A Novel Approach for Road Lane and Vehicle Detection

Soney R Nadh

Abstract— Applications with autonomous navigation of vehicles has attracted increasing attention in recent years due to the challenging scenarios it has to face while road scene analysis. The system introduced in this paper has a robust detection of road lane marking and vehicles. But challenges like large appearance variation in lane marking caused by occlusion, shadows and changing lighting conditions of the scene makes the problem harder to solve. The main objective of the system is to address these issues through some learning based approach using usual inputs captured from a camera mounted in front of a moving vehicle. In order to make the system more intelligent and automated vehicle detection using a new descriptor based on log-Gabor functions is included. Log-Gabor filter banks are proven to yield better results than Gabor filter banks.

Index Terms— Intelligent transportation, Boosting, context, hierarchical descriptor, log-gabor filters.

1. Introduction

Annually some 1.2 million people die worldwide as a result of traffic accidents. By maintaining an awareness of the road environment for deriver assistance the above rate can be reduced. Researches on sensing system for vehicle safety gives more better system for driver assistance. Researches and studies in computer vision for on-road safety have introduced monitoring the interior of vehicle, the exterior or both. This paper is focusing on monitoring the exterior of the vehicle. Monitoring the exterior can consist of estimating lanes, pedestrians,vehicles, or traffic signs. Taking a human-centred approach is integral for providing driverassistance; using the visual modality allows the driver to validate the systems output andto infer context. Many prior research studies monitoring the vehicle exterior address oneparticular on-road concern.

Within the last few years, research into intelligent vehicles has expanded into applications which work with or for the human user. Human factors research is merging with intelligent vehicle technology to create a new generation of driver assistance systems thatgo beyond automated control systems by attempting to work in harmony with a human operator.Systems which monitor the drivers state, predict driver intent, warn drivers of lane departures, and/or assist in vehicle guidance are all emerging. With such a wide variety of system objectives, it is important to examine how lane position is detected and measure performance with relevant metrics in a variety of environmental condition.

• Soney R Nadh, currently pursuing M.Tech Degree in Computer Science and Information technology from FISAT, Angamaly, Kerala, India
• E-mail: soneyrnadh@yahoo.co.in

The detection of road lane and vehicle is a hard problem due to variations present in the (a) the type of road in which the vehicle is traveling, (b) the appearance of lane and vehicle appearance (c) Time of scene analysis (at day time features of the road scene will be more clear than night time), and (d) the presence of shadows due to other objects. To deal with the above problems several approaches have been proposed in the literature.

2. RelatedWorks

The road scene analysis using image input is an active research area since last two decades. A complete survey with recent works is done by Mc Call and Trivadi[1], which provides a summary of existing approaches. Almost all of the works follow the following steps:-

(1). Extracting features to initialize the lane markings such as edges[2],texture[3], colour[4] andfrequency domain features[5];
(2). Post-processing the extracted features to remove outliers using techniques like Houghtransform[6] and dynamic programming[7], along with computational models explainingthe structure of the road using deformable contours[8], and regions with piecewiseconstant curvatures[9];

Several real time systems have been introduced recently. GOLD by Bertozzi and Broggi[10] is using a stereo-vision based hardware and software architecture for the detection. Y. Wang et.al. [11] proposed a B-Snake based lane detection and tracking algorithm without any camera parameters. F. Heimes and H.-H. Nagel[12] combined passive GPS and map based route guidance with model based machine vision. Also machine learning methods with a single classification boundary such as neural networks and support vector machines [13], have also been used for detection. As the visual properties of the lane markings change time to time, the use of local features to describe their appearance and learning the decision boundary from a single classifier may not be robust and scalable. The main contributions of this paper are:-

(1). A pixel hierarchy feature descriptor to model the context information of lane marking pixels.
(2). An outlier robust boosting algorithm to learn the relevant spatial features for detecting lane markings.
(3). A gabor-filter based vehicle detection.

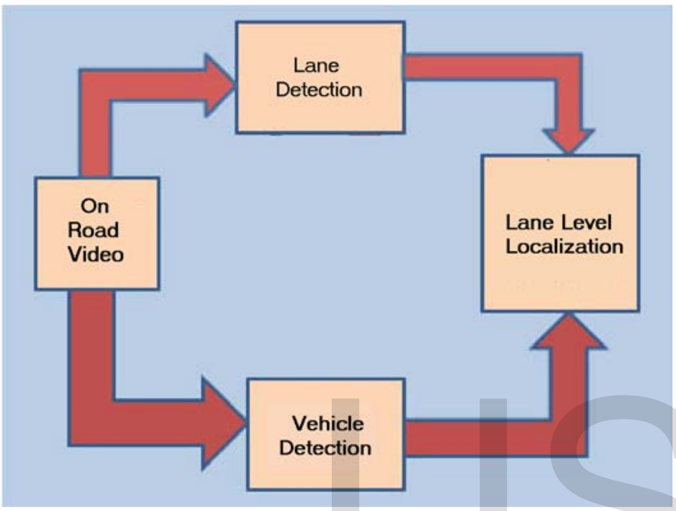

Figure 1 : Overview of this system

*Organization of this paper:*We discuss the detection of road lane marking by using boosting algorithm from the extracted spatial context information of the road scene in section 3. Section 4 includes the vehicle detection part, where a log-gabor filter based approach is used. The section 5 includes the experimental validation details and section 6 gives the conclusion. Figure 1 is block diagram for overview of the system.

3. Lane Marking Detection

Here is a learning-based approach to detect road lane markings and vehicles from a roadscene without requiring a predefined road model. A data-driven discriminative approach is used to classify lane markings O from non-lane markings O' .Given a set of M labelled training samples $\{x_i\}_{i=1}^{M} = 1$belonging to $O$and O', we first extract M_1contextual features from every x_i .Then train a classifier on these featuresusing a boosting-based machine learning algorithm, through which the detection of lane markings is performed on test images.

A. Feature Extraction

Since the lane markings share a rich neighbour-hood with the road regions understanding the role of context for this problem attains prominence. Classification of low-level contextual sources belongs to one of the following categories; 1) top-down methods that computethe gist of the scene by computing some global image statistics and 2) bottom-up methodsthat correlate the properties of a pixel with its immediate adjoining region. Both haverelative advantages/disadvantages depending on the application of interest. This system, propose a hierarchical descriptor that encodes information of both types.

Given a pixel x corresponding to O or O', considers a hierarchy of image regions represented by concentric circles cantered at that pixel. Let $R = \{R_i\}_{i=1}^{M_2}$ represent the regionsenclosed by circles of increasing radius. It uses the visual information present in R to obtain contextual measurements $h^x = f^x(I_o, I_s)$) for a pixel x, where $I_o$denotes the pixel x in isolation (i.e., the region enclosed by circle of zero radius, centred at x), I_s corresponds to the regions enclosed by all other circles of positive radius centred at x, and f represents a collection of filters that computes context between$I_o and\, I_s$. But the exact definition of visual information depends on its application. The basic contextual feature for a pixel x is computed as

$$h_j^x = f_j^x(I_o, I_s): R \times F \to R \qquad \forall j = 1\; to\; M_1 (1)$$

by analysing the pattern of F on different regions R. A set of rectangular Haar-like filtersare used here for this purpose. Typically, the number of features $M_1$is on the order of 1000s depending on the precision with which the positive and negative rectangular patterns of Harr filters are varied.

B. Training the Classifier

Now it require a principled way of selecting relevant features among , which are most discriminative in classifying pixels x corresponding to O from O'. This requirement paves the way to adapting the principles of boosting, a machine learning method that determines the optimal set of features to classify objects with provable detection error bounds.We know given a set of labeled training samples belonging to two classes with thesame initial weights and a pool of weak learners, the goal of boosting is to find a sequence of best weak classifiers by adaptively increasing (and decreasing) the weights of wrongly (and correctly) classified samples in each stage. Before getting into the details of detecting lane markings using boosting, let us address an important problem in the learning stage of boosting algorithms. Considering the basic version of the Adaboost algorithm [14] it is found that of the presence of outliers in the training set leads to error [15], i.e., when there are outliers present in the training set, for example, due to mislabelled samples or due to samples that are very different from other neighbours of their class, this process will result in a substantial increase in their weights and thereby force the weak learners to concentrate much more on these samples.

This might end up being detrimental to the performance of Adaboost.

So the observations made on the existing boosting algorithm are (1) all of them starts with equal initial weights for all the samples before learning the classifier, and (2) the error, which is minimized to select the weak learners, does not account for the unbalanced weight distribution of the samples. These two issues are addressed in this paper by modifying the boosting algorithm as follows given in Algorithm I.

Learning prior information about the training samples: Statistics is used here to prioritize the training data rather than assuming uniform initial weights for all the samples. Here we perform kernel discriminant analysis on the training data and analyse the distance of projection of samples with respect to their projected class means to determine the initial weights. Given the set of labeled training data $\{(x_i, y_i)\}_{i=1}^{M}$, we first determine the projection directions α that maximize the following criterion:

$$J(\alpha) = \frac{\alpha^T S_B \alpha}{\alpha^T S_W \alpha} \quad (2)$$

S_B and S_B are the between-class scatter and within class scatter matrices computed using the kernel trick. Fisher directions α are then used to project x_i to obtain z_i (i.e., $z_i = \alpha^T x_i$). Now analyse these z_is to learn prior information on the training sample weights. Let $\mu_{y_i}, y_i \in \{-1,1\}$ denote the class mean of the projected samples. Then, for each sample z_i, computes a parameter

$$\epsilon_i = \frac{|z_i - \mu_{y_i}|}{\sum_{\forall k: y_k = y_i} |z_k - \mu_{y_k}|} \quad (3)$$

which is a function of the distance between a sample and its class mean in the projectedspace. Then, if $\omega_i = \frac{1}{M}$ denotes the uniform initial weights of all training samples x_i, the new initial weights ($\widetilde{\omega}_i$) are obtained by

$$\widetilde{\omega}_i = \omega_i \exp(-\delta \epsilon_i) \quad (4)$$

where δ is the factor controlling the importance of weights learned from (3). δ is the total classification accuracy4 of kernel discriminant analysis on the training samples, which estimates the reliability of the learned weights. This new set of weights is then normalized to make it a distribution, and the classification function is learned using boosting, as described in Algorithm I.

New cost function for ϵ_t^*: The error ϵ_t^* that is minimized to select the weak learners $\hat{h}_t$ is, in its basic form, a function of the classification rate. However, as mentioned before,the problem of outliers leads to a situation where the weights of certain samples becomesignificantly higher than others. Since the undesirable condition caused by outliers isan uneven distribution of the sample weights, is solved by minimizing the following costfunction at every t^{th} iteration:

$$g_t = \frac{(M - \sum_{i=1}^{M} D_i(i) y_i \hat{h}_t^{x_i})}{M} + \lambda_R f_P(D_{t+1}) \quad (5)$$

where the first term measures the error in classification, and the second term $f_P(.)$ measureshow sparse the distribution D_{t+1} produced by the weak learner $\hat{h}_t$ will be. λ_R is a regularization parameter. From the study of the problem of outliers it is found that $f_P(.)$ should not be sparse. In other words, the weights should not be concentrated on only a few training samples. Hence, it is defined as:

$$f_P(D_{t+1}) = \frac{\sum_{i=1}^{M} I(D_{t+1}(i) < \lambda_{cost})}{M} \quad (6)$$

where $I(.)$ is an indicator is function, and λ_{cost} is a threshold. The values of λ_R and λ_{cost} are learned using cross-validation. With thesetwo modifications an outlier-robust boosting algorithm is presented in Algorithm 1. Atevery iteration t, the detection statistic g_t (5) is evaluated on all the weak learners ℍto select the one with the minimum cost (10). The sets of weak learners selected up toTiterations are then linearly combined to generate the final decision value g^* (13) thatclassifies a pixel x in the test image as a lane marking or otherwise.

Algorithm I: Proposed boosting algorithm that reduces over fitting and the effect of outliers in the training set

---

Given: $\{(x_i, y_i)\}_{i=1}^{M}$, where $x_i \in \{-1, +1\}$ its class label, and a pool of weak learners $H = \{(h_j,)\}_{j=1}^{M_2}$.

Initialize the weight distribution of training samples D_i from the weights learned from (4)

$$D_i(i) = \frac{1}{M} \exp(-\delta \epsilon_i) \; \forall i = 1, \ldots, M \quad (7)$$

For iterations $t = 1, \ldots, T$:

(i). $\forall h_j \in H$, compute the classification error,

$$E_{h_j} = \frac{M - \sum_{i=1}^{M} D_i(i) y_i h_j^{x_i}}{M} \quad (8)$$

(ii). Compute an intermediate weight distribution ${D_{t+1}}^{h_j}$, which the weak classifiers $h_j \in H$ will produce,

$$D_{t+1}^{h_j}(i) = \frac{D_i exp(-\alpha_t y_i h_j^{x_i})}{Z_t} \quad (9)$$

Where $\alpha_t \in R$, and Z_t is a normalization term to make $D_{t+1}^{h_j}$ a distribution.

(iii). Select the weak learner $\hat{h}_t$ that has the minimum cost g_t, and compute

$$\vec{\epsilon_t} = \min_{h_j \in H} E_{h_j} + \lambda_R f_P \left(D_{t+1}^{h_j}\right) \quad (10)$$

$$\hat{h}_t = arg \min_{h_j \in H} E_{h_j} + \lambda_R f_P \left(D_{t+1}^{h_j}\right) \quad (11)$$

(iv). Compute the new weight distribution,

$$D_{t+1}(i) = \frac{D_i \exp\left(-\alpha_t y_i h_j^{x_i}\right)}{Z_t}$$

Output the final classifier as

$$g*(x) = sign\left(\sum_{t=1}^{T_i} \alpha_t \hat{h}_t^x\right)$$

Which is a binary value corresponding to whether the test pixel x belong to lane markings or non-lane markings.

C. Detection Phase

Here lane markings are localized in a test image by computing the pixel-hierarchy descriptor f for all pixels and then classifying those using (3). $g^*(x) = 1$ if the test pixel $x \in O$ (lane markings), and $g^*(x) = -1$ otherwise. The subsets of pixels in a test image classified as lane markings are then grouped and parameterized by a second-order polynomial using the generalized Hough transform [16] as follows:

$$L_i = p_2\bar{x}^2 + p_1\bar{x} + p_0 \quad (14)$$

where L_i denotes the i^{th} lane marking, and $\bar{x}$ its horizontal coordinates in the image plane. It is denoting the final detection result.

4. Vehicle Detection

While dense traffic has been reported as challenging for various lane detection and vehicle detection systems, few studies have explored integration of lane and vehicle detection. The study showed that coupling the two could improve vehicle detection rates for vehicles in the ego-lane. Integrating lane and vehicle detection can provide robustness in dense traffic scenarios, improving detection performance for vehicles and lanes. Most of the reported methods address vehicle detection in two stages, namely hypothesis generation and hypothesis verification. In the former, a quick search is performed so that potential locations of the vehicles in the image are hypothesized. The search is typically based on some expected feature of vehicles, such as color, shadow, vertical edges, or motion. The aim of the second stage is to verify the correctness of the vehicle candidates provided by the hypothesis generation stage.

Traditionally, fixed or deformable models have been used for vehicle verification. And some widespread descriptors include Gabor filters, principal component analysis (PCA), and histograms of oriented gradients (HOG). In particular, Gabor filters have been broadly used for image-based vehicle verification.Traditionally, a Gabor filter bank at different scales and orientations is used for feature extraction. So in this system vehicle detection is done based on Gabor filters which have been reported to show good performance in this task. Although Gabor filters have been extensively applied for a broad range of applications, they involve a number of drawbacks.

- The bandwidth of a Gabor filter is typically limited to one octave, thus a large number of filters are needed to obtain wide spectrum coverage. The amplitude of natural images falls off in average by a factor of roughly$1/f$. This is in contrast to the properties of Gabor filters.
- A big extent of the Gabor response concentrates on the lower frequencies, which in turn results in redundant information of the filters and the high frequency tail of the images is not captured.

Here we propose a new descriptor based on the alternative family of log-Gaborfunctions for vehicle detection, as opposed to existing Gabor filter-based descriptors. Thesefilters are theoretically superior to Gabor filters as they can better represent the frequencyproperties of natural images.

A. Design of the Descriptor

Prior to defining the Gabor and log-Gabor descriptors, let us make some considerationsrelated to the filtering process. Naturally, the input space is discrete, therefore in orderto transfer the image and the filter to the frequency domain, the Discrete Fourier Transform (DFT) is required. The filtered image in the input space is eventually obtained by applying the Inverse DFT (IDFT) to the product of the image and the filter inthe frequency domain. At this point it is interesting to recall that the product of DFTs isequivalent to the circular convolution of the corresponding functions in the spatial domain.Thus, the discontinuity between the intensity in the different borders of the original image affects the filtering and produces artefacts in the output image contour. Since the filter is oriented in the horizontal axis, artefacts arise in the left and right boundaries of the image.

In order to avoid this effect in both Gabor and log-Gabor filtering, here proposes to enlarge the original image by replicating its boundaries (i.e., values outside the bounds of the image are assumed to equal the nearest border value). The artificially imposed continuity at the borders results in a low response to the filter, while the artefacts due to circular convolution are shifted to the new image boundaries. Those are conveniently discarded after obtaining the filtered image through the IDFT. The size of the enlarged image is $(R+E)\times(C+E)$, where E must be greater than the width of the filter envelope, $E>\sigma_x$, $E>\sigma_y$. In addition here Gabor filters have been forced to have a zero DC component by setting $G_{m,n}(0,0)=0$, as done in other approaches to Gabor filtering in order to reduce the sensitivity of the filter to absolute intensity values.

B. Feature Extraction

Several strategies can be taken to define the feature vector from the result of Gabor filtering. It includes the use of raw Gabor responses, thresholded Gabor features, Gabor energy features and grating cell operator features. The former two deliver poor results, whereas good performance is achieved by Gabor energy features and grating cell operator features. Gabor energy features (which combine the response of symmetric and antisymmetric Gabor filter) are thus selected. These features result in very large vectors (as large as the size of the image) and thus entail heavy training and classification. Therefore, statistical moments are usually preferred and will also be adopted in this system. In particular, three moments are analysed: the mean(μ), the standard deviation(σ), and the skewness(γ), of the data distribution:

$$\mu_{m,n}=\frac{1}{R.C}\sum_x\sum_y|J_{m,n}(x,y)|$$

$$\sigma_{m,n}=\sqrt{\frac{1}{R.C}\sum_x\sum_y\left(|J_{m,n}(x,y)|-\mu_{m,n}\right)^2}$$

$$\gamma_{m,n}=\frac{1}{R.C}\sum_x\sum_y\left(\frac{|J_{m,n}(x,y)|-\mu_{m,n}}{\sigma_{m,n}}\right)^3$$

where $J(m,n)(x,y)$ represents the input image $I(x,y)$ filtered by one of the filters in theGabor filter bank, $G(m,n)$, or the log-Gabor bank, $LG_{m,n}$, and $|.|$denotes the modulus of a complex number. The mean and the variance are deemed essential and are therefore always included in the feature vector. The skewness, in turn, may be included if the gain in performance justifies the computational overhead. To assess the amount of information conveyed by this parameter, let us analyse its distribution along the Gabor-filtered images. Specifically, a reference Gabor filter bank with parameters $N=4, K=6$, $F_{0=0.4}$ and $a=2$ is selected, and the skewness is computed for each filtered image. The feature vector is thus composed of the mean and variance of the images resulting of applying the Gabor or log- Gabor filter bank to the input image:

$$v\simeq\left[\mu_{0,0},\ \sigma_{0,0},\ \mu_{1,0},\sigma_{1,0},\ldots,\mu_{N-1,K-1},\sigma_{N-1,K-1}\right]$$

5. Experiments

First we evaluated the proposed system using the lane marking and non-lane marking images in both daytime and night time. The system was trained using two datasets, first using a lane marking dataset and secondis a non-lane dataset. The lane marking dataset include 859 road scene images with clear lane marking. And non-lane marking dataset include 85 road scene images without lane marking. Here all the images in the datasets are processed by extracting the Spatial Context features for selected 24 regions. Now these features are given to Adaboost for training a model.

The testing set was of 50 images. During testing counted the fraction of lane marking pixels in the regions corresponding to boosting results (which intersects with those hand marked by the user) to determine the correct detection rate and counted the points that are not marked by the user that have been classified as the lane marking class by the algorithm to compute the false positive rate. And in that evaluation the system has shown an 85 percentage of accuracy.

6. Conclusion

This system puts forward a learning approach for detection of road lane using visual inputs from a camera mounted in front of a vehicle. Apart from conventional feature models a pixel-hierarchy descriptor in which different visual features such as intensity patterns, texture, and edges are analysed in a hierarchy of regions surrounding each pixel corresponding to the object. This system thereby introduces modelling spatial context information through an outlier-robust boosting formulation. At the core of this approach is the importance placed on the quality of data. In order to make the system more intelligent and automated vehicle detection is included. Most importantly, in contrast to typical approaches using Gabor filter banks, a new descriptor based on log-Gabor functions has been proposed. In particular, log-Gabor filter banks are proven to yield better results than Gabor filtersbanks using the same number of filters due to their more effective coverage of the spectrum, and to scale better as the number of filters decreases. To obtain robust performance under varied road conditions, one could use complementary information from different sensing modalities such as the vehicles inertial sensors, GPS information, and models for road geometry.

References

[1] J. McCall and M. Trivedi, “Video-based lane estimation and tracking fordriver assistance: Survey, system, and evaluation,” *IEEE Trans. Intell.Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.
[2] Y. Otsuka, S. Muramatsu, H. Takenaga, Y. Kobayashi, and T. Monj, “Multitype lane markers recognition using local edge direction,” in *Proc.IEEE Intell. Veh.Symp.*, Jun. 2002, pp. 604–609.

[3] C. Rasmussen, "Combining laser range, color, and texture cues for autonomous road following," in *Proc. IEEE Int. Conf. Robot.Autom.*,Aug. 2002, pp. 4320–4325.

[4] R. Tapia-Espinoza and M. Torres-Torriti, "A comparison of gradient versus color and texture analysis for lane detection and tracking," in *Proc. Latin Amer. Robot.Symp.*, Oct. 2009, pp. 1–6.

[5] C. Kreucher and S. Lakshmanan, "LANA: A lane extraction algorithm that uses frequency domain features," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 343–350, Apr. 1999.

[6] Q. Li, N. Zheng, and H. Cheng, "Springrobot: A prototype autonomous vehicle and its algorithms for lane detection," *IEEE Trans. Intell. Transp.Syst.*, vol. 5, no. 4, pp. 300–308, Dec. 2004.

[7] D. Kang and M. Jung, "Road lane segmentation using dynamic programming for active safety vehicles," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3177–3185, Dec. 2003.

[8] Y. Wang, E. Teoh, and D. Shen, "Lane detection and tracking using B-snake," *Image Vis. Comput.*, vol. 22, no. 4, pp. 269–280, Apr. 2004.

[9] S. Nedevschi, R. Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, and C. Pocol, "3D lane detection system based on stereovision," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Oct. 2004, pp. 161–166.

[10] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo visionsystem for generic obstacle and lane detection," *IEEE Trans. ImageProcess.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.

[11] Y. Wang, E. Teoh, and D. Shen, "Lane detection and tracking using B-snake," *Image Vis. Comput.*, vol. 22, no. 4, pp. 269–280, Apr. 2004.

[12] F. Heimes and H.-H. Nagel, Towards active machine vision-based driver assistance for urban areas, Int. J. Comput. Vis., vol. 50, no. 1, pp. 534, Oct. 2002.

[13] Z. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 16–26, Mar. 2008.

[14]Y. Freund and R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119139, Aug. 1997.

[15]Gopalan, R. , Tsai Hong, Shneier, M. and Chellappa, R, "A learning approach towards detection and tracking of lane markings," *IEEE Trans. Intell.Transp. Syst.*, vol. 13, no. 3, Sep. 2012.

[16] S. Divvala, D. Hoiem, J. Hays, A. Efros, and M. Hebert, An empirical study of contextin object detection, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 12711278.