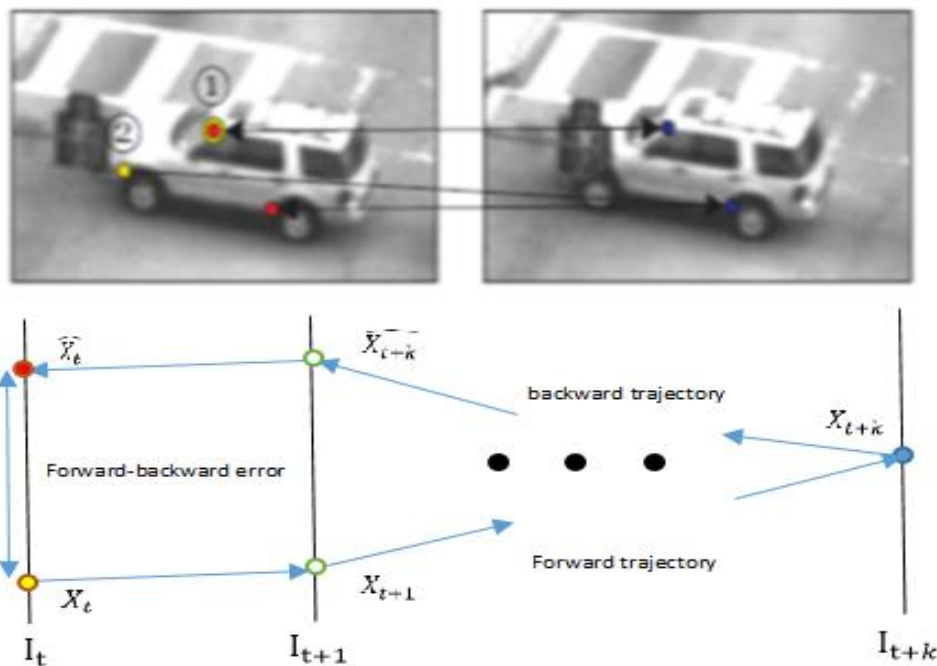
شکل ۷. الف) جهت و اندازه گرادیان محاسبه شده پیرامون نقاط کلیدی. ب) هیستوگرام برای جهت‌های گرادیان در هر همسایگی [۲۰]

Fig. 7. A) The Direction and Magnitude of Gradients Computed Around Keypoints. B) Histogram of Gradient Directions in Each Surrounding Area.

خواهد شد که در شکل ۷ نشان داده شده است. در نتیجه یک بردار عارضه
سیفت شامل یک بردار با $4 \times 4 \times 8 = 128$ عنصر خواهد شد.
ضرب داخلی بین بردارهای هر کدام از توصیف‌گرهای به‌دست آمده
از فریم اول که شامل یک بردار با ۱۲۸ آرایه می‌باشد و تمام بردارهای
توصیف‌گر نقاط کلیدی به‌دست آمده از تصویر دوم محاسبه می‌شود، حاصل
یک ماتریس برداری $1 \times n$ خواهد شد که در این جا n تعداد نقاط کلیدی

۲-۷- تشکیل توصیف‌گر برای نقاط کلیدی

به ازای هر نقطه اصلی پیکسل‌های مجاور (4×4) آن انتخاب می‌شوند
و در هر پیکسل ضریب زاویه‌ای آن (گرادیانت) تعیین شده و در مرکز ۱۶
پیکسل 4×4 ، برآیند هر یک بر روی چهار جهت اصلی و نیمساز آن‌ها ترسیم
می‌شود و تشکیل هیستوگرام می‌دهد. هر هیستوگرام شامل ۸ شاخه با توجه
به ۸ جهت اصلی خواهد شد و هر توصیف‌گر شامل یک آرایه ۴ هیستوگرامی



شکل ۸. ردیابی پیش‌رو و پس‌رو برای دو مشخصه از تصویر [۱۹]

Fig. 8. Forward and Backward Tracking for Two Features of the Image

و تفاوت بین این دو خط سیر اندازه‌گیری می‌شود. ردیابی نقطه یک روش معمول در بینایی کامپیوتر است. با توجه به موقعیت نقطه داده شده در زمان t ، هدف تخمین مکان آن در زمان $t+1$ می‌باشد. در عمل، چون نقاط از نظر ظاهری به طور چشمگیری تغییر می‌کنند یا از دید دوربین خارج می‌شوند، تحت این شرایط، نتیجه ردیابی اغلب با مشکل مواجه می‌شود. در شکل ۸ ردیاب قادر است نقطه شماره ۱ را در ردیابی پیش‌رو و پس‌رو به درستی به یک ناحیه محدود کند. نتیجه ردیابی این نقطه به صورت پیش‌رو یا پس‌رو خط سیر یکسانی خواهد داشت. از طرف دیگر، ردیاب نقطه شماره ۲ را در ردیابی پس‌رو به ناحیه‌های متفاوتی نسبت به نقطه واقعی نگاشت می‌کند [۱۹].

اگر $S = (I_t, I_{t+1}, \dots, I_{t+k})$ دنباله‌های تصویر باشد و X_t موقعیت نقطه در زمان t باشد. با استفاده از یک ردیاب اختیاری، نقطه‌ی X_t به صورت رو به جلو در k گام ردیابی می‌شود. خط سیر حاصل $T_f^k(x_t, x_{t+1}, \dots, x_{t+k})$ می‌باشد که برای پیش‌رو به کار می‌رود و k طول دنباله‌ها را نشان می‌دهد. هدف، تخمین خطا از خط سیر حاصل T_f^k با توجه به دنباله‌های تصویر S می‌باشد. نقطه‌ی X_{t+k} به صورت رو به عقب ردیابی می‌شود تا به فریم اول برسد و $T_b^k(\hat{x}_t, \hat{x}_{t+1}, \dots, \hat{x}_{t+k})$

استخراج شده از ناحیه چهره در فریم جاری می‌باشد. این ماتریس برداری به صورت صعودی مرتب می‌شود. مفهوم ضرب داخلی در اینجا بر اساس آنالیز برداری، همان مفهوم حداقل فاصله اقلیدسی را دارد لذا نسبت اولین عنصر کمینه به دومین عنصر کمینه در بردار ضرب داخلی را محاسبه کرده که این مقدار نباید از یک آستانه کمتر باشد (مقدار آستانه با استفاده از تصاویر آموزشی مشخص می‌گردد). در صورتی که حاصل این ضرب داخلی در دو مجموعه کمتر از آستانه مورد نظر بود، آن دو نقطه به عنوان دو نقطه متناظر در نظر گرفته می‌شود. این الگوریتم برای تمامی نقاط کلیدی به دست آمده در تصویر دوم تکرار می‌شود تا نقاط تطبیقی به دست بیاید. پس از همترازی تصاویر چهره و استخراج موقعیت دقیق نقاط لندمارک، عملیات ردیابی نقاط لندمارک در فریم‌های ویدئویی انجام می‌شود. در ادامه الگوریتم مورد استفاده برای ردیابی نقاط کلیدی آورده شده است [۲۰].

۲-۸- ردیابی نقاط کلیدی مربوط به لندمارک چهره در فریم‌های متوالی

پس از استخراج نقاط کلیدی سیفت برای لندمارک‌های چهره، از روش ارائه شده برای ردیابی، توسط خطای پیش‌رو - پس‌رو می‌توان استفاده کرد. یعنی ردیابی به طور هم‌زمان به صورت رو به جلو و رو به عقب اجرا می‌شود



شکل ۹. نمونه‌ای از تصاویر پایگاه داده YawDD

Fig. 9. Sample Images from the YawDD Database

خواب‌آلودگی می‌باشد. آستانه مورد نظر که برای تشخیص خواب‌آلودگی مورد استفاده قرار می‌گیرد ۳۷ پیکسل می‌باشد که به صورت تجربی در آزمایشات به دست آمده است.

۳- پایگاه داده

در این تحقیق از پایگاه داده تشخیص خمیازه استاندارد، برای ارزیابی روش پیشنهادی استفاده شده است. این پایگاه داده با نام Yawn-ing Detection Dataset (YawDD) برای ارزیابی سیستم‌های تشخیص خواب‌آلودگی عرضه شد [۲۱]. در تولید این پایگاه داده ۱۶ مرد و ۱۳ زن مشارکت داشته‌اند. در پایگاه داده YawDD ویدئوها در دو دسته در حالت خمیازه کشیدن و حالت عادی چهره قرار گرفته‌اند (شکل ۹).

۴- پیاده سازی روش پیشنهادی

الگوریتم پیشنهادی با استفاده از نرم افزار MATLAB 2020 و بر روی سیستمی با مشخصات Intel Core i5 با قدرت پردازش 3.2GHz و حافظه اصلی 8GB پیاده سازی و اجرا شد. زمان اجرای الگوریتم یکی از مواردی است که باید مورد توجه قرار بگیرد. در شکل ۱۰ فرآیند آشکارسازی لندمارک‌های چهره برای یک نمونه ویدئو مشخص شده است. برای هر ویدئو، سرعت پردازش بر حسب فریم بر ثانیه در کنار کادر خروجی نمایش

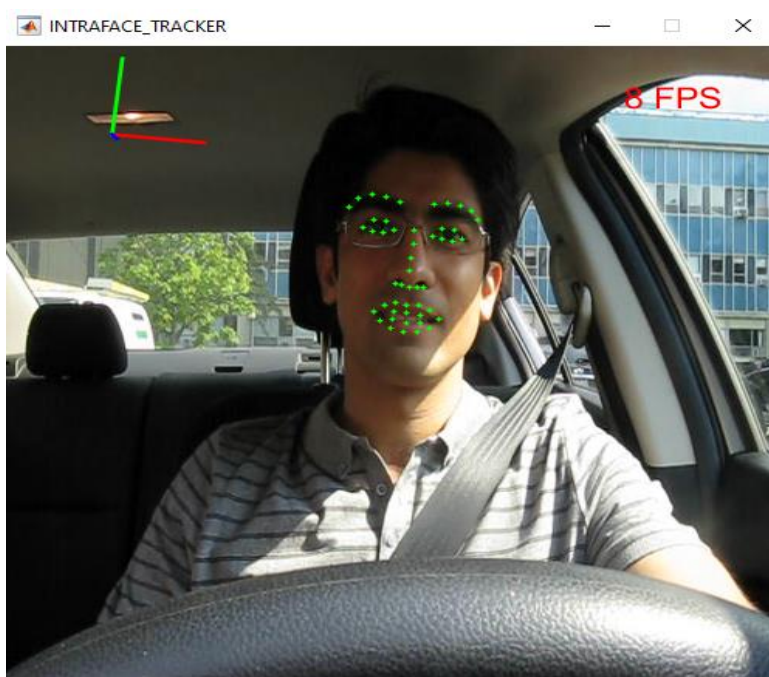
تولید شود، $\hat{X}_{t+k} = X_{t+k}$ و خطای پیش‌رو - پس‌رو، با استفاده از فاصله بین این دو خط سیر به صورت معادله ۹ تعریف می‌شود.

$$FB(T_b^k|S) = \text{distance}(T_f^k, T_b^k) \quad (9)$$

در این گزارش، از فاصله اقلیدسی بین نقطه اولیه و نقطه‌ی نهایی برای اعتبارسنجی خط سیر به صورت رابطه ۱۰ استفاده شده است.

$$\text{distance}(T_f^k, T_b^k) = \|x_t - \hat{x}_t\| \quad (10)$$

در صورتی که این فاصله اقلیدسی از مقدار اپسیلون بیشتر باشد (یعنی عملیات تطبیق با مشکل مواجه شود)، فرآیند آشکارسازی توسط الگوریتم سیفت مجدد انجام می‌شود و نقاط لندمارک به‌روزرسانی می‌شوند. پس از استخراج لندمارک‌های چهره و ردیابی آن‌ها، در هر فریم ویدئویی، برای تشخیص خواب‌آلودگی در این تحقیق از فاصله بین لندمارک‌های قرار گرفته روی لب‌ها در فریم‌های ویدئویی استفاده می‌شود. به این صورت که پس از استخراج لندمارک‌های لب، فاصله اقلیدسی لندمارک، لب بالایی و لب پایینی محاسبه می‌شود. این فاصله اقلیدسی به عنوان معیاری برای تشخیص



شکل ۱۰. آشکارسازی لندمارک‌های چهره در فریم ویدئویی اول

Fig. 10. Facial Landmarks Detection in the First Video Frame

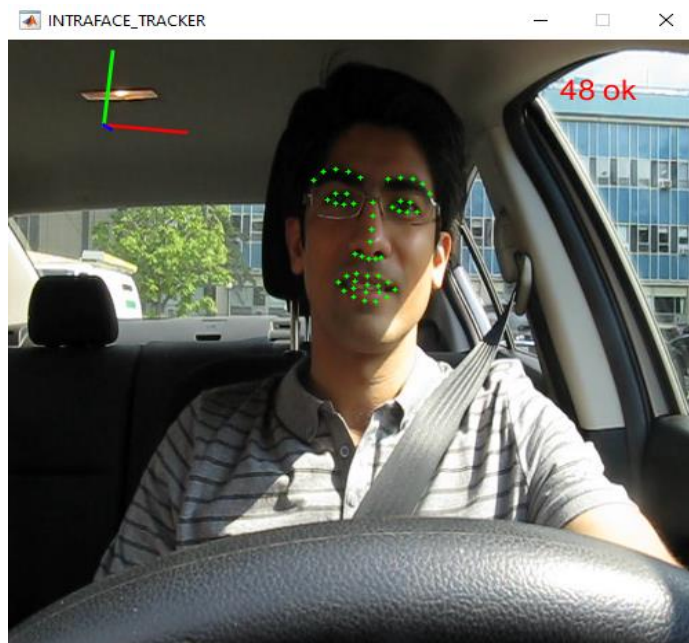
مراحل کد نویسی به شرح ذیل می‌باشد: ۱- پس از دریافت فریم‌های ویدئویی، ابتدا فرآیند آشکارسازی چهره در فریم اول با استفاده از الگوریتم ویولا - جونز انجام می‌شود. ۲- پس از آشکارسازی چهره، نقاط کلیدی با استفاده از الگوریتم سیفت از ناحیه چهره استخراج می‌شود. ۳- با توجه به اینکه آشکارسازی و ردیابی نقاط در مراحل بعدی نیازمند همترازی تصاویر چهره و تبدیل تصاویر به تصویر از مقابل می‌باشد، از نقاط کلیدی استخراج شده برای همترازی تصاویر چهره استفاده می‌شود. در این مرحله نقاط کلیدی استخراج شده با استفاده از تبدیل افاین به نقاطی نگاشت می‌شوند که در صورت زاویه دار بودن سر شخص، تصویر آن به تصویر از مقابل تبدیل شود. ۴- یکی دیگر از ویژگی‌های الگوریتم ویولا-جونز استخراج نواحی مربوط به لب‌ها، دهان و چشم‌ها می‌باشد. پس از آشکارسازی ناحیه مربوط به دهان شخص، برای افزایش سرعت پردازش، تنها نقاط کلیدی مربوط به لب بالایی و لب پایینی ردیابی می‌شوند. ۵- حال برای تشخیص خواب آلودگی فاصله اقلیدسی نقاط کلیدی لب بالایی و لب پایینی محاسبه می‌شود. ۶- در صورتی که فاصله اقلیدسی محاسبه شده در مرحله قبل از حد آستانه تعریف شده کمتر باشد، حالت شخص راننده عادی بوده و ردیابی نقاط کلیدی ادامه پیدا می‌کند، در غیر این صورت، نشان دهنده خواب آلودگی راننده بوده به علائم هشدار دهنده فعال می‌شوند.

داده می‌شود. در فریم اول با توجه به اینکه باید فرآیند آشکارسازی موقعیت چهره با استفاده از الگوریتم ویولا جونز انجام شده و نقاط کلیدی سیفت از نواحی چهره استخراج شوند، سرعت اجرا نسبتاً پایین (برای این نمونه ویدئو ۸ فریم بر ثانیه) می‌باشد.

با گذشت زمان، سرعت اجرای الگوریتم پیشنهادی (در ویدئوی نمونه به طور میانگین به ۴۸ فریم بر ثانیه) افزایش می‌یابد که نشان دهنده اجرای بالادرنگ روش پیشنهادی می‌باشد.

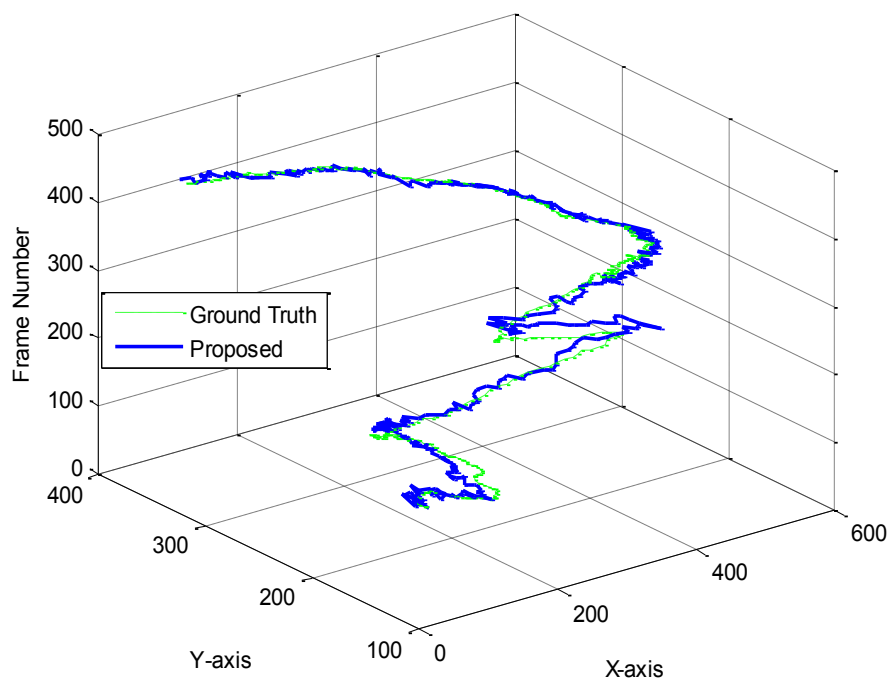
با توجه به اینکه در فریم‌های بعدی فرآیند آشکارسازی چهره و استخراج نقاط کلیدی حذف شده و تنها ردیابی لندمارک‌های چهره انجام می‌شود، منجر به افزایش سرعت پردازش خواهد شد که سرعت پردازش در شکل ۱۱ نیز نشان داده شده است.

برای نشان دادن صحت ردیابی و نحوه عملکرد الگوریتم ردیابی در فریم‌های متوالی، موقعیت دهان با استفاده از الگوریتم ویولا-جونز استخراج شد. موقعیت‌های به دست آمده توسط الگوریتم ویولا-جونز به عنوان معیار در نظر گرفته می‌شود. سپس خط سیر به دست آمده از ردیابی موقعیت دهان در فریم‌های متوالی با خروجی حقیقی مقایسه گردید. حاصل این مقایسه در شکل ۱۲ نشان داده شده است. همانطور که مشخص است، الگوریتم ردیابی می‌تواند با دقت خوبی عملیات ردیابی را انجام دهد.



شکل ۱۱. افزایش سرعت پردازش الگوریتم پیشنهادی

Fig. 11. Enhancement of the Processing Speed of the Proposed Algorithm



شکل ۱۲. خط سیر الگوریتم ارائه شده در مقایسه با خروجی حقیقی برای یک نمونه فریم ویدئویی

Fig. 12. A Comparison of the Algorithm's Trajectory with the Ground Truth for a Sample Video Frame.

جدول ۱. مقایسه عملکرد روش پیشنهادی و روش‌های مطرح در تشخیص خواب‌آلودگی روی پایگاه داده YawDD

Table 1. Comparison of Drowsiness Detection Performance Between the Proposed Method and Established Techniques on the YawDD Database.

خطا	دقت تشخیص	نوع الگوریتم
۱۷	۸۳	تشخیص خمیازه توسط آنالیزهای توصیف‌گر تغییرات [22]
۱۴	۸۶	پایش آنی خواب آلودگی راننده تحت کنترل Descent Method [21]
۸	۹۲	تشخیص خواب آلودگی راننده توسط شبکه یادگیری عمیق سلسله مراتبی زمانی [23]
۸	۹۲	تشخیص خواب آلودگی راننده توسط مجموعه شبکه های عصبی کانولوشنال [24]
۶	۹۴	روش پیشنهادی این تحقیق

۴-۱- مقایسه روش پیشنهادی با روش‌های مطرح

در جدول ۱ حاصل مقایسه روش پیشنهادی این تحقیق با روش‌های مطرح برای تشخیص خواب‌آلودگی که از پایگاه داده YawDD استفاده کرده‌اند، آورده شده است.

نتایج آزمایشات نشان می‌دهد روش پیشنهادی قادر است با دقت خوبی خواب‌آلودگی را نسبت به روش‌های مطرح تشخیص دهد. خطاهای ایجاد شده به دلیل در نظر گرفتن آستانه دستی "فاصله بین لب بالایی و پایینی" هنگام تشخیص خواب‌آلودگی می‌باشد. شخص هنگام صحبت کردن ممکن است دهان خود را بیش از حد معمول باز کند و یا هنگام خمیازه کشیدن، دهان را کمتر از حالت عرف باز کند، با توجه به اینکه برای تشخیص خواب‌آلودگی و خستگی از فاصله بین لب‌های بالا و پایین استفاده شده است. ممکن است آستانه در نظر گرفته شده برای تشخیص خواب‌آلودگی با خطا مواجه شود. پیشنهاد می‌شود، آستانه در نظر گرفته شده توسط یک سیستم هوشمند آموزش داده شود تا خطای تشخیص خواب‌آلودگی کاهش یابد.

برای ارزیابی بهتر روش به کار گرفته شده از معیار ارزیابی Recall استفاده می‌شود که در ادامه توضیح داده شده است. برای محاسبه و استفاده از مقادیر True Positives (TP)، True Negatives (TN)، False Positives (FP) و False Negatives (FN)، نیاز است در ابتدا بفهمیم که این مقادیر چه معنایی دارند و چطور به دست می‌آیند.

True Positives (TP): تعداد مواردی است که مدل به درستی به عنوان مثبت تشخیص داده است، یعنی نمونه‌هایی که واقعاً مثبت بوده‌اند (مثلاً خمیازه یا خواب‌آلودگی) و مدل به درستی آن‌ها را شناسایی کرده است.

True Negatives (TN): تعداد مواردی است که مدل به درستی به عنوان منفی تشخیص داده است، یعنی نمونه‌هایی که واقعاً منفی بوده‌اند (مثلاً راننده در حالت طبیعی و بدون خستگی) و مدل به درستی آن‌ها را شناسایی کرده است. False Positives (FP): تعداد مواردی است که مدل به اشتباه به عنوان مثبت تشخیص داده است، یعنی مواردی که مدل اشتباه کرده و آن‌ها را به عنوان خمیازه یا خواب‌آلودگی شناسایی کرده، در حالی که اینطور نبوده است. False Negatives (FN): تعداد مواردی است که مدل به اشتباه به عنوان منفی تشخیص داده است، یعنی مواردی که راننده واقعاً خمیازه زده یا خواب‌آلود بوده، اما مدل نتوانسته این وضعیت را تشخیص دهد.

برای بدست آوردن مقادیر TP، TN، FP، FN، باید آزمایشی انجام دهید که در آن نتایج واقعی رفتار راننده (مثبت یا منفی) با نتایج پیش‌بینی شده توسط مدل مقایسه شود. در هر بار پیش‌بینی مدل، یکی از چهار حالت فوق رخ می‌دهد. برای هر نمونه در داده‌های آزمایشی، مدل یا درست پیش‌بینی می‌کند یا اشتباه. در این تحقیق تعداد داده‌ها ۱۵۳ عدد بوده است و پارامترهای محاسبات (یعنی FN، FP، TN، TP) بر اساس نتایج، به ترتیب عبارتند از ۸۰، ۶۴، ۳ و ۶ حال باید معیارهای دقت، صحت و Recall محاسبه گردد. دقت (Accuracy) به معنای درصد پیش‌بینی‌های صحیح (چه مثبت و چه منفی) از کل نمونه‌ها است. صحت (Precision) نشان می‌دهد از میان تمام مواردی که مدل به عنوان مثبت شناسایی کرده، چه درصدی واقعاً مثبت بوده‌اند. معیار Recall نشان می‌دهد از تمام موارد مثبت واقعی، چند درصد توسط مدل شناسایی شده‌اند (روابط ۱۱، ۱۲ و ۱۳).

استفاده شود. همچنین با توجه به اینکه حواس پرتی از طریق چرخاندن سر به اطراف یا صحبت کردن یکی دیگر از عواملی است که رانندگی را با خطر مواجه می‌کند، پیشنهاد می‌شود برای تحقیقات آینده روی موضوع حواس پرتی تمرکز گردد.

منابع

- [1] R. W. Wohleber, G. Matthews, G. J. Funke, and J. Lin, "Considerations in Physiological Metric Selection for Online Detection of Operator State: A Case Study," in Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience, vol. 9743, D. D. Schmorrow and C. M. Fidopiastis, Eds., in Lecture Notes in Computer Science, vol. 9743, Cham: Springer International Publishing, 2016, pp. 428–439. doi: 10.1007/978-3-319-39955-3_40.
- [2] M. J. Parsa, M. Javadi, and A. H. Mazinan, "Fatigue level detection using multivariate autoregressive exogenous nonlinear modeling based on driver body pressure distribution," Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 236, no. 1, pp. 168–184, Jan. 2022, doi: 10.1177/09544070211014290.
- [3] R. E. P. Soares, "Driver monitoring systems of fatigue based on eye tracking," PhD Thesis, 2017. Accessed: Apr. 24, 2024. [Online]. Available: <https://repositorium.sdum.uminho.pt/handle/1822/54750>
- [4] G. Matthews, R. Wohleber, J. Lin, G. Funke, and C. Neubauer, "Monitoring Task Fatigue in Contemporary and Future Vehicles: A Review," in Advances in Human Factors in Simulation and Modeling, vol. 780, D. N. Cassenti, Ed., in Advances in Intelligent Systems and Computing, vol. 780, Cham: Springer International Publishing, 2019, pp. 101–112. doi: 10.1007/978-3-319-94223-0_10.
- [5] H. Singh, J. S. Bhatia, and J. Kaur, "Eye tracking based driver fatigue monitoring and warning system," in India International Conference on Power Electronics 2010 (IICPE2010), IEEE, 2011, pp. 1–6. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/>

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \quad (11)$$

$$\frac{80 + 64}{80 + 64 + 3 + 6} = \frac{144}{153} \approx 0.941$$

$$Precision = \frac{TP}{TP + FP} = \frac{80}{80 + 3} = \frac{80}{83} \approx 0.964 \quad (12)$$

$$Recall = \frac{TP}{TP + FN} = \frac{80}{80 + 6} = \frac{80}{86} \approx 0.930 \quad (13)$$

۵- جمع بندی

تشخیص خواب‌آلودگی رانندگان یکی از راه‌های موثر در کاهش تصادفات رانندگی می‌باشد. روش‌های تشخیص خواب‌آلودگی به سه دسته جامع روش‌های مبتنی بر علائم فیزیولوژی، روش‌های مبتنی بر عملکرد راننده و روش‌های مبتنی بر ظاهر و حالت راننده تقسیم می‌شوند. از میان این سه دسته، روش‌های مبتنی بر ظاهر راننده به دلیل عدم ایجاد مزاحمت برای رانندگان و دقت مناسب، بیشتر از دو روش دیگر مورد توجه محققین قرار گرفته‌اند.

در این تحقیق سیستمی جامع برای تشخیص خواب‌آلودگی رانندگان براساس پردازش تصاویر چهره ارائه شد. با توجه به اینکه برای تشخیص خواب‌آلوده بودن شخص ابتدا باید ناحیه چهره در تصویر مشخص شود، در این تحقیق از الگوریتم ویولا-جونز برای آشکارسازی چهره استفاده می‌شود. سپس لندمارک‌های چهره با استفاده از الگوریتم سفت استخراج خواهد شد. برای افزایش سرعت پردازش نیز به جای ردیابی کل چهره، از ردیابی نقاط کلیدی استخراج شده استفاده شد. در نهایت از فاصله اقلیدسی نقاط کلیدی لب بالایی و لب پایینی برای تشخیص خواب‌آلودگی راننده استفاده شد. مجموع الگوریتم‌های استفاده شده در این تحقیق نشان می‌دهد، روش پیشنهادی قادر است با دقت خوبی عملیات تشخیص خواب‌آلودگی راننده را انجام دهد. از دیگر مزایای روش پیشنهادی می‌توان به اجرای بلادرنگ ۴۷ فریم بر ثانیه و دقت ۹۴ درصد بر روی پایگاه داده YawDD اشاره کرد. پیشنهاد می‌گردد در تحقیقات آتی، با توجه به ایجاد تغییرات در سیگنال‌های مغزی در هنگام خواب‌آلودگی، از ترکیب عوامل ظاهری با سیگنال‌های دریافت شده از مغز، برای تشخیص دقیق‌تر خواب‌آلودگی در هنگام رانندگی

- [13] J. Jose, J. S. Vimali, P. Ajitha, S. Gowri, A. Sivasangari, and B. Jinila, "Drowsiness Detection System for Drivers Using Image Processing Technique," in 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Jun. 2021, pp. 1527–1530. doi: 10.1109/ICOEI51242.2021.9452864.
- [14] V. Naren Thiruvalar and E. Vimal, "A comparative analysis on driver drowsiness detection using CNN," International Journal of Nonlinear Analysis and Applications, Dec. 2021, doi: 10.22075/ijnaa.2021.5894.
- [15] K. Balasubramanian, G. D. K, and Ramya, "Drowsiness detection and safety monitoring using image processing", doi: 10.6703/IJASE.202209_19(3).001.
- [16] Y. Albadawi, A. AlRedhaei, and M. Takruri, "Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features," Journal of Imaging, vol. 9, no. 5, Art. no. 5, Apr. 2023, doi: 10.3390/jimaging9050091.
- [17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, Ieee, 2001, p. I–I. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/990517/>
- [18] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
- [19] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in 2010 20th international conference on pattern recognition, IEEE, 2010, pp. 2756–2759. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/5596017/>
- [20] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [21] N. Neshov and A. Manolova, "Drowsiness monitoring in real-time based on supervised descent method," in [abstract/document/5728062/](https://ieeexplore.ieee.org/abstract/document/5728062/)
- [6] J. Xu, J. Min, and J. Hu, "Real-time eye tracking for the assessment of driver fatigue," Healthc. technol. lett., vol. 5, no. 2, pp. 54–58, Apr. 2018, doi: 10.1049/htl.2017.0020.
- [7] R. Bhardwaj, P. Natrajan, and V. Balasubramanian, "Study to determine the effectiveness of deep learning classifiers for ECG based driver fatigue classification," in 2018 IEEE 13th international conference on industrial and information systems (ICIIS), IEEE, 2018, pp. 98–102. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8721391/>
- [8] L. Wang, H. Wang, and X. Jiang, "A new method to detect driver fatigue based on EMG and ECG collected by portable non-contact sensors," Promet-Traffic&Transportation, vol. 29, no. 5, pp. 479–488, 2017.
- [9] R. Bhardwaj, S. Parameswaran, and V. Balasubramanian, "Comparison of Driver Fatigue Trend on simulator and on-road driving based on EMG correlation," in 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), IEEE, 2018, pp. 94–97. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8721431/>
- [10] X. Hu and G. Lodewijks, "Exploration of the effects of task-related fatigue on eye-motion features and its value in improving driver fatigue-related technology," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 80, pp. 150–171, Jul. 2021, doi: 10.1016/j.trf.2021.03.014.
- [11] S. Ananthi, R. Sathya, K. Vaidehi, and G. Vijaya, "Drivers Drowsiness Detection using Image Processing and I-Ear Techniques," presented at the 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 2023, pp. 1326–1331.
- [12] M. Hossain, M. Hasan, K. Abid, and S. Ahmed, "Developing an Image Processing Based Real-Time Driver Drowsiness Detection System," International Journal of Scientific and Research Publications, vol. 13, Apr. 2023, doi: 10.29322/IJSRP.13.04.2023.p13622.

- Workshops, vol. 10118, C.-S. Chen, J. Lu, and K.-K. Ma, Eds., in Lecture Notes in Computer Science, vol. 10118. , Cham: Springer International Publishing, 2017, pp. 117–133. doi: 10.1007/978-3-319-54526-4_9.
- [24] R. M. Salman, M. Rashid, R. Roy, M. M. Ahsan, and Z. Siddique, “Driver Drowsiness Detection Using Ensemble Convolutional Neural Networks on YawDD,” Dec. 19, 2021, arXiv: arXiv:2112.10298. Accessed: Apr. 24, 2024. [Online]. Available: <http://arxiv.org/abs/2112.10298>.
- [25] S. Eskandari, A.M.Rahimi, E.R.Khansari, “Study of cognitive variables for driver fatigue to reduce severity and prevent accidents,” Master’s Thesis, Department of Engineering, University of Zanjan, pp.34-42, Aug 2023.
- 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IEEE, 2017, pp. 660–663. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8095173/>
- [22] B. Akrouit and W. Mahdi, “Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness,” in 2016 International Image Processing, Applications and Systems (IPAS), IEEE, 2016, pp. 1–5. Accessed: Apr. 24, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7880127/>
- [23] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, “Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network,” in Computer Vision – ACCV 2016

چگونه به این مقاله ارجاع دهیم

S. Eskandari, A. M. Rahimi, E. Ramezani-khansari, *Detecting Driver Drowsiness by Combining Viola-Jones Image Processing Method and Image Landmarks Detection*, *Amirkabir J. Civil Eng.*, 57(3) (2025) 489-506.

DOI: [10.22060/ceej.2025.23159.8122](https://doi.org/10.22060/ceej.2025.23159.8122)

