MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Visually Salient Features for Highway Scene Analysis

Anurag Singh
Center for Advanced Computer Studies
University of Louisiana at Lafayette, Lafayette, LA, USA
axs2573@cacs.louisiana.edu

Chee Hung Henry Chu
Informatics Research Institute and Center for Advanced Computer Studies
University of Louisiana at Lafayette, Lafayette, LA, USA
cice@cacs.louisiana.edu

Michael A. Pratt
W.H. Hall Department of Electrical and Computer Eng
University of Louisiana at Lafayette, Lafayette, LA, USA
mpratt@louisiana.edu

Abstract

In applications involving autonomous vehicle or camera assisted driving it is important to have a generic prior understanding of the highway scene. We present a visually salient feature-based approach for road image understanding. We use salient features for object localization, near-far distinction and urban-rural scene classification; these tasks have such applications as in adjusting a vehicle's speed. We empirically verify the efficacy and assess the performance of each task.

1. Introduction

There has been tremendous progress in the field of machine vision along with low-cost sensing and computing devices which together hold promising future for "smart" cars. Many modern cars today come with driver assistance facility which helps understanding the road and its environment for the driver. These capabilities enable services such as automatic or assisted adjustment of the vehicle's speed. In this paper we present machine vision tasks that contribute to the capabilities based on early identification of the focus of attention at a location. One of the early processes for finding the location of visual attention is visual saliency detection.

The first application of visual attention to an image acquired from a driver's viewpoint is to localize the attention on the road objects. There are many visually prominent components in a road scene but we focus on the objects occurring on the road region where line markers, shadows, other cars, and other obstacles are most important for detection. Salient features can highlight the car and markers as they stand out with respect to the road.

After localizing the object it is important to find out if the object is close to the camera. A typical solution is to use multiple cameras to deduce the distance of an object using stereo vision. Another solution is to use range sensing devices such as Lidar. We propose a solution based on a single image that uses linear regression on the saliency measure to predict closeness of the object to the vehicle or camera. The properties of perspective projections helps to easily differentiate a near object from a far object based on aggregated saliency value.

Awareness of the environment, such as whether the surrounding is a rural or an urban setting, is useful in generally setting the vehicle's driving mode. Salient feature works well for this purpose as an urban scene attracts different region compared to a rural scene.

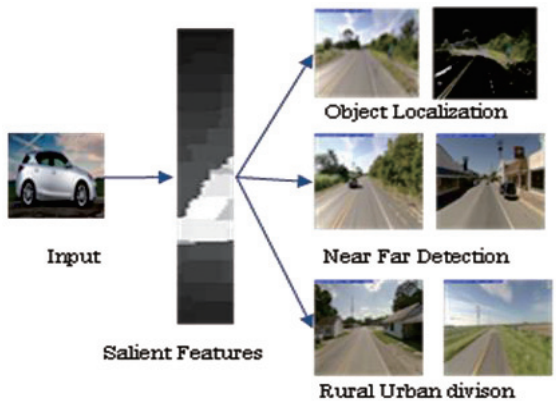

Figure 1. Highway scene analysis examples.

We use a collection of images acquired from a camera mounted on the front of a vehicle that traversed U.S. state-level highways through rural and urban environments. The overall view of the project is shown in Figure 1.

Visual saliency has been used for several highway applications such as automatic target detection, road sign detection, for capturing the gist of the scene [1,4], for off road obstacle detection. Object localization gives bounding boxes which are used for robust object detection. A saliency based bounding box was proposed by [10] where saliency generation is seen as a sampling problem.

Road environment classification was proposed in a recent work [7] where the histogram of oriented gradient (HOG) was used to classify a road image into urban and rural setting. HOG based classifiers have been successfully used in many scene recognition systems but we advocate in our work use of simple, early features. Early detection has the advantage of the usability of features.

The rest of this paper is organized as follows: we next briefly present a description of the saliency algorithm used to generate salient features. In section 4 a description is given in detail for the use of salient features to assist in highway scene analysis. Finally we conclude by describing the experimental methodologies and future works.

2. Salient Feature Detection

A pixel or, more usefully, a group of pixels is salient if it stands out from the others in terms of some pixel properties, such as color or texture. State-of-the-art methods include those that have a biologically motivated approach [11], those that are based in the frequency domain [12], and those that are comparison based [13]. In comparison based methods, the image properties of a pair of pixels or a pair of windows each around a pixel are compared. More recently, a so-called superpixel representation of an image is found by clustering pixels to form possibly irregularly sized and shaped superpixels. Superpixels are more responsive to edge intensity and also carry local information well. A comparison based saliency detection algorithm can then be applied to this representation by comparing the image properties of a pair of superpixels [14]. A superpixel representation reduces the computational cost for comparison based saliency detection methods.

In the following, we describe our saliency detection work that broadly follow the work in [13,14]. Nevertheless we note that the highway scene analysis methods are based on salient features that are independent of the choice of a saliency detection algorithm.

In a contrast-based method, two pixels or superpixels are compared based on the two essential properties of position and color. Consider the unit of comparison to be either a pixel or a superpixel. The distance between two pixels is trivially the Euclidean distance between them. When we consider the superpixel as a unit of comparison, the spatial position of each superpixel in image plane is given by calculating the mean position value of all the constituent pixels in the superpixel.

If the unit under consideration is a pixel, then the color difference can be the Euclidean distance between the pixel pair. If the unit is a window around a pixel or if the unit is a superpixel, we need to compare the color of two pixel groups. There are various ways to compare the color difference between two groups of pixels. These methods vary from average color [13], histogram based or the MPEG-7 color descriptors [14]. The scalable color descriptor and the dominant color descriptor can work well to compare two superpixels with different shapes and sizes. Due to the uniformity of pixels in a superpixel there is negligible deviation of each constituent pixel from the mean color value in each channel, the average value of color is used to represent the overall color value for a superpixel. We compute the color difference in the CIE L*a*b* color space which supports chromatic double opponency [3].

The general argument for using the color and position properties is as follows. A unit is salient if its color is different from all other units. The color difference between two units is inversely scaled by their positional distance for the reason that two units that make up a salient region would tend to be located close to each other. On the other hand, the fact that a unit has a large color difference from another unit that is far away from it is not sufficient evidence that the unit being considered is salient. A measure of dissimilarity between units p_i and p_j that supports saliency detection that follows this line of argument is:

$$d(p_i, p_j) = \frac{d_{color}(p_i, p_j)}{1 + \alpha d_{position}(p_i, p_j)} \qquad (1)$$

where $d_{color}(p_i, p_j)$ is the color difference between the units (pixels or superpixels), $d_{position}(p_i, p_j)$ is the position difference between the units, and α is a normalizing constant, set to balance the overall scales of the color and position differences. Equation (1) measures the difference of unit p_i relative to p_j. A measure of whether p_i is salience, denoted $S(p_i)$ can therefore be determined by

$$S(p_i) = 1 - \exp\left\{ -\frac{1}{N} \sum_{j=1}^{N} d(p_i, p_j) \right\}, \qquad (2)$$

where N is the number of units in the image. The saliency of a pixel or superpixel has a finite range, such as [0, 1]. In our work, we use a superpixel as the unit of comparison. Figure 2 illustrates the saliency map detection.

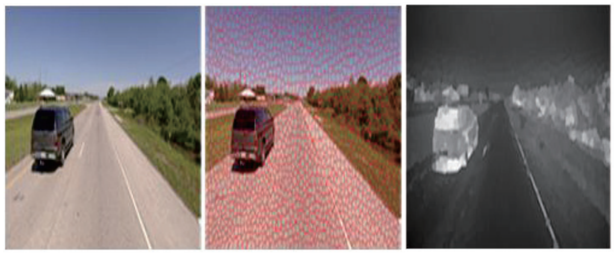

Figure 2. The saliency detection process. An input image (*left*) is over segmented to form a set of superpixels (*middle*). The saliency of each superpixel is determined to form a saliency map (*right*), in which the brighter the pixel the higher the saliency value is.

3. Highway Scene Analysis

In the following, we describe methods for highway scene analysis that are based on the saliency detection output, typically as a saliency map. We collected image data acquired from a camera mounted on the dashboard of a vehicle. We have generated PASCAL style annotation for the dataset to be made available. There are 100 images along with 50 images that have one or more cars in it. The algorithms for various highway analyses were implemented in OpenCV and C++. We use a superpixel representation of each image. The superpixel generation was done using a publicly available code by Veksler *et al.* [9]. Neural network implementation of WEKA [8] was used for classification purposes.

In a society that uses right-hand traffic, the road side scene in a highway image is defined by the portion of the image on the right side of the road which is below the horizon. This area ignores the sky. The road is detected using a vanishing point algorithm [5, 6]. The vanishing point algorithm gives a good approximation of the road region. In an urban setting it is often difficult to see the horizon due to occlusion. In such cases, it is suitable to only take the image portion below the vanishing point on the right hand side of the image. The left hand side of the road is not used as it can be occluded due to incoming vehicles. The right side of the road is used to describe a scene associated with the road image. Figure 3 illustrates

Figure 3. Road estimation and scene region extraction. In an input image (*left*), the traveling lane (*center*) can be identified to find the vanishing point, from which the road side image region can be estimated (*right*).

the process of road estimation and scene extraction.

3.1. Road Object

There can be many salient objects on a road image: cars, lane markers, shadows, etc. Of all of the objects the most interesting ones are other cars and the least interesting ones are shadows. The important characteristic of the object on the road is they can occur as a single object or a set of cluttered objects. Understanding the objects on road and their interaction has many applications especially for self-driven cars. We use visual saliency maps to assist in understanding objects on road by doing object localization and near-far object discrimination.

Object localization is an important step in object detection and recognition whether we are doing object cat-category recognition or individual object recognition. Visual saliency map are used to localize the salient pixels and using localized objects bounding boxes are made around them. Figure 4 shows an input sequence with corresponding saliency map, localized object and their bounding boxes.

3.2. Near-Far Discrimination

In such applications as those for self-driven cars there is a great amount of interest in finding how close or far the objects are on the road. Due to the perspective projections the objects nearer to the camera which in turn are nearer to the car are larger than ones that are far away. Figure 5 shows examples of near and far object as seen from camera mounted on the car.

We use saliency map generated to build a two class linear discriminant function which is used to differentiate near or far objects in a sequence of road images and the two class linear discriminant function is given by

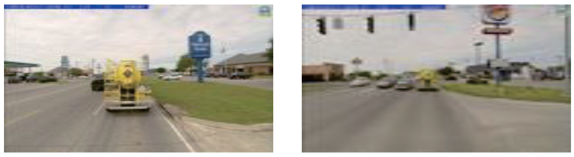

Figure 5. Examples of an image with a near object (*left*) and an image with a far object (*right*).

$$y(x) = wx + w_o \qquad (3)$$

where $y(x)$ is a two class variable representing near and far, x is the average aggregated saliency value for each image, w is the weight term and w_o is the bias. The aggregated

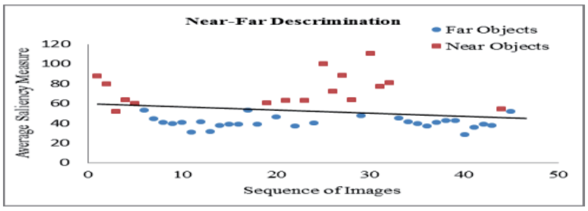

Figure 6. Near-far aggregated saliency plot road.

value of x is in the range of 22 to 110 with closer objects having higher aggregated value. This is because closer salient objects will have more salient pixels relative to other pixels in the road region.

Figure 6 shows the response of aggregated saliency values for different image sequences with image with far objects is denoted by blue dot and image with near objects is denoted by red dots. The black line represents the linear regression line dividing the near and far object. The overall accuracy is 95.5 % for this model tested on an expert-labeled sample training set of 50 images.

3.3. Scene Classification

Humans are very good at putting information in categories. It is normal for humans to drive by farm, rural village or an urban town by comparing the visual similarity to the places they have been before. They can seamlessly categorize the place into rural or urban place.

Automatic scene classification is a very important step in scene understanding. There are various methods of scene analysis [4] as like object based scene analysis, region based scene analysis, context-based scene analysis and biologically plausible scene recognition. For early recognition biologically possible scene recognition is advocated. We build a 2-class classifier using salient features to classify urban and rural scenes.

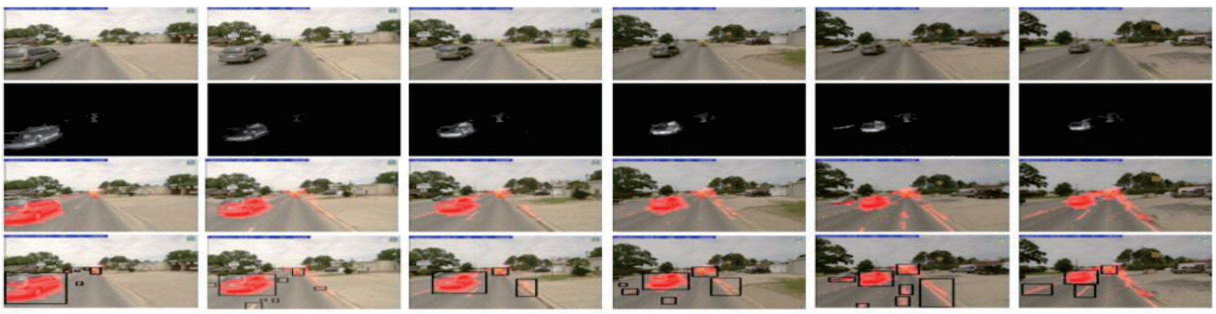

Figure 4. Road object localization for highway sequence. The top row shows input image, 2nd row shows saliency maps, 3rd row shows corresponding object localization, and last row shows bounding boxes.

a: Urban Scenes b: Rural Scenes

Figure 7. Rural-urban scene classification results.

Urban-Rural Classification: A training set of 100 highway images are collected with 50 urban and 50 rural images.

Feature Extraction: From this training set scene superpixels are extracted and their saliency values are computed using Equation 2. The saliency value associated with each superpixel is used as features. We consider the superpixels in the right roadside region (as illustrated in Figure 3). We collect the K highest saliency values from each image to form a feature vector, where K is set in our work to be the smallest number of superpixels in the training set, which in this case is 932.

Classification: For classification purpose WEKA's [8] implementation of multi-layer back prorogation algorithm is used. The final scene classification results into 89.6 % accuracy on training set of 100 images. How does dissimilarity vector of an urban scene differs from a rural scene? An urban scene typically has a lot of man-made structure which causes more pop out segments when compared to the saliency profile of a rural scene. This makes the dissimilarity vector for urban scene different from rural scene thus accounting for a higher accuracy in scene recognition. Figure 7 shows examples of training set, correctly classified and incorrectly classified.

4. Conclusions

We show through some applications that salient features are useful for early scene analysis. Our application domain is in highway images. To best of our knowledge use of contrast based salient features for highway scene analysis has not been robustly explored. Salient feature holds a promising use in autonomous vehicle and camera based assistance in vehicular applications.

Our ongoing work is in refining the methods. In the near future we plan to explore the use of salient feature for safety mechanism. Salient feature can be used to find distractions on highway like improper placement of sign boards, neon signs for advertisements competing for attention with road signs, etc.

Acknowledgments

The authors thank the reviewers for their suggestions. This work was supported in part by the Louisiana Board of Regents through grant no. LEQSF (2011-14)-RD-A-28.

References

[1] C.K. Chang, C. Siagian, and L. Itti, "Mobile robot vision navigation & localization using gist and saliency," in *IEEE IROS,* pp. 4147-4154, 2010.

[2] P. Federico, P. Krahenbuhl, Y. Pritch, and A. Hornung., "Saliency filters: Contrast based filtering for salient region detection," in *IEEE CVPR,* 2012, pp.733- 740, 2012.

[3] S. Engel, X. Zhang, B.Wandell, "Colour tuning in human visual cortex measured with functional magnetic resonance imaging," *Nature*, vol. 388, No. 6637, pp. 68-71, 1997.

[4] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," in *IEEE TPAMI* vol. 29, no. 2, pp. 300-312, 2007.

[5] K. Hui, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *IEEE CVPR*, pp. 96, 2009.

[6] T. Kuhnl, F. Kummert, and J. Fritsch, "Monocular road segmentation using slow feature analysis," in *IEEE Intelligent Vehicles Symposium (IV)*, 2011, pp. 800-806.

[7] D. Aubert, R. Brémond, A. Cord, E. Dumont, D. Gruyer, N. Hautière, P. Nicolle, J.-P. Tarel, V. Boucher, P. Charbonnier, P. Foucher F. Fournela, F. Greffier, V. Muzet, "Digital imaging for assessing and improving highway visibility," in *Proc. TRA*, 2014

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten ; "The WEKA Data Mining Software: An Update," in *SIGKDD Explorations*, Volume 11, Issue 1, 2009.

[9] O. Veksler, Y. Boykov, P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *ECCV*, pp. 211-214, 2010.

[10] P. Siva, C. Russell, T. Xiang, and L.Agapito, "Looking beyond the image: Unsupervised learning for object saliency and detection," in *IEEE CVPR*, pp. 3238-3245, 2013.

[11] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE TPAMI,* vol. 35, no. 1, pp. 185–207, Jan. 2013.

[12] J. Li, M.D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE TPAMI,* vol. 35, no. 4, pp. 996–1010, 2013.

[13] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE TPAMI,* vol. 354, no. 10, pp. 1915-1926, Oct. 2012.

[14] A. Singh, C.H. Chu and M.A. Pratt. "Multiresolution superpixels for visual saliency detection," in *IEEE CIMSIVP,* pp. 1-8, 2014.