International Journal of Current Engineering and Technology
©2017 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Improving Context Enhanced Object Tracking and Road Surface Information Analysis using Computer Vision

Chala Simon* **and Shilpa Gite**

Department of Computer Science and Engineering, Symbiosis International university, Pune, India

Received 15 Sept 2017, Accepted 19 Nov 2017, Available online 20 Nov 2017, **Vol.7, No.6 (Nov/Dec 2017)**

Abstract

A scene is a real-world view of the environment that contains different surfaces and objects, organised in a meaningful way. Since low-level features obtained from the video stream are insufficient and limited to describe the scene, it's difficult for the classifier to identify the objects on the scene due to less information about the image. Image fusion is the procedure of combining multiple image information into one to produce more steady and useful information. Using high-level semantic descriptor can also help the classifier to perform classification easily. The aspects explored in this paper to use fused information of the low-level features along with a high-level semantic descriptor of an image sequence from vehicle dashboard-mounted camera to a better understanding of the scene. Here we have proposed to develop vision-based road awareness and object tracking for driving assistance purpose.

Keywords: *ADAS, Context-enhanced, Computer vision, feature extraction, multi-resolution analysis based image fusion*

1. Introduction

Advanced driver-assistance systems (ADAS) is part of the intelligent transportation system (ITS) which is developed to automate and also enhance vehicle systems for safety and better driving. Today, one of the fastest growing sectors in automotive electronics is driver assistance system (Riches, I, 2014). Lane detection, parking assistance, adaptive cruise control, detecting front vehicles and estimating their position, speed, and acceleration, traffic signs and light recognitions are systems under development of autonomous vehicles.

With the aim of increasing the level of driving automation in Intelligent transportation, it is important to address critical issues, including road monitoring for irregularities and visual information perception. Computer vision is an interdisciplinary field that deals with how the machines can be made for acquisition of high-level understanding from digital images or videos. Computer vision paired with different algorithms that can let the machine to recognise images, interprets, and learn and identify patterns.

The context would give more information about the surrounding environment like a person, location, activities of the object. Using the contextual information of the image along with the low-level information of the image will make the machine learning algorithms to perform well in image region classification and tracking.

This research aims to identify the road surface and tracking foreground objects for driving assistance purposes using computer vision towards the final goal of implementing vision-based driver assistance system.

2. Related Work

Road surface content analysis and foreground object tracking's are the main tasks within current problem formulation. There are many approaches stated in the literature which can solve the problem of Road surface content analysis and front object detection as well as tracking.

Road scene awareness is achieved in paper (Altun, M. *et al*, 2017) by using video frame obtained from single camera which is mounted on vehicle dashboard. Both Low level features and high-level descriptors are used to get better the segmentation and foreground object detection. An adaptive maximum likelihood classifier selects road surface regions and Kalman filter is used to track the object.

In (Yan, Gang, *et al*, 2016), hypothesis generation and hypothesis verification are proposed to detect a real-time vehicle. The knowledge-based method uses features of the vehicle's appearance, such as its shadow, edge, texture, symmetry, colors and vehicle lights are used in hypothesis generation. According to this paper, Hypothesis is generated using pre-knowledge shadows underneath the vehicles. In Hypothesis verification, HOG feature was applied to train the SVM classifier and the AdaBoost classifier

*Corresponding author's ORCID ID: ORID ID: 0000-0002-9993-4529 individually. By comparing the result between two classifiers, they concluded that the AdaBoost classifier performed better regarding accuracy and runtime. The accuracy of classification was (97.24%).

(Huang, Deng-Yuan, *et al*, 2017), This paper also uses hypothesis generation and hypothesis like (Yan, Gang, *et al*, 2016) but it uses the C-SVM classifier to classify front vehicles and non-vehicles. They can achieve up to 94.08%, in different scenes of urban/suburb using a single camera In (Fang, Chiung-Yao, *et al*, 2013), to detect and track vehicles vehicle shadows (Otsu's method) and horizontal edges for Day times and vehicle tail light features (Cr component of the YCrCb color model and the hue component of the HSI color model) for night time. The result they have got was during night time, the precision of vehicle detection is similar, but the recall result is lower with daytime.

Paper (Li, Xin, *et al*, 2010), used Kalman filter motion model to achieves efficient tracking of multiple moving objects under the confusing situations. Kalman filter predicts object's position by using current object information. They reduce search scope and search time of moving the object to achieve fast-tracking.

Paper (Venkateshkumar SK, *et al,* 2015) proposes a speculation of HPMs that gets rid of the requirement for having root filters for non-leaf nodes by regarding them as latent variables inside a Dynamic Programming based optimization scheme.

From a large number of research papers on vision-based driver assistance system, there are only limited numbers of papers works with feature fusion and combination with temporal attributes.

Table 1: Combination of temporal analysis and feature fusion in prior papers

Papers	Approach	Information
(Yan, Gang, *et al*, 2016)	Pre-knowledge shadows to detect front vehicle	Spatial
(Huang, Deng-Yuan, *et al*, 2017),	Shadows and horizontal edges for vehicle detection	Spatial
(Fang, Chiung-Yao, *et al*, 2013)	Shadow and edge feature combined to detect the vehicle, color model to segment road into a different region. And also, the hue, saturation and intensity used to vehicle detection during night time	Spatial and spectral
(Li, Xin, *et al*, 2010)	To detect moving object using shape and pixel feature	temporal and spatial
(Venkateshkumar SK, *et al,* 2015)	Color and pixel-based similarity used to segment road scene	temporal and spatial
(Palaniappan, Kannappan, *et al.,* 2010)	Features fusion such as gradients, eigenvectors and temporal are used to detect vehicle	Spatial and temporal
(Miksik, Ondrej, *et al*, 2013)	Optical flow and a pixel-based similarity metrics are utilised to segment road scenes.	Temporal and spectral
(Yang, Yang, *et al*, 2009)	Optical flow for traffic analysis using a fixed camera directed at an intersection.	Temporal and spatial
(Alvarez, *et al*, 2012)	Texture descriptor which is based on a combination of color plane and local binary patterns are designed for road scene segmentation	Spatial and spectral
(Bolovinou A, *et al*, 2013)	Road scenes are classified by motion boundary histograms which used large feature vectors	Spatial and temporal
(Fernández, Carlos, *et al.* 2015)	Detecting road segment is detected using texture anisotropy feature	Spectral and spatial

Paper (Bolovinou A, *et al*, 2013; Fernández, Carlos, *et al.* 2015) presents that Extraction of the texture and motion information and segmentation of an image which is based on feature vectors obtained for all pixels uses structural tensors. These are parametric models, and they require apriori information such as some segments and class means. Such model requires the number of segments and class means information in prior. This a reason why unsupervised clustering is difficult. Commonly, these models applied to only segment the image into a small number of regions.

In this paper, we are going to use the spatial, spectral and temporal information which is used by (Altun, M. *et al*, 2017). Fusion method of these information's is based on human perception model. Visual stimulus is made from detected contours from colors, intensity, texture, and motion (Altun, M. *et al*, 2017). A map draws attention is one that has combined the output of these visual features that are sufficiently different from its surrounding components. Since the information we are going to use is image feature/ descriptor, better feature level fusion algorithm will provide quality of information and less noise in the fused image formed.

3. Proposed Work

3.1 Feature Extraction

Features are the information extricated from images in the way of numerical values that are difficult to comprehend and associate by a human. The information extracted from images are called feature or sometimes called descriptor. Our feature extraction methods are based on human perception model.

In human perception model, visual stimulus framed by the retinal receptor layer by the end goal to yield better division and scene understanding by using every one of the three visual data prompts: spatial, spectral and temporal. Human eyes have a photoreceptor in its retina which is sensitive to intensity and color. So that, human eyes it can able to cluster comparative hues and intensities in unconstrained vision which can help to draw an object contour to sketch the object. The human eye can also simply identify the appearance and changes in texture. Textures also clustered as color and intensity, and Human eyes can detect the boundaries between different textures. Accordingly, the visual stimulus will be made from detected contours from colors, intensity, texture, and motion. The combined yield of these visual highlights frames a guide demonstrating where a large portion of human consideration centers in a scene. Human attention can also be caught by tracking neurons in the visual system that fire upon detection of motion. Like the Human eye, computer vision methods can also extract color, intensity and motion information.

Spatial information

An image is composed of small elements called pixels. Every pixel compares to anyone value called pixel intensity. The intensity of an image varies according to the location of a pixel. Suppose I am an image on (x,y) position of any pixel, then I(x,y) can be represented as an image with a given function and where x and y's are integers. Along these lines an image I(x,y) is a matrix of pixels. These matrices are a measurement of the intensity of the color components. Spatial information's are color and intensity based feature which is acquired by analysing the video frames. In the extraction of spatial information from a video frame, the first step is identifying the pixels and clustering them based on their color. Local variance is also spatial descriptor which is solid contour descriptor. Road surface and foreground objects video frame will be clustered into different colors, and the edge of each object will be identified.

Spectral information

In spectral information extraction, we are going to find textures by using energy levels in local regions. Image texture is used as a description for region segmentation. Image energies can be calculated in different methods. In this paper, two energy transform methods will be used. Such as which helps to separate the image into parts of differing importance (concerning the image's visual quality) called Discrete Cosine Transform (DST) and which transforms a signal or image from the spatial domain to the frequency domain called Discrete Fourier Transform (DFT). Because by using this method, the calculation will be easy and fast (Altun, M. *et al*, 2017).

Temporal information

The pattern of evident movement of image protests between two continuous frames caused by the object or camera is called ***optical flow***. The optical stream is an intense apparatus for identifying movement in a video. Optical flow calculation is used to portray and measure the movement of objects in a video stream, frequently for object tracking system. Therefore, by computing the optical flow of video frame, temporal analysis can be achieved.

3.2 Information Fusion

Image fusion is a prevalent decision for different image enhancement applications, for example, overlay of two image items, refinement of image resolutions for arrangement, and image mix for feature extraction and target acknowledgement. Image fusion is comprehensively categorized into three types of given the image representation stage in which combination happens: (Nirmala, *et al*, 2015).

a. Pixel level
   - Pixel level is identified with the pixel area which joins the input image visual data into the single image in light of the first-pixel area.

b. Feature level
   - Feature level fusion is a combination of feature sets relating to multiple modalities which are region-based fusion approach.

c. Decision level
   - Decision level fusion is combination of classification results of previous classification stages.

Feature level combination requires the first extraction of the feature; those features can be distinguished by attributes, for example, contrast, color, size, shape, and texture (Nirmala, *et al*, 2015). Since such characteristics are going to be used in our system, the feature level fusion can be the best suitable stage in this system.

To achieve feature level fusion, we are going to use multi-resolution analysis based image fusion. Multi-resolution methods, such as pyramid transformation, have been implemented for data fusion since the early 1980s (Jiang, Dong, *et al,* 2011). Later, Discrete wavelet transform (DWT) is developed which has been successfully used in many fusion schemes (Lewis, J. J., *et al*, 2004). The advantages of DWT over pyramid transformation,

- more guiding information,
- no blocking artifacts,
- better signal-to-noise ratios and,
- improved perception model.

The problem with DWT is, sub-sampling occurs at each level will make it shift variant nature. The Dual-Tree Complex Wavelet Transform (DT-CWT) can deliver both good shift invariance and directional selectivity over the DWT, although there are an increased memory and computational cost. Therefore, this paper is aimed to use Dual-Tree Complex Wavelet Transform (DT-CWT) fusion algorithm.

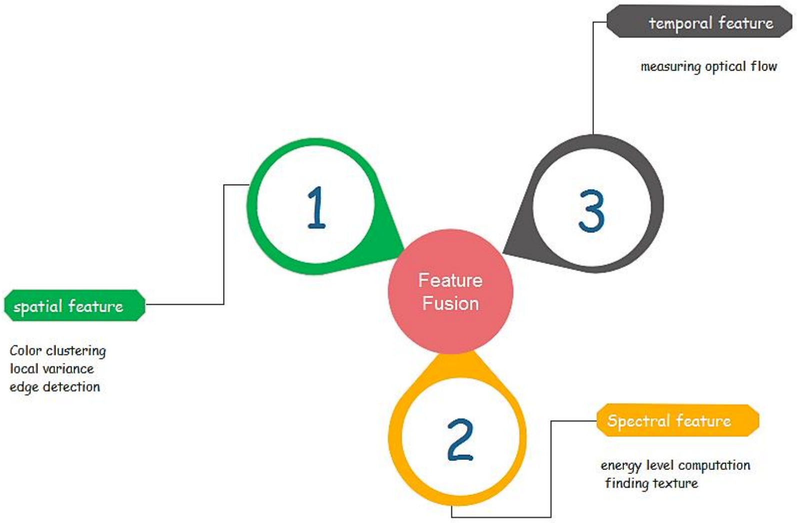

Figure 1: Feature level information fusion

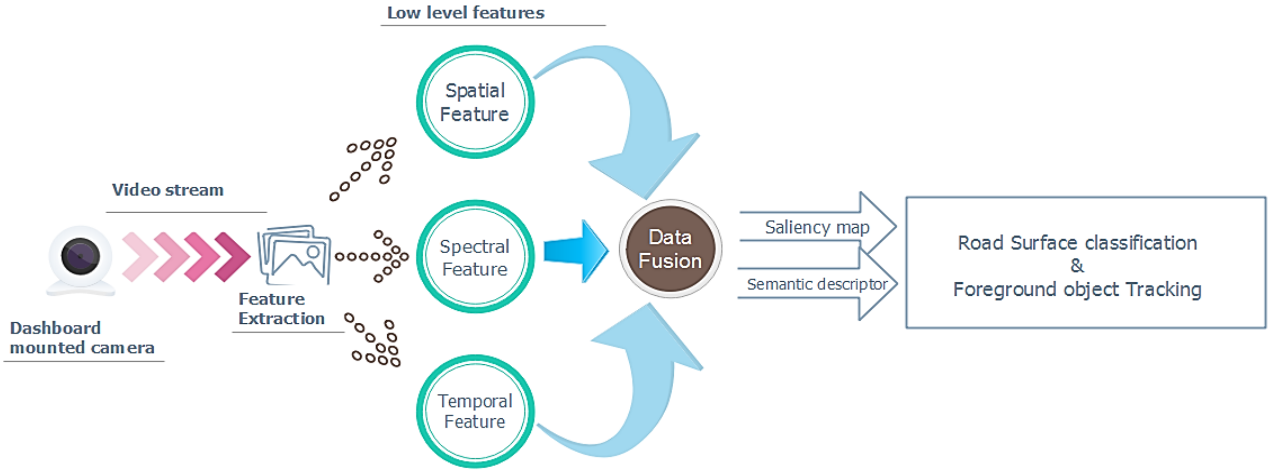

Figure 2: Overall architecture of the proposed system

3.3. Road surface identification and foreground object tracking

Image region segmentation can be performed by using contours in obtained saliency map. The high-order description is obtained from detected image segments, and it's used by machine learning algorithm to have better scene understanding.

Several papers have been published by many researchers to classify the road surface and foreground object detection like vehicles and pedestrian. However, Bayesian learning and inference algorithm is more compatible human learning than any other learning (Bishop CM. *et al*, 2006). Similarly, Kalman filter, which is based on the Bayesian model, have high potential to track the fast-moving object. (Li, Xin, *et al*, 2010). Kalman also can fill the missed data, so it can easily handle the problem of the missing object scene in the video frame while tracking.

Therefore, this paper uses the Bayesian-based model to classify the road region and track foreground objects from salience image formed alongside the high-level image descriptor.

The classifier uses ontological features which include properties such as shape, position, and color for better classification. Since more contextual information of the image is provided alongside with the high-quality image contour for the classification algorithm, road and non-road regions, as well as the foreground objects like vehicles and pedestrians can be detected with high accuracy.

Conclusion and future work

We have presented computer vision approach to identify road surface and track foreground objects for driver assistance purpose. This can be achieved by using the video frame acquired from the camera which is mounted on the vehicle dashboard. The low-level features such as spatial, spectral and temporal information will be extracted from image sequence from the camera based on human perception model. Using multi-resolution analysis based image fusion, Dual-Tree Complex Wavelet Transform (DT-CWT) algorithm aggregates these low-level features to give high-level image descriptor and saliency map.

Bayesian learning and inference algorithm used to identify the road surface and Kalman filter can track foreground objects using saliency image and semantic descriptor provided. The saliency map and semantic descriptor of the image let the classifier algorithms to understand the scene easily. Another possible future work is bridging the semantic gap between low-level features and high-level descriptor using advanced machine learning algorithms which help the system to mimic human perception model highly.

References

Altun, M., & Celenk, M. (2017). Road Scene Content Analysis for Driver Assistance and Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*.

Yan, G., Yu, M., Yu, Y., & Fan, L. (2016). Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification. *Optik-International Journal for Light and Electron Optics*, *127*(19), 7941-7951.

Huang, D. Y., Chen, C. H., Chen, T. Y., Hu, W. C., & Feng, K. W. (2017). Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads. *Journal of Visual Communication and Image Representation*, *46*, 250-259.

Fang, C. Y., Liang, J. H., Lo, C. S., & Chen, S. W. (2013, April). A real-time visual-based front-mounted vehicle collision warning system. In *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2013 IEEE Symposium on* (pp. 1-8). IEEE.

Wang, H., & Cai, Y. (2015). Monocular based road vehicle detection with feature fusion and cascaded Adaboost algorithm. *Optik-International Journal for Light and Electron Optics*, *126*(22), 3329-3334.

Li, X., Wang, K., Wang, W., & Li, Y. (2010, June). A multiple object tracking method using Kalman filter. In *Information and Automation (ICIA), 2010 IEEE International Conference on* (pp. 1862-1866). IEEE.

Venkateshkumar, S. K., Sridhar, M., & Ott, P. (2015, December). Latent Hierarchical Part Based Models for Road Scene Understanding. In *Computer Vision Workshop (ICCVW), 2015 IEEE International Conference on* (pp. 115-123). IEEE.

Palaniappan, K., Bunyak, F., Kumar, P., Ersoy, I., Jaeger, S., Ganguli, K., ... & Seetharaman, G. (2010, July). Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video. In *Information fusion (FUSION), 2010 13th Conference on* (pp. 1-8). IEEE.

Yang, Y., Liu, J., & Shah, M. (2009, September). Video scene understanding using multi-scale analysis. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 1669-1676). IEEE.

Fernández, C., Izquierdo, R., Llorca, D. F., & Sotelo, M. A. (2015, September). A comparative analysis of decision trees based classifiers for road detection in urban environments. In *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on* (pp. 719-724). IEEE.

Miksik, O., Munoz, D., Bagnell, J. A., & Hebert, M. (2013, May). Efficient temporal consistency for streaming video scene analysis. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on* (pp. 133-139). IEEE.

Brox, T., Rousson, M., Deriche, R., & Weickert, J. (2010). Colour, texture, and motion in level set based segmentation and tracking. *Image and Vision Computing*, *28*(3), 376-390.

Alvarez, J. M., Gevers, T., LeCun, Y., & Lopez, A. M. (2012, October). Road scene segmentation from a single image. In *European Conference on Computer Vision* (pp. 376-389). Springer, Berlin, Heidelberg.

Bolovinou, A., Kotsiourou, C., & Amditis, A. (2013, July). Dynamic road scene classification: Combining motion with a visual vocabulary model. In *Information Fusion (FUSION), 2013 16th International Conference on* (pp. 1151-1158). IEEE.

Fernández, C., Izquierdo, R., Llorca, D. F., & Sotelo, M. A. (2015, September). A comparative analysis of decision trees based classifiers for road detection in urban environments. In *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on* (pp. 719-724). IEEE.

Malik, S. S., Shivprasad, B. J., Maruthi, G. B., Kuriakose, J., & Amruth, V. (2013, August). Feature level image fusion. In *Proc. Int. Conf., Emerg. Res. Comput., Inf., Commun. Appl.(ERCICA)* (pp. 566-574).

Kootstra, G., & Kootstra, G. W. (1978). Visual Attention and Active Vision From Natural to Artificial Systems.

Jiang, D., Zhuang, D., Huang, Y., & Fu, J. (2011). Survey of multispectral image fusion techniques in remote sensing applications. In *Image fusion and its applications*. Intech.

Riches, I. (2014). Strategy Analytics: Automotive Ethernet: Market Growth Outlook| Keynote Speech 2014 IEEE SA: Ethernet & IP@ Automotive Technology Day. *PDF). IEEE. Retrieved*, 11-23.

Li, L., Song, J., Wang, F. Y., Niehsen, W., & Zheng, N. N. (2005). IVS 05: new developments and research trends for intelligent vehicles. *IEEE Intelligent Systems*, *20*(4), 10-14.

Nirmala, D. E., & Vaidehi, V. (2015, March). Comparison of Pixel-level and feature level image fusion methods. In *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on* (pp. 743-748). IEEE.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Lewis, J. J., O'callaghan, R. J., Nikolov, S. G., Bull, D. R., & Canagarajah, C. N. (2004, June). Region-based image fusion using complex wavelets. In *Seventh International Conference on Information Fusion (FUSION)* (Vol. 1, pp. 555-562).