

## Research Article

# Illumination Invariant Motion Detection and Tracking Using SMDWT and a Dense Disparity-Variance Method

V. A. Deepambika <sup>1,2</sup> and M. Abdul Rahman <sup>3,4</sup>

<sup>1</sup>Karpagam Academy of Higher Education, Department of ECE, 641021, Tamil Nadu, India

<sup>2</sup>LBS Institute of Technology for Women, Department of ECE, 695012, Kerala, India

<sup>3</sup>Karpagam Academy of Higher Education, 641021, Coimbatore, Tamil Nadu, India

<sup>4</sup>APJ Abdul Kalam Kerala Technological University, 695016, Kerala, India

Correspondence should be addressed to V. A. Deepambika; [deepambika.va@gmail.com](mailto:deepambika.va@gmail.com)

Received 29 June 2018; Revised 5 October 2018; Accepted 31 October 2018; Published 30 December 2018

Academic Editor: Xinyu Liu

Copyright © 2018 V. A. Deepambika and M. Abdul Rahman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The navigation management systems in autonomous vehicles should be able to gather solid information about the immediate environment of the vehicle, discern ambulance from a delivery truck, and react in a proper manner to handle any difficult situation. Separating such information from a vision controlled system is a computationally demanding task for heavy traffic areas in the real world environmental conditions. In such a scenario, we need a robust moving object detection tracking system. To achieve this, we can make use of stereo vision-based moving object detection and tracking, utilizing symmetric mask-based discrete wavelet transform to deal with illumination changes, low memory requirement, and fake motion avoidance. The accurate motion detection in complex dynamic scenes is done by the combined background subtraction and frame differencing technique. For the fast motion track, we can employ a dense disparity-variance method. This SMDWT-based object detection has a maximum and minimum accuracy of 99.62% and 94.95%, respectively. The motion track has the highest accuracy of 79.47% within the time frame of 28.03 seconds. The lowest accuracy of the system is 62.01% within the time frame of 34.46 seconds. From the analysis, it is clear that this proposed method exceptionally outperforms the existing monocular and dense stereo object tracking approaches in terms of low computational cost, high accuracy, and in handling the dynamic environments.

## 1. Introduction

Recently, object tracking has a vital role in most of the computer vision applications like man-machine interface, robotic vision, and intelligent security system. In medical imaging, object tracking is used for the diagnosis of diseases [1]. Even though heavily investigated for decades by many researchers, an accurate and high-performance motion track approach is still far away from most of the real-time requirements. The degree of difficulty in this task highly depends on the dynamic changes in the environment as well as the type of object to be detected and tracked. In the complex dynamic traffic situations, the driver assistance system (DAS) must

accurately figure out the scene and react properly to handle any difficult situation. The navigation management system in the autonomous vehicles normally depend on various range measurement techniques like GPS, laser rangefinder, video cameras, and several subsystems to discern where the vehicle is going and to take crucial decisions according to the data coming from these sensors [2].

Similar to human eyes, cameras and their related subsystem can cognize the scenes, which no other sensors can handle. Conventional approaches for moving object detection are background subtraction [3, 4], temporal differencing [5], statistical methods [6], and optical flow method [7, 8]. In the real world scenario, these methods are sensitive to

radiometric variations, sensor noises, and fake motions. Often, the discrete wavelet transform (DWT) [9] and low resolution techniques are used to handle these situations [10]. The conventional DWT schemes can be performed only at a higher computational cost. The increase in the level of decomposition used to meet the low memory requirements often ends in incomplete object detection. The symmetric mask-based discrete wavelet transform (SMDWT) is an apt choice to meet these demands, as it is always associated with attractive characteristics such as shorter path, independent subband coding, and fast computation [11, 12]. Generally, these approaches are done on monocular videos and it often face challenges such as multiple object occlusion, shadow interference, and radiometric changes. Also, the horizon of a single camera should be wide enough to cover the objects in motion. A multicamera system can accurately locate the moving objects without any ambiguity [13]. Accurate localization and segmentation of the objects of interests can be done with the help of depth information extracted from the stereo views [14–16]. This study is about a fast and reliable moving object tracking system using a dense disparity-variance method. Precise motion detection is accomplished by combining the background subtraction and the frame differencing method, incorporating the advantages of SMDWT.

## 2. Related Work

In the recent developments in video acquisition technology, multiple view systems dominate over the monocular ones. With the use of stereo vision technology, 3D information of the objects can be estimated accurately at a sub-pixel level, which can be utilized for more reliable object tracking. Stereo systems can be of either wide-baseline or short-baseline systems. Feature-based correspondence is usually done on wide-baseline systems [17, 18] that results in sparse disparity. The primary challenge associated with the wide-baseline systems is the large viewpoint-related occlusions. The dense depth map obtained from short-baseline systems [19–21] can be used for precise segmentation of the objects of interests.

The stereo vision system is used in multiobject tracking [22]. Single camera and the multiple sensor fusion are used to track humans [23]. For object tracking at traffic roundabouts, a stereo vision system is employed [24]. The dense stereo vision system is used for pedestrian tracking [25]. Stereo inputs are utilized for identifying and segmenting moving objects [26]. Stixel World representation of the traffic scenes are used for autonomous systems [27]. By incorporating the epipolar constraints, Andre et al. [28] used wide-baseline stereo vision for object tracking. Kalman filter is used for stereo video-based object tracking [29]. The adaptive particle filter is employed for stereo-based motion tracking [30]. Object detection based on the classification of feature points in the 3D space is suggested for self-driving cars [31]. To manage the vehicle tracking in urban traffic, cuboids estimated from stereo views are used [32]. Stereo vision and LiDAR technology for autonomous driving vehicles is

presented in [33]. Dense stereo is used for vehicle tracking in urban road intersections [34]. RANSAC algorithm-based tracking [35] and moving object detection [36] used stereo pairs for motion estimation. Vehicle tracking suggested in [37] uses AdaBoost detectors and optical flow. In most of the existing tracking methods, illumination variations, fake motions, dynamic scenes, accuracy, and computational cost are the main challenges.

## 3. Features and Contributions

This is the first stereo vision-based fast motion object detection and tracking method for dynamic environments. It incorporates the advantages of symmetric mask-based discrete wavelet transform, combined background subtraction and frame differencing technique, and dense disparity-variance method. The main contributions of this study include the following:

- (i) The use of SMDWT to deal with the illumination changes, low memory requirement, and fake motion avoidance
- (ii) The accuracy and the reliability of motion detection are increased by combining background subtraction and frame differencing methods. This combined method can handle dynamic scene changes, and it does not need to frequently update the background frame. Instead, it is to be updated only on the base of a threshold distance covered by the sensor
- (iii) The dense stereo disparity is applied for removing unwanted outliers in the region of interest and thus enhancing the accuracy of object detection. Finally, the dense stereo disparity-variance method is used for fast object tracking and the use of the dense depth map increases the accuracy of this distance-based object tracking

Apart from these, the SMDWT-based object detection attains an accuracy of 99.62%. The lowest detection accuracy of the system is 94.95% for the compressed input image size of  $80 \times 110$ . It gives tracking results within a low time frame compared to the lifting-based discrete wavelet transform (LDWT) method. This novel dense disparity-variance method uses only the maximum variance value of pixels within a window, instead of taking pixel by pixel values which gives faster motion tracking results when compared with the current approaches. This system integrates the advantages of combined background subtraction and frame differencing technique, SMDWT, and dense stereo disparity-variance object track. It has a highest tracking accuracy of 86.4% within a time frame of 35.23 seconds, and the lowest tracking accuracy of the system is 62.01% within a time frame of 34.46 seconds. The results indicate a better performance of the entire system, in terms of accuracy, computational cost, and in handling dynamic scenes than the usual DWT-based approaches and the existing dense stereo object tracking approaches.

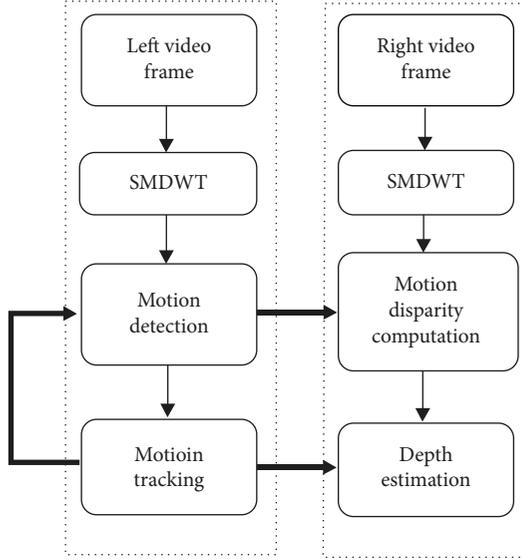


FIGURE 1: Block schematic of the motion detection and tracking system using dense disparity-variance technique.

#### 4. Motion Detection and Tracking Using SMDWT and a Dense Disparity-Variance Method

This stereo vision-based object tracking method incorporates the features of SMDWT in dealing with radiometric invariance, low memory requirement, and fake motion avoidance. The salient features of the system include accurate object detection without any ambiguity, fast object tracking of moving objects, and accurate region of interest (ROI) selection. The accuracy of this distance-based tracking is perfected by the use of the stereo vision-based dense depth map. Further, this system does not need to frequently update the reference frame. It is to be updated only if the depth value of the ROI becomes greater than that of a threshold value.

The stereo information is utilized in two modules: (i) in ROI outlier removal to improve accuracy and (ii) in the disparity-variance-based motion tracking. For the disparity computation, the sum of absolute difference (SAD) algorithm is used. Robust object detection in dynamic traffic scenes is done using combined background subtraction and frame differencing technique. Absolute object detection and fast motion tracking are accomplished by the dense disparity-variance method. The entire 3D visual tracking module consists of moving object detection, depth estimation using SAD stereo matching, and the tracking of multiple moving objects. Figure 1 shows the system overview.

**4.1. Symmetric Mask-Based Discrete Wavelet Transform (SMDWT).** Convolutional or FIR filter bank structure discrete wavelet transform (DWT) used to detect moving parts in real world environments is done at a high computational cost, and hence, they are not an apt choice for high-speed

TABLE 1:  $LL_1$  subband coefficients of SMDWT mask.

$-1/32$	$1/16$	$3/16$	$1/16$	$-1/32$
$-3/32$	$3/16$	$9/16$	$3/16$	$-3/32$
$-1/32$	$1/16$	$3/16$	$1/16$	$-1/32$
$1/64$	$-1/32$	$-3/32$	$-1/32$	$-1/64$

video processing applications [38]. In comparing with convolution-based transform, lifting-based DWT (LDWT) needs only 50% less computational complexity from video compression methods [39, 40]. Though LDWT has excellent reconstruction property, it requires more transpose memory. To reduce this requirement, the mask-dependent processing SMDWT can be used for motion detection and tracking [41, 42]. In this work, symmetric mask-based discrete wavelet transform (SMDWT) is applied on both stereo frames using the coefficient derived from the 2-D 5/3 integer LDWT. In order to conduct spatial filtering at less computational cost, four masks,  $3 \times 3$ ,  $5 \times 3$ ,  $3 \times 5$ , and  $5 \times 5$ , are used. On further optimization, these four subband processings will lead to a reduced temporal memory requirement and speedy results. Using these masks,  $LL_1$ ,  $LH_1$ ,  $HL_1$ , and  $HH_1$  subbands are obtained. Of these, the  $LL_1$  subband is devoid of high-frequency components such as fake motions and fast illumination changes. Therefore, in this stereo object tracking system, the  $LL_1$  subband is utilized for spatial filtering. Table 1 represents the coefficients of  $LL_1$  mask. Figure 2 shows the schematic of two-dimensional SMDWT.  $LL_1$  subband can be calculated using the following equation.

$$\begin{aligned}
 LL_1(i, j) = & \left(\frac{9}{16}\right)x(2i, 2j) \\
 & + \left(\frac{1}{64}\right)\sum_{u=0}^1\sum_{v=0}^1x(2i-2+4u, 2j-2+4v) \\
 & + \left(\frac{1}{16}\right)\sum_{u=0}^1\sum_{v=0}^1x(2i-1+2u, 2j-1+2v) \\
 & + \left(-\frac{1}{32}\right)\sum_{u=0}^1\sum_{v=0}^1x(2i-1+2u, 2j-2+4v) \\
 & + \left(-\frac{1}{32}\right)\sum_{u=0}^1\sum_{v=0}^1x(2i-2+4u, 2j-1+2v) \\
 & + \left(\frac{3}{16}\right)\sum_{u=0}^1[x(2i-1+2u) + x(i, 2j-1+2u)] \\
 & + \left(-\frac{3}{32}\right)\sum_{u=0}^1[x(2i-1+4u) + x(2i, 2j-2+4u)].
 \end{aligned} \tag{1}$$

**4.2. Motion Detection.** The steps involved in motion detection are

- (a) background subtraction and frame differencing

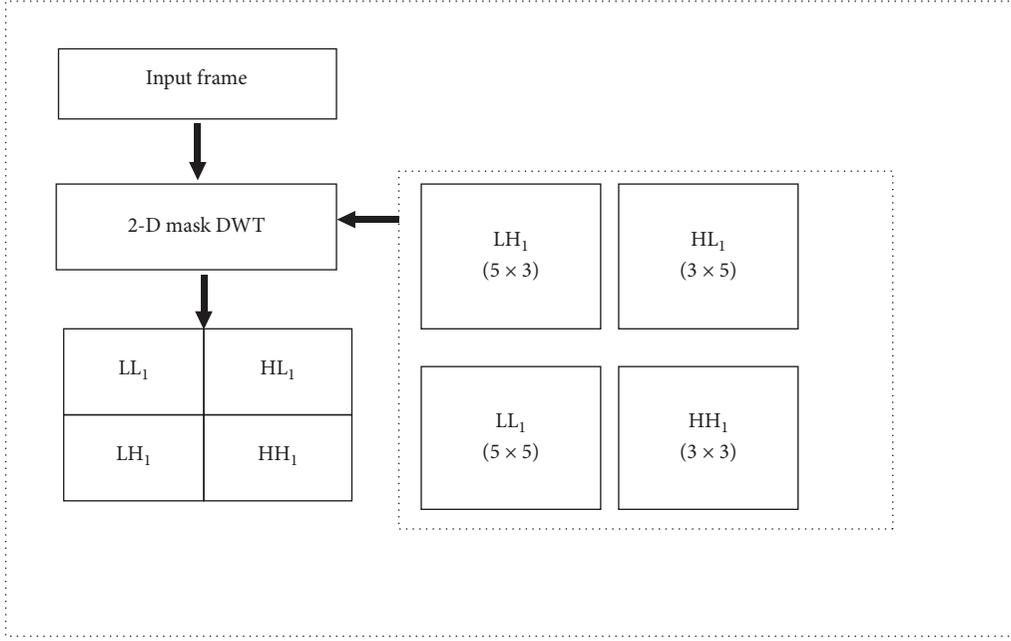


FIGURE 2: Block schematic of 2-D SMDWT. LL<sub>1</sub>: low-low-frequency subband of SMDWT. HL<sub>1</sub>: high-low-frequency subband of SMDWT. LH<sub>1</sub>: low-high-frequency subband of SMDWT. HH<sub>1</sub>: high-high-frequency subband of SMDWT.

(b) binarization and OR operation

(c) morphological operations

Motion detection is followed by motion disparity computation which can be used to increase the accuracy of motion tracking.

**4.2.1. Combined Background Subtraction and Frame Differencing Method.** The background subtraction (BS) technique detects the moving objects from the present frame based on a reference frame. Though the motion mask vector can be estimated from the background subtraction, it is susceptible to illumination changes and other radiometric changes. This can be overcome by incorporating the frame differencing (FD) method which can handle dynamic scene changes. Thus, in this proposed method, the motion mask data is obtained from the combined BS and FD method. Only the steady-state objects in the background are considered for the reference frame, and the current frame is subtracted from the reference for getting the objects in motion. In FD, the moving parts are separated out from the successive frames in a video stream. The background subtraction and frame differencing can be done by using the following equations.

$$\begin{aligned} \{BS(x, y)\}_k &= \{Fr(x, y)\}_k - \{Fr(x, y)\}_{k=0}, \\ \{FD(x, y)\}_k &= \{Fr(x, y)\}_k - \{Fr(x, y)\}_{k-1}, \end{aligned} \quad (2)$$

where  $x = 1, 2, 3, \dots, m$ ,  $y = 1, 2, 3, \dots, n$ , and  $k = 1, 2, 3, \dots, q$ .  $m$  and  $n$  represent the number of rows and columns in a frame, and  $q$  is the number of frames in the video.  $\{BS(x, y)\}_k$  is the background frame, and  $\{FD(x, y)\}_k$  is the frame difference.

**4.2.2. Binarization and Logical OR Operation.** The data obtained from the BS and FD method is combined by the binarization and logical OR operation. The threshold value for binarization may change from one frame to another. The equation (3) is used to calculate the mean  $\mu_k$  and standard deviation  $\sigma_k$ .

$$\begin{aligned} \mu_k &= \frac{1}{m * n} \sum_{x=1}^m \sum_{y=1}^n FD(x, y)_k, \\ \sigma_k &= \sqrt{\frac{1}{m * n} \sum_{x=1}^m \sum_{y=1}^n ((FD(x, y)_k - \mu_k))^2}. \end{aligned} \quad (3)$$

From these values, the global threshold “ $T_h$ ” for a particular frame can be calculated using the following equation:

$$T_h = \sigma_k * 0.085. \quad (4)$$

FD is the result of frame differencing,  $\mu_k$  is the mean value of FD,  $\sigma_k$  is the standard deviation,  $T_h$  is the threshold value, and 0.085 is the constant value obtained from trial and error method.  $O_r$ , the result of OR operation, can be obtained using the following equation:

$$\{O_r(x, y)\} = \{BS(x, y)\}_k \cup \{FD(x, y)\}_k. \quad (5)$$

**4.2.3. Morphological Operations.** Undesirable features in a binary image can be eliminated by morphological operations. Some motions are too feeble, which may be taken as nonmotion pixels and will create holes in the motion mask. To generate accurate motion mask, the morphological closing is applied.

**4.3. Motion-Based Disparity Computation.** The motion mask obtained after the morphological operation may contain the outlier parts from its surroundings. Accordingly, these outlier portions lead to the incorrect estimation of the motion parameters. The following steps are involved in the motion-related region of interest (ROI) generation and the motion-based disparity computation.

**4.3.1. ROI Extraction.** From the motion mask and input images, ROI can be generated using the following equations:

$$I_{Lroi}(x, y) = \begin{cases} I_{Left(x,y)}, & \text{if } MM_k == 1, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

$$I_{Rroi}(x, y) = \begin{cases} I_{Right(x,y)}, & \text{if } MM_k == 1, \\ 0, & \text{otherwise.} \end{cases}$$

Here,  $MM_k$  is the motion mask obtained from the morphological operation.  $I_{Left}$  and  $I_{Right}$  are the left and right input images.  $I_{LROI}$  and  $I_{RROI}$  are the left and right ROI images. The unwanted outliers in the misaligned ROI can be removed on the basis of disparity, i.e., the estimated motion vectors were converted into disparity vectors.

**4.4. Sum of Absolute Differences (SAD) Stereo Correspondence.** In the SAD-based correlation approach, the intensity of pixels within a window taken from the reference image is matched with its corresponding one in the target image [42]. The 2-D correspondence search can be brought down to a 1-D search (along the  $x$  direction only) if the input image pairs are rectified [43].

Here, all the epipolar lines in two views are in parallel with the baseline, and the search is to be done along the horizontal ( $x$ ) direction only. Let  $N$  represent the amount of pixels in either direction from the centre pixel ( $I, j$ ) within a window of size  $(2N + 1)$ . The cost function SAD used for similarity measurement is given by the equation (7). Disparity ( $d$ ), the shift in  $x$  coordinate of the conjugate pair, can be obtained using equation (8).

$$SAD(x, y, d) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} |I_{Lroi}(x, y) - I_{Rroi}(x, y)|, \quad (7)$$

where  $I_{LROI}(x, y)$  be the intensity of the pixel at  $(x, y)$  in the left frame  $I_{RROI}(x, y)$  and the right frame. SAD values are computed for all pixels within the window, and the ‘‘winner’’ is the disparity corresponding to the SAD minimum.

$$\text{Disparity}(x, y) = \min_{d \in D} (SAD(x + d, y)). \quad (8)$$

The disparity value of the pixel window around the centre pixel  $(x, y)$  changes from zero to maximum disparity value ‘‘ $d_{max}$ .’’ By sliding the reference window over the right frame, the search can be carried out. This process repeats for individual block until the disparity value is

TABLE 2: Video frames used for the analysis.

Input video	Total number of frames	Image resolution
Video 1	629	640 × 880
Video 2	500	320 × 440
Video 3	165	640 × 880
Video 4	95	640 × 880
Video 5	700	320 × 440
Video 6	150	640 × 880

calculated for the entire frame and the resultant image is called disparity space image (DSI). Here, the DSI corresponding to moving object is represented as  $DSI(M_k(x, y))$ . The images obtained after SMDWT are used as the input sequence for this correspondence search. The depth ( $Z$ ) can be computed by triangulation based on the following equation.

$$Z = \frac{f \cdot b}{x_l - x_r} = \frac{f \cdot b}{d}, \quad (9)$$

where  $f$ ,  $b$ , and  $d$  represent the camera focal length, the baseline distance between the camera centres, and the computed disparity value, respectively. This depth value can be used to identify moving objects that come below the threshold distance. In this work, this threshold distance is taken as 5 m.

**4.4.1. Removal of Outlier in ROI.** The unwanted outlier in ROI can be removed based on the following equation:

$$M_k = \begin{cases} \text{if } DSI(M_k(x, y)) > T_k, & \text{the feature is removed,} \\ \text{otherwise,} & \text{the feature is selected,} \end{cases} \quad (10)$$

where  $M_k$  represents the motion mask of the  $k^{\text{th}}$  frame. The disparity of the motion vector ( $DSI(M_k)$ ) is used to eliminate the outliers, and  $T_m$  is the threshold level of the motion mask vector. The mean value and variance of each frame are updated periodically.

**4.5. Variance-Based Motion Tracking.** In this variance-based motion tracking, the variance of pixels is calculated for a window of size five taken from the ROI. The variance value can be calculated based on the following equation:

$$\sigma_k^2 = \frac{1}{m * n} \sum_{x=1}^m \sum_{y=1}^n M_k^2(x, y), \quad (11)$$

where  $M_k$  is the motion mask obtained after removing the unwanted outliers. The window slides horizontally in order to cover the entire ROI. Here, only the maximum variance value of pixels within a window is considered, instead of taking pixel by pixel values. This window-based approach gives faster motion tracking results, and the aggregation

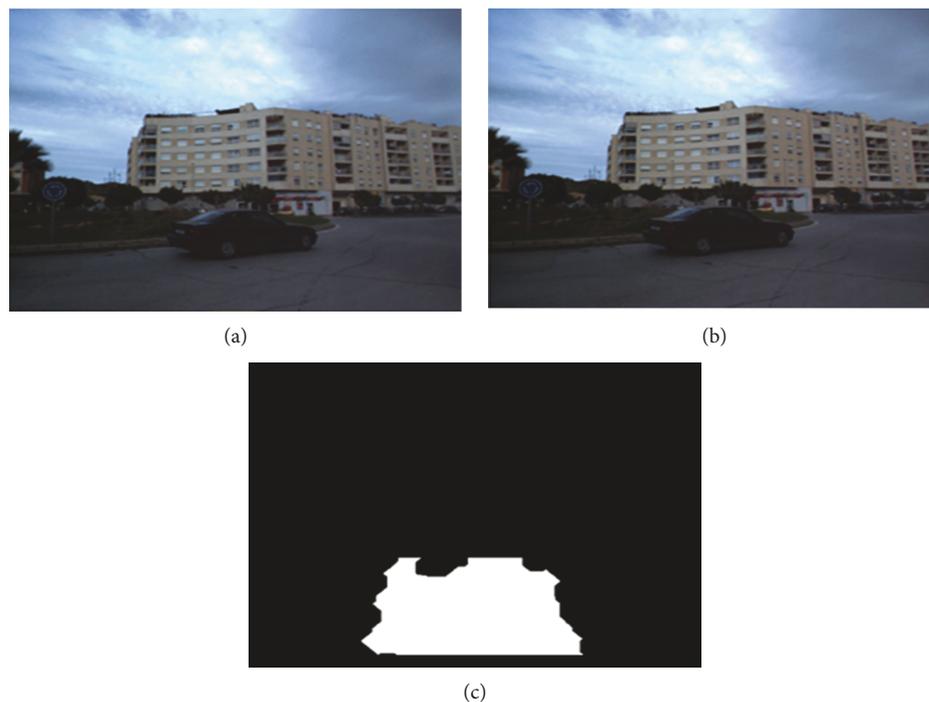


FIGURE 3: Generation of motion detection mask. (a) Left frame. (b) Right frame. (c) Generated mask using the  $LL_1$  subband image and combined BS and FD method.

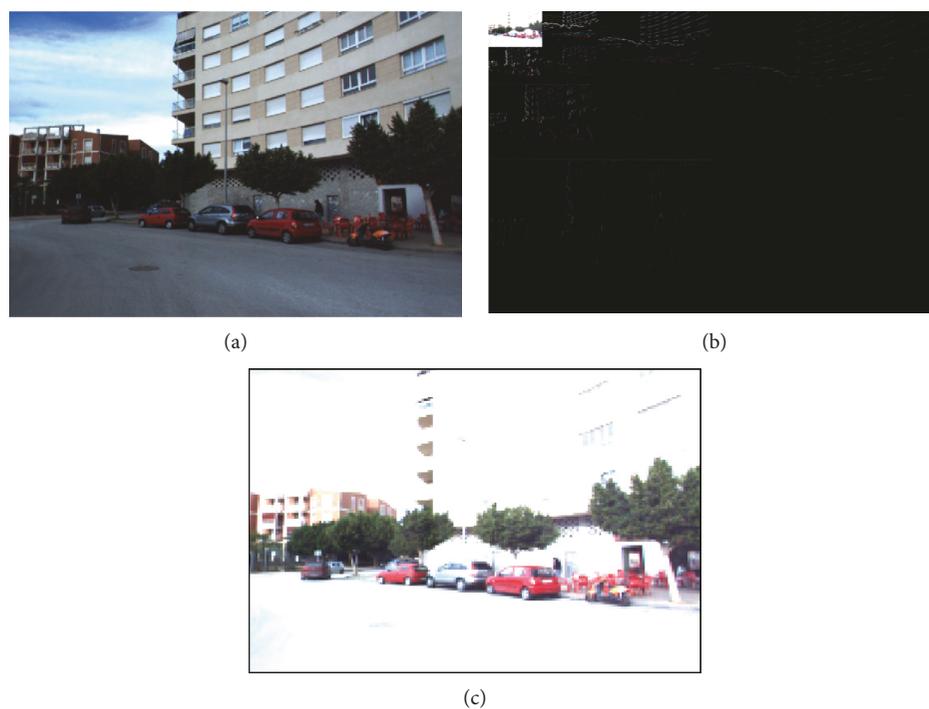


FIGURE 4: SMDWT results. (a) Input video frame (left view). (b) SMDWT result of the 3rd level decomposition. (c)  $LL_3$  subband image.

of this “variance-max” is considered as a moving object. A boundary is fixed for locating the moving object to point out the object in the present input frame. Finally, the depth of this object from the camera is calculated using the disparity value obtained and a decision can be taken depending on this depth value for driver assistance.

## 5. Results and Analysis

In this work, the algorithm is tested and evaluated using the data sets available on KITTI Vision Benchmark Suite [44] and DAIMLER pedestrian segmentation benchmark [45] using Intel® 2.10 GHz Core™ i3 2310M CPU, 4 GB RAM,

TABLE 3: Accuracy comparison of moving object detection using SMDWT, LDWT, and a median filter.

Input video	Decomposition level/resolution	SMDWT	LDWT	Median filter
Video 1	LL <sub>1</sub> (320 × 440)	99.62	99.62	99.01
	LL <sub>2</sub> (160 × 220)	99.01	99.01	92.03
	LL <sub>3</sub> (80 × 110)	94.83	94.83	87.13
Video 2	LL <sub>1</sub> (160 × 220)	99.23	99.23	99.05
	LL <sub>2</sub> (80 × 110)	94.95	94.95	92.03
	LL <sub>3</sub> (40 × 55)	93.82	93.82	84.15

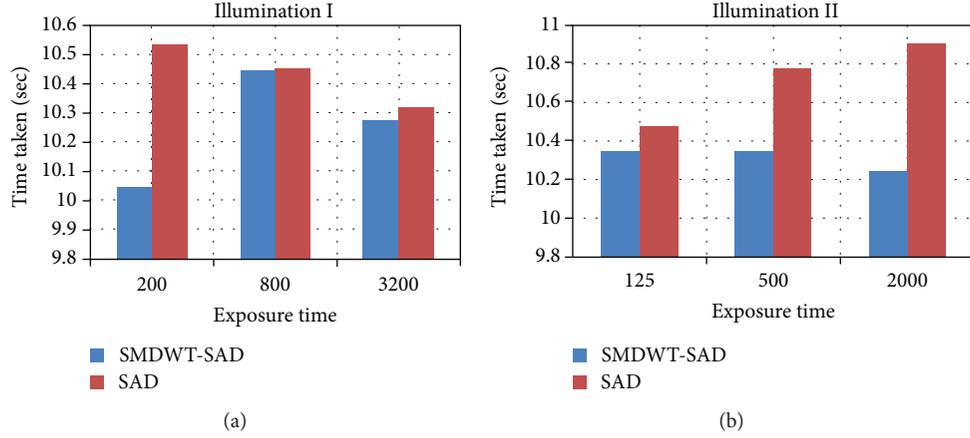


FIGURE 5: RMS error comparison of SAD and SMDWT-SAD under (a) illumination I (b) illumination II.

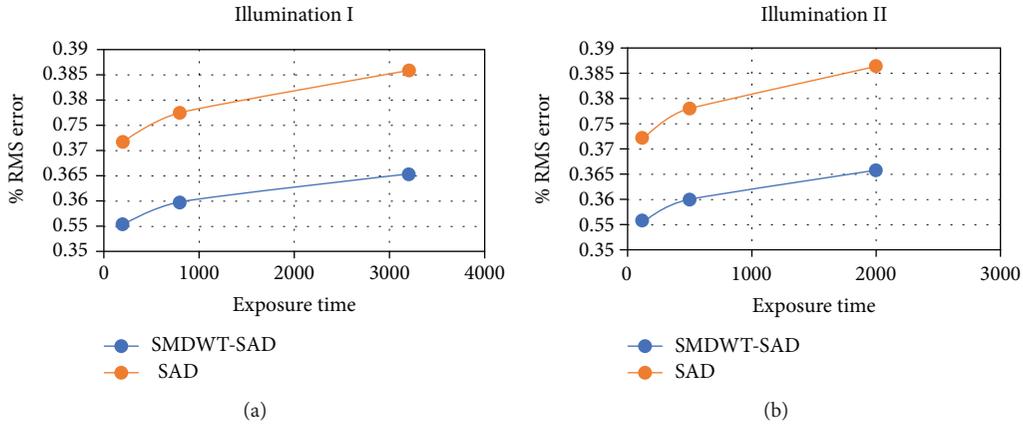


FIGURE 6: Comparison between computational time requirements for disparity computation of SAD and SMDWT-SAD under (a) illumination I (b) illumination II.

and MATLAB. The original image frame sizes are in the RGB system. Table 2 details the video frames used for the analysis. By applying SMDWT, every input image is compressed to a size of  $80 \times 110$ . For example, for the input video of size  $320 \times 440$ , this work uses the 2nd level decomposition and for the video size of  $640 \times 880$ , the 3rd level decomposition was used. Even though the increase in the level of decomposition gives faster results, it will decrease the accuracy. There always exists a trade-off between accuracy and speed.

**5.1. Result and Analysis of SMDWT.** By applying SMDWT, the noise and fake motions in the input video sequence are removed. Figure 3 shows the generated motion mask using the combined BS and FD method and LL<sub>1</sub> band of SMDWT. From the Figure 3(c), it is clear that the generated motion mask failed to retain the shape of the object. This can be overcome by deploying LL<sub>3</sub> subband instead of LL<sub>1</sub>, since LL<sub>3</sub> subband can retain the moving objects more effectively. Figure 4 shows the third level decomposition result of SMDWT. Figure 4(a) shows the left view of input video

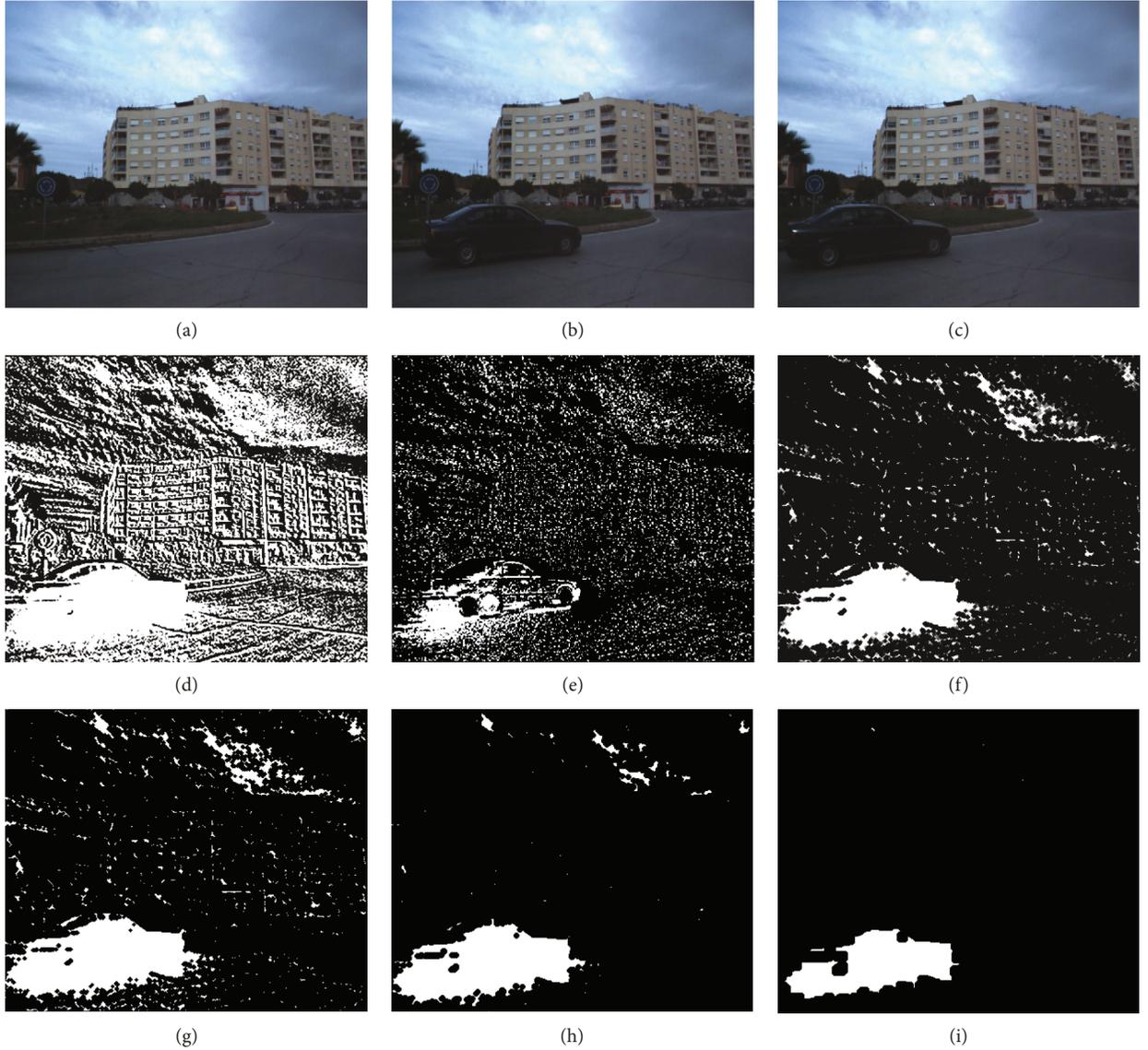


FIGURE 7: Generation of motion mask. (a) Reference frame used. (b, c) Image of two consecutive frames from the left view. (d) Result of BS. (e) Result of FD. (f) Result of the combined BS and FD method. (g) Motion mask (LL<sub>1</sub>). (h) Motion mask (LL<sub>2</sub>). (i) Motion mask (LL<sub>3</sub>).

frame, Figure 4(b) shows the 3rd level decomposition result of SMDWT, and Figure 4(c) shows the LL<sub>3</sub> subband image. Object detection performance, expressed in terms of precision rate, can be calculated using the following equation:

$$P_R = \frac{\text{No. of object detected}}{\text{Total no. of objects in frame}} * 100. \quad (12)$$

Table 3 illustrates the accuracy comparison of moving object detection using SMDWT, LDWT, and a median filter. It can be seen that the system attains a maximum detection accuracy of 99.62% and minimum accuracy of 94.95% for the decomposed input image size of  $80 \times 110$ , which is higher than the median filter.

**5.2. Results of SMDWT-SAD Correspondence.** The accuracy of SMDWT-SAD correspondence is tested by using the rectified stereo image pairs provided by Middlebury stereo data

[46]. The images are first preprocessed by SMDWT, and then, the SAD correspondence is done on images taken under different exposure times and different illumination conditions. Here, the maximum disparity level is taken as one-third of the maximum value of ground truth disparity [47].

**5.3. Performance Evaluation.** The performance of stereo algorithm and the quality of results can be evaluated by varying certain parameters [48]. Root-mean-squared error (RMSE) can be calculated using the following equation.

$$\text{RMSE} = \left( \frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{1/2}, \quad (13)$$

where  $d_C(x,y)$  is the obtained disparity,  $d_T(x,y)$  is its ground truth value, and  $N$  represents the total number of pixels in the disparity map.



FIGURE 8: Motion tracking results. (a–c) Tracking results using video 1 of different frames. (d–f) Tracking results using video 2 of different frames.



FIGURE 9: Results of ROI-based motion tracking using video 1.

Figure 5 shows the comparison of RMS error in the disparity map obtained from SMDWT-SAD and SAD, using the input images taken under different exposure times and different illumination conditions. Figure 6 shows their

corresponding comparison between the computational time requirements for disparity computation. From the comparison, it is evident that SMDWT-SAD is more accurate when compared to the conventional SAD correspondence. Also,

TABLE 4: Comparison of SMDWT, LDWT, and a median filter based on speed and tracking accuracy.

Input video	% accuracy			Time frame (sec)		
	SMDWT	LDWT	Median filter	SMDWT	LDWT	Median filter
Video 1	62.01	62.01	52.01	34.46	65.54	70.43
Video 2	73.33	73.33	64.34	32.38	57.02	63.25
Video 3	79.47	79.47	67.04	35.55	58.63	64.33
Video 4	65.21	65.21	50.36	30.48	63.46	70.64
Video 5	86.40	86.40	74.8	35.23	59.87	65.48
Video 6	82.74	82.74	72.32	35.74	56.54	62.11

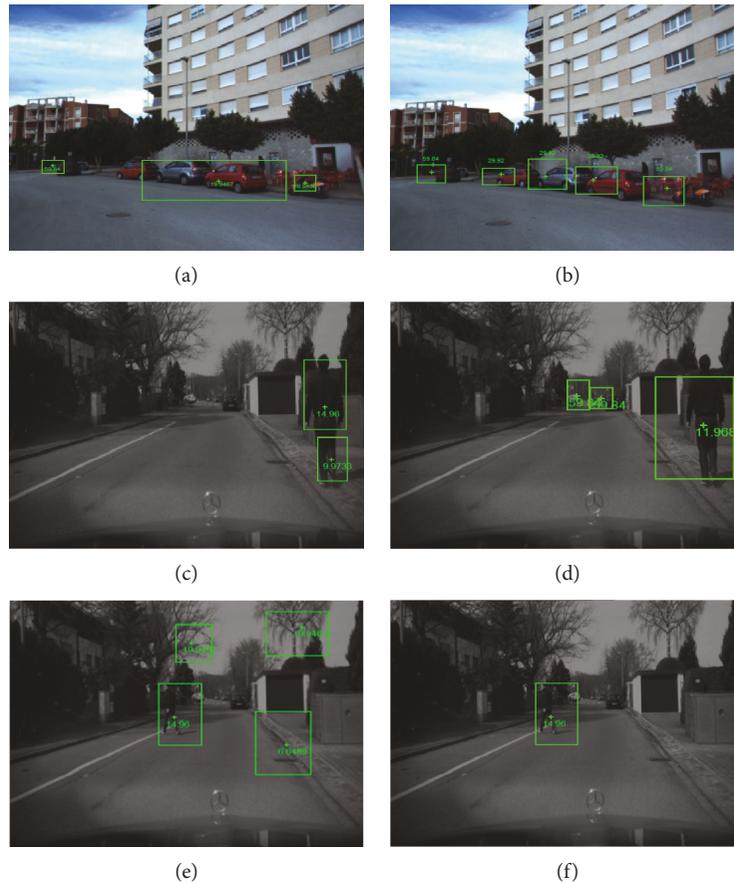


FIGURE 10: False detection avoidance using SMDWT. (a) False detections due to close and occluded vehicles. (b) Corrected result. (c) Object fragmentation due to fake motions and bad illumination. (d) Corrected result. (e) False detections and erroneous tracking. (f) Corrected result.

the time taken for the SMDWT-SAD method is much less compared to SAD. Thus, this novel SMDWT-SAD method is more accurate and fast for illumination invariant disparity computation in a stereo-based object tracking approach.

**5.4. Results of Object Tracking.** The proposed system is tested and analyzed qualitatively and quantitatively using various input video frames. Figures 7(a)–7(i) show the generation of motion mask using the combined BS and FD method and SMDWT.

Figure 8 demonstrates motion tracking results. Here, Figures 8(a)–8(c) show the tracking results of three consecutive frames of video 1. Figures 8(d)–8(f) show the tracking results of different frames in video 2. In this, the value

TABLE 5: Comparison of tracking accuracy.

Input video	SMDWT variance method	SMDWT mean method
Video 1	62.01	58.91
Video 2	73.33	71.33
Video 3	79.47	62.63
Video 4	65.21	59.78
Average tracking accuracy	<b>70.05</b>	63.16

TABLE 6: Tracking accuracy comparison with the existing methods.

Input video	Monocular LSK [49]	Stereo LSK [29]	Particle filter [51]	$\ell_1$ method [50]	Proposed method
Video 1	53.13	56.33	05.55	09.94	<b>62.01</b>
Video 2	69.62	67.37	61.20	59.01	<b>73.33</b>
Video 3	75.58	75.54	64.40	15.58	<b>79.47</b>
Video 4	28.81	51.87	19.38	05.42	<b>65.21</b>
Average tracking accuracy	56.7	62.77	37.63	22.48	<b>70.05</b>

indicated at the centre of the boundary boxes shows the disparity value at that point.

Figure 9 shows the results of ROI-based motion tracking in video 1. Here, the red circle indicates moving ROI. Table 4 shows the comparison of SMDWT, LDWT, and a median filter based on speed and tracking accuracy. The video 5 attains the maximum precision rate of 86.4% within a time frame of 35.23 seconds. The lowest accuracy of the system is 62.01% within a time frame of 34.46 seconds. From this, it is evident that even though the tracking accuracy of SMDWT and LDWT is the same, SMDWT is faster than LDWT.

**5.5. Qualitative Analysis.** Without using SMDWT, there are many false detections due to close and occluded vehicles, which are shown Figure 10(a). By incorporating SMDWT, these issues are eliminated as shown in Figure 10(b). In Figure 10(c), the false detection errors occurred due to fake motions and bad illumination conditions of video 2, i.e., even though the objects under motion are tracked, they are fragmented. Incorporation of  $LL_2$  subband image avoids this false detection, which is shown in Figure 10(d). Many false detections and erroneous tracking occurred in Figure 10(e) are eliminated using the SMDWT- $LL_2$  subband and variance method which is shown in Figure 10(f).

**5.6. Quantitative Analysis.** Table 5 demonstrates the results obtained from the quantitative analysis of our system. The motion tracking accuracy of SMDWT variance and SMDWT mean is computed and compared for various video sequences. The system attains a tracking accuracy (TA) of 79.47% by using the SMDWT variance method. Compared with the SMDWT mean tracking method, the SMDWT variance method is more accurate.

**5.7. Comparison with the Existing Method.** Table 6 shows the tracking accuracy comparison of the proposed system with other existing monocular and stereo methods. The monocular system detects moving objects based on colour histogram and local steering kernel for the object tracking [49]. Stereo LSK [29] uses a disparity histogram for object tracking. The particle filter method incorporating the appearance-based models has been used for visual tracking [50]. In the  $\ell_1$  method [51], the particle filter approach is used for visual tracking. From the comparison table, it is evident that our SMDWT dense disparity-variance method is more accurate than the other methods. The tracking accuracy of our proposed method is higher for all the input video sequences compared to monocular and stereo approaches.

## 6. Conclusion and Future Works

This work presents a novel stereo vision-based moving object tracking system for dynamic environments. This fast motion tracking approach employs SMDWT to deal with illumination changes and fake motions. In this, the reliable motion detection is achieved with combined background subtraction and frame differencing technique. The incorporation of the threshold level in this combined method can handle dynamic scene changes without frequently updating the reference frame. The dense disparity-based ROI selection increases the accuracy of the object detection. Apart from all the other features, the dense disparity-variance-based motion tracking improves the tracking accuracy and speed. The performance analysis proved the effectiveness of the system in terms of accurate object detection in dynamic scenes, fake motion avoidance, fast motion tracking, and accurate ROI selection. Thus, this fast illumination invariant system can be used for real-time motion tracking of dynamic scenes taken, even under bad environmental conditions.

Most of the motion tracking systems in real-time applications such as autonomous vehicles and video surveillance systems demand accuracy and speed simultaneously. As the demand for the driver assistance system in autonomous technology is increasing day by day, the incorporation of cognitive approaches in the vision technology will help to tackle the uncertainties and technical challenges in this area. An improvement over this system looks forward to a fast deep learning approach in dealing with object discrimination and occlusion handling. Another extended approach can be object tracking with the appearance-based object model using multiple stereo cameras. Such approaches will provide immense growth in the field of autonomous vehicle technology.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Y. Wang, B. Georgescu, T. Chen et al., "Learning-based detection and tracking in medical imaging: a probabilistic approach," in *Deformation Models*, pp. 209–235, Springer, Dordrecht, Netherlands, 2013.

- [2] Y.-C. Lim, J. Kim, C.-H. Lee, and M. Lee, "Stereo-based tracking-by-multiple hypotheses framework for multiple vehicle detection and tracking," *International Journal of Advanced Robotic Systems*, vol. 10, no. 7, pp. 1–13, 2013.
- [3] A. M. McIvor, "Background subtraction techniques," in *Image and Vision Computing*, vol. 4, pp. 3099–3104, Elsevier, 2000.
- [4] Y. Liu, H. Ai, and G.-y. Xu, "Moving object detection and tracking based on background subtraction," in *Proceedings SPIE 4554, Object Detection, Classification, and Tracking Technologies*, pp. 62–66, Wuhan, China, 2001.
- [5] M. Weng, G. Huang, and X. Da, "A new interframe difference algorithm for moving target detection," in *3rd International Congress on Image and Signal Processing (CISP)*, vol. 1, Yantai, China, 2010.
- [6] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.
- [7] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 266–280, 2000.
- [8] L. Wixson, "Detecting salient motion by accumulating directionally-consistent flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 774–780, 2000.
- [9] F.-H. Cheng and Y.-L. Chen, "Real time multiple objects tracking and identification based on discrete wavelet transform," *Pattern Recognition*, vol. 39, no. 6, pp. 1126–1139, 2006.
- [10] B. Sugandi, H. Kim, J. K. Tan, and S. Ishikawa, "Tracking of Moving Objects by Using a Low Resolution Image," in *Second International Conference on Innovative Computing, Information and Control (ICICIC 2007)*, pp. 408–408, Kumamoto, Japan, 2007.
- [11] C.-H. Hsia, J.-M. Guo, and J.-S. Chiang, "Improved low-complexity algorithm for 2-D integer lifting-based discrete wavelet transform using symmetric mask-based scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 8, pp. 1202–1208, 2009.
- [12] H. B. Resmi, V. A. Deepambika, and M. A. Rahman, "Symmetric mask wavelet based detection and tracking of moving objects using variance method," *Procedia Computer Science*, vol. 58, pp. 58–65, 2015.
- [13] V. Lippiello, B. Siciliano, and L. Villani, "Robust visual tracking using a fixed multi-camera system," in *IEEE International Conference on Robotics and Automation, ICRA'03*, vol. 3, pp. 3333–3338, Taipei, Taiwan, 2003.
- [14] C. Li, L. Lu, G. D. Hager, J. Tang, and H. Wang, "Robust object tracking in crowd dynamic scenes using explicit stereo depth," in *Asian Conference on Computer Vision- ACCV 2012*, pp. 71–85, Springer, Berlin, Heidelberg, Germany, 2012.
- [15] G. Catalin and S. Nedeveschi, "Object tracking from stereo sequences using particle filter," in *4th International Conference on Intelligent Computer Communication and Processing*, pp. 279–282, Cluj-Napoca, Romania, 2008.
- [16] L. Zhao and C. E. Thorpe, "Stereo-and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 148–154, 2000.
- [17] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. H. Matthies, "A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle," *International Journal of Robotics Research*, vol. 28, no. 11-12, pp. 1466–1485, 2009.
- [18] A. Barth and U. Franke, "Estimating the driving state of oncoming vehicles from a moving platform using stereo vision," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 560–571, 2009.
- [19] T.-N. Nguyen, B. Michaelis, A. al-Hamadi, M. Tornow, and M.-M. Meinecke, "Stereo-camera-based urban environment perception using occupancy grid and object tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 154–165, 2012.
- [20] A. Gaschler, D. Burschka, and G. Hager, "Epipolar-based stereo tracking without explicit 3d reconstruction," in *20th International Conference on Pattern Recognition*, pp. 1755–1758, 2010.
- [21] R. Muñoz-Salinas, M. García-Silvente, and R. Medina Carnicer, "Adaptive multi-modal stereo people tracking without background modelling," *Journal of Visual Communication and Image Representation*, vol. 19, no. 2, pp. 75–91, 2008.
- [22] L. Cai, L. He, Y. Xu, Y. Zhao, and X. Yang, "Multi-object detection and tracking by stereo vision," *Pattern Recognition*, vol. 43, no. 12, pp. 4028–4041, 2010.
- [23] T. Zhao, M. Aggarwal, R. Kumar, and H. Sawhney, "Real-time wide area multi-camera stereo tracking," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 976–983, San Diego, USA, 2005.
- [24] M. Muffert, D. Pfeiffer, and U. Franke, "A stereo-vision based object tracking approach at roundabouts," *IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 2, pp. 22–32, 2013.
- [25] C. G. Keller, M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnorr, and D. M. Gavrilu, "The benefits of dense stereo for pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1096–1106, 2011.
- [26] A. Wedel, A. Meißner, C. Rabe, U. Franke, and D. Cremers, "Detection and segmentation of independently moving objects from dense scene flow," in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 14–27, Springer, Berlin, Heidelberg, 2009.
- [27] P. David and U. Franke, "Efficient representation of traffic scenes by means of dynamic stixels," in *2010 Intelligent Vehicles Symposium (IV)*, pp. 217–224, San Diego, USA, 2010.
- [28] G. Andre, D. Burschka, and G. Hager, "Epipolar-based stereo tracking without explicit 3D reconstruction," in *2010 20th International Conference on Pattern Recognition*, pp. 1755–1758, Istanbul, Turkey, 2010.
- [29] O. Zoidi, N. Nikolaidis, A. Tefas, and I. Pitas, "Stereo object tracking with fusion of texture, color and disparity information," *Signal Processing: Image Communication*, vol. 29, no. 5, pp. 573–589, 2014.
- [30] H. Sun, H. Zou, S. Zhou, C. Wang, and N. El-Sheimy, "Surrounding moving obstacle detection for autonomous driving using stereo vision," *International Journal of Advanced Robotic Systems*, vol. 10, no. 6, p. 261, 2013.
- [31] Y. Zeng, Y. Hu, S. Liu et al., "RT3D: Real-Time 3-D Vehicle Detection in LiDAR Point Cloud for Autonomous Driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3434–3440, 2018.
- [32] R. Danescu, S. Nedeveschi, M. M. Meinecke, and T. Graf, "Stereo-vision based vehicle tracking in urban traffic

- environments,” in *2007 IEEE Intelligent Transportation Systems Conference*, pp. 400–404, Hawaii, United States, 2007.
- [33] M. Allogri, A. Broggi, D. Giaquinto, M. Patander, and A. Prioletti, “Machine learning in tracking associations with stereo vision and lidar observations for an autonomous vehicle,” in *2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 648–653, Gothenburg, Sweden, 2016.
- [34] B. Alexander, D. Pfeiffer, and U. Franke, “Vehicle tracking at urban intersections using dense stereo,” in *3rd Workshop on Behaviour Monitoring and Interpretation, BMI*, pp. 47–58, Karlsruhe, Germany, 2009.
- [35] Y.-C. Lim and M. Kang, “Stereo vision-based visual tracking using 3D feature clustering for robust vehicle tracking,” in *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics*, vol. 2, pp. 788–793, Vienna, Austria, 2014.
- [36] L. Chen, L. Fan, G. Xie, K. Huang, and A. Nuchter, “Moving-object detection from consecutive stereo pairs using slanted plane smoothing,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3093–3102, 2017.
- [37] T. Kowsari, S. S. Beauchemin, and J. Cho, “Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost,” in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1255–1260, Washington, DC, USA, 2011.
- [38] K. Andra, C. Chakrabarti, and T. Acharya, “A VLSI architecture for lifting-based forward and inverse wavelet transform,” *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 966–977, 2002.
- [39] I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting steps,” *Journal of Fourier Analysis and Applications*, vol. 4, no. 3, pp. 247–269, 1998.
- [40] T. Acharya and C. Chakrabarti, “A survey on lifting-based discrete wavelet transform architectures,” *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 42, no. 3, pp. 321–339, 2006.
- [41] C.-H. Hsia, J.-M. Guo, J.-S. Chiang, and C.-H. Lin, “A novel fast algorithm based on SMDWT for visual processing applications,” in *2009 IEEE International Symposium on Circuits and Systems*, pp. 762–765, Taipei, Taiwan, 2009.
- [42] V. A. Deepambika and A. Rahiman, “Stereo correspondence algorithms for robotic applications under ideal and non ideal lighting conditions,” *International Journal of Image Processing (IJIP)*, vol. 9, no. 3, pp. 89–101, 2015.
- [43] A. Fusiello, E. Trucco, and A. Verri, “A compact algorithm for rectification of stereo pairs,” *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.
- [44] <http://www.cvlibs.net/datasets/kitti/>.
- [45] [http://www.gavrilina.net/Datasets/Daimler\\_Pedestrian\\_Benchmark\\_D/daimler\\_pedestrian\\_benchmark\\_d.html](http://www.gavrilina.net/Datasets/Daimler_Pedestrian_Benchmark_D/daimler_pedestrian_benchmark_d.html).
- [46] <http://vision.middlebury.edu/stereo/data/>.
- [47] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, pp. 1–8, Madison, USA, 2003.
- [48] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1/3, pp. 7–42, 2002.
- [49] O. Zoidi, A. Tefas, and I. Pitas, “Visual object tracking based on local steering kernels and color histograms,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 5, pp. 870–882, 2013.
- [50] X. Mei and H. Ling, “Robust visual tracking using  $\ell 1$  minimization,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 1436–1443, Kyoto, Japan, 2009.
- [51] S. K. Zhou, R. Chellappa, and B. Moghaddam, “Visual tracking and recognition using appearance adaptive models in particle filters,” *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1491–1506, 2004.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

