Vehicle Detection Based on LiDAR and Camera Fusion

Feihu Zhang, Daniel Clarke, Alois Knoll

Abstract— Vehicle detection is important for advanced driver assistance systems (ADAS). Both LiDAR and cameras are often used. LiDAR provides excellent range information but with limits to object identification; on the other hand, the camera allows for better recognition but with limits to the high resolution range information. This paper presents a sensor fusion based vehicle detection approach by fusing information from both LiDAR and cameras. The proposed approach is based on two components: a hypothesis generation phase to generate positions that potential represent vehicles and a hypothesis verification phase to classify the corresponding objects. Hypothesis generation is achieved using the stereo camera while verification is achieved using the LiDAR. The main contribution is that the complementary advantages of two sensors are utilized, with the goal of vehicle detection. The proposed approach leads to an enhanced detection performance; in addition, maintains tolerable false alarm rates compared to vision based classifiers.

Experimental results suggest a performance which is broadly comparable to the current state of the art, albeit with reduced false alarm rate.

I. INTRODUCTION

Within the past few years, vehicle detection has become an important task for advanced driver assistance systems (ADAS). Various algorithms for vehicle detection have been developed principally based on computer vision techniques. Current vision based vehicle detection techniques extract different features to segment Regions Of Interest (ROI) and detect objects within those regions [1]. Khammari et al. present a vision based vehicle detection system using a gradient and Adaboost classification technique [2]. Miller et al. [3], Paragios et al. [4] have achieved gabor filtering and adaptive contour algorithm. Vehicle profile symmetry and shadows underneath a vehicle are also used to detect vehicles in [5]. However, vision based methods often suffer from variations in light intensity and limited fields of view. In addition, it is also difficult to extract accurate range information, which is often critical for vehicle detection.

LiDAR is also widely used for ADAS related applications. It provides high precision range information with wide fields of view [6]. Dominguez et al. present a LiDAR based perception solution for autonomous vehicle [7]. However, the major challenge for LiDAR based system is that LiDAR points suffer from association issues, which is a challenge to distinguish objects (especially for points cloud data).

The integration of LiDAR and cameras to improve the performance of object detection, has attracted the attention of the research communities [8], [9]. Premebida et al. fuse the information from LiDAR and camera for vehicle detection and tracking [10]. Fawzi Nashashibi et al. utilize a transferable belief model to combine the data from both LiDAR and cameras with the goal of classification [11]. Spinello et al. present a HOG-SVM classifier based on the object's position detected by the LiDAR [12]. In their work, a Bayesian decomposed expression is used as the reasoning fuse rule. In the above work, range information is first provided by LiDAR while the vision based technique is used to detect the objects. This kind of sensor fusion system is called the classic LiDAR-camera fusion system.

In this paper, the focus is on vehicle detection by fusing data from LiDAR and camera sensors. There are two phases in our approach: a hypothesis generation phase and a hypothesis verification phase. In hypothesis generation phase, the image is cropped to extract the ROI where potentially represent vehicles. Regions are then transformed from image coordinates to LiDAR coordinates to extract the corresponding measurements. In hypothesis verification phase, those measurements are used to estimate the object's contour parameters by utilizing the Random Hypersurface Models (RHM) [13]. Finally, the shape parameters are used for the Support Vector Machine (SVM) to classify vehicle and non-vehicle objects [14].

The contributions of the proposed approach can be concluded as follows: First, a sensor fusion methodology is proposed. The stereo camera provides features for object detection in the hypothesis generation phase, whereas the LiDAR provides range information for object classification in hypothesis verification phase. Second, using shape parameters for object classification is first proposed. In this paper, RHM model is applied to extract object's shape parameter for the classification. Third, false detection rate is greatly reduced. The experiment indicates that the proposed approach achieves high reliability for vehicle detection with a lower false alarm rate compared to vision-based techniques.

The proposed approach is evaluated under real traffic scenarios provided by an off-the-shelf platform [15]. The related suites of sensors include a Velodyne 64-beam laser and a stereo camera. The calibration between LiDAR and the camera has already been achieved [16].

The remainder of this paper is structured as follows: Sec. II briefly describes the hypothesis generation phase. Sec. III introduces more details about the hypothesis verification phase. Sec. IV presents experimental results under traffic scenarios. Finally, the paper is concluded in Sec.V.

Feihu Zhang and Alois Knoll are with the Technische Universität München, Garching bei München, Germany, e-mail: feihu.zhang@tum.de, knoll@in.tum.de.

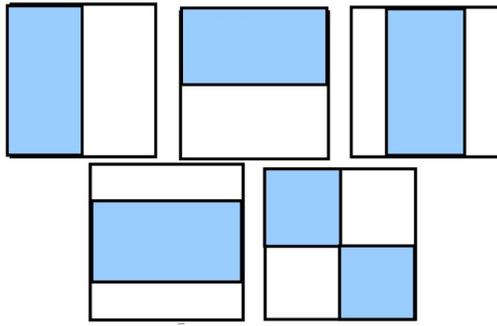Daniel Clarke is with the fortiss GmbH, München, e-mail: clarke@fortiss.org.

Fig. 1. Haar features set

II. HYPOTHESIS GENERATION PHASE

The generation phase consists of two steps: the detection step and the extraction step.

In detection step, regions where potentially represent vehicles are detected; in extraction step, the related LiDAR measurements on objects' surfaces are extracted.

The goal of the hypothesis generation phase is to extract potential objects' range measurements.

A. Vehicle Detection

Vision based vehicle detection is a well established field of research. Current techniques utilize various features to detect vehicles [17], e. g. HoG features [5], [18], Haar-like features [19], [20], edge features [21] and optical flow [22], [23], [24].

In this paper, we use Haar-like feature detector and AdaBoost to detect vehicles. Furthermore, input images from the stereo cameras are utilized to calculate a dense depth map based on the semi-global matching algorithm [25]. Range measurements from the potential objects are acquired since the calibration process has already been done.

- Haar Features

Fig. 1 exhibits the basic templates in Haar features. In this paper, we only utilize templates which contain two and three rectangles to compute the Haar features since most of the vehicles can be assumed as a rectangular shape. The diagonal template (four rectangle template) doesn't represent the effective information for this shape.

- Combining Haar Features with AdaBoost

The AdaBoost algorithms was first proposed by Viola et al. for face detection in real time environment [26]. They introduced the Haar features combined with AdaBoost algorithms as classifiers for pedestrian detection. A single Haar feature is considered as a classifier since it can classify the whole group of images into positive ones and negative ones. By combining several classifier the proposed AdaBoost algorithm achieves high precision detection rate. In this paper, Haar features are utilized to detect vehicles in urban environment.

B. Measurement Extraction

Once a potential vehicle has been detected, the associated LiDAR measurements on surface are required for the shape
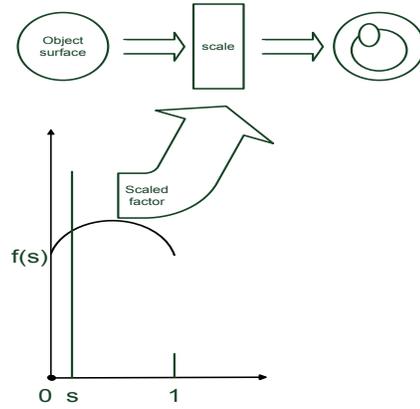

Fig. 2. Random Hypersurface Model for an ellipse

estimation in verification phase. The only issue is to calculate the locations of the potential vehicles with the goal of measurement extraction.

In this paper, we use the semi-global matching algorithm to calculate the depth map of the images [25]. Based on this depth map, the vehicles positions are acquired form the ROI. The corresponding range measurements from LiDAR sensor are also acquired according to the coordinates transformation.

Finally, those measurements are used to estimate the shape information (here we only use the LiDAR points on object's surface to estimate the shape parameters). More details can be found in Sec. III.

III. HYPOTHESIS VERIFICATION PHASE

The verification phase has two steps: the shape estimation step and the classification step. The estimation step calculates contour parameters by utilizing the information from LiDAR, whereas the classification step verifies the objects by utilizing the support vector machine.

A. Shape Estimation

In hypothesis generation phase, LiDAR points from individual objects are acquired. In this section, the corresponding points from object's rear side is extracted to estimate the shape parameter in a $2D$ coordinates.

- Random Hypersurface Model

In this section, a Bayesian solution called Random Hypersurface Model (RHM) is applied with the goal of parameter estimation [13].

As illustrated in Fig. 2, the RHM assumes a single measurement source is a randomly scaled element of the shape boundary. The horizontal axis represents the scaled factor while the vertical axis represents the boundary value calculated by the scaled factor with a random rule. Furthermore, the scaled factor is randomly drawn within the range $[0, 1]$ and independent of the object shape. The definition of the Random Hypersurface Model is given as follows:

As a star-convex shape, $S(p_k)$ is defined as the object surface which consists of the center m_k and the parameter
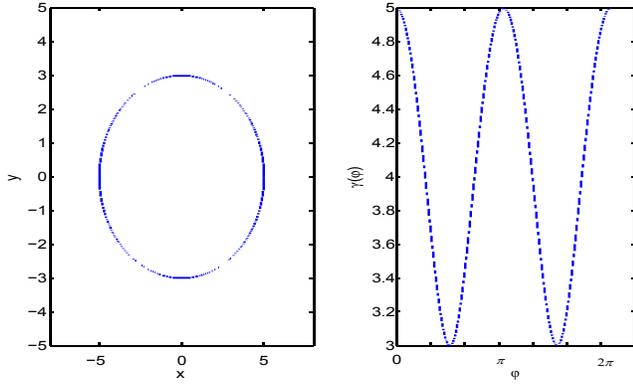
Fig. 3. Example for the shape of star-convex object

vector p_k. The measurement source is assumed to be an scaled boundary element if s is a random draw from the range $[0, 1]$:

$$m_k + s \cdot (S(p_k) - m_k) \qquad (1)$$

where each measurement source lies on the object boundary once the scaled factor equal to 1.

As illustrated in Fig. 3, $r(\phi)$ is represented as a radial function which calculates the distance from the boundary to the center on angel ϕ considered as the polar representation of the contour [27]. The representation is also adaptable for RHM since it is restricted to star-convex shapes.

Assuming $r(\phi)$ is characterized by a parameter vector $\bar{b}_k$ with the center m_k, the shape parameter is considered as $p_k = [\bar{b}_k, m_k]^T$ while the object surface is given by

$$S(p_k) = \{s \cdot r(\bar{b}_k, \phi) \cdot e(\phi) + m_k | \phi \in [0, 2\pi], s \in [0, 1]\} \quad (2)$$

where $e(\phi) := \begin{bmatrix} \cos \phi \\ \sin \phi \end{bmatrix}$ is the unit vector with angle ϕ.
The radial function $r(\bar{b}_k, \phi)$ is considered as the expansion of Fourier series in ϕ. Since $r(\bar{b}_k, \phi)$ is a periodic function with period $[0, 2\pi]$, the Fourier series expansion of degree N_F becomes

$$r(\bar{b}_k, \phi) = a_k^0 + \sum_{j=1}^{N_F} a_k^j \cos(j\phi) + b_k^j \sin(j\phi) \qquad (3)$$

where the shape parameter vector $\bar{b}_k$ is given by

$$\bar{b}_k = [a_k^0, a_k^1, b_k^1, \dots, a_k^{N_F}, b_k^{N_F}]^T \qquad (4)$$

Furthermore, (3) becomes linear in $\bar{b}_k$ if angle ϕ fixed, i. e. ,

$$r(\bar{b}_k, \phi) = R(\phi) \cdot \bar{b}_k \qquad (5)$$

where

$$R(\phi) = [1, \cos(\phi), \sin(\phi), ..., \cos(N_F\phi), \sin(N_F\phi)] \quad (6)$$

Therefore, for a certain type of star-convex shape, the Fourier coefficients $\bar{b}_k$ is considered as the shape parameters of the surfaces.

Lower indices Fourier coefficients encode rough shape information while Fourier coefficients with higher indices give more detailed information. In this paper, 13 Fourier coefficients ($N_F = 6$) are considered to represent most details of the shape information.

• Bayesian State Estimator

A Bayesian state estimator is utilized to calculate the Fourier coefficients with the measurements on object's surface.

The noisy measurement y_k is originated from the single measurement source z_k according to

$$y_k = z_k + v_k \qquad (7)$$

where v_k is considered as zero-mean Gaussian noise.

The state x_k is considered as the Fourier descriptors. Since the shape parameter doesn't drift against time, the dynamic model can be considered as

$$x_k = A_k x_{k-1} + w_k \qquad (8)$$

where A_k denotes the identity matrix, w_k is zero-mean Gaussian noise.

With (2) the sensor model (7) becomes

$$\begin{aligned} y_k =& z_k + v_k \qquad (9)\\ =& s \cdot r(\bar{b}_k, \phi) \cdot e(\phi) + m_k + v_k \\ :=& h(x_k, v_k) \end{aligned}$$

which maps the state x_k to the measurement y_k. According to (5), $r(\bar{b}_k, \phi)$ becomes linear for given angle ϕ while the sensor model is therefore turned out to be

$$y_k = s \cdot R(\phi) \cdot \bar{b}_k \cdot e(\phi) + m_k + v_k \qquad (10)$$

Algebraic manipulations on (10) can be used for estimating Fourier coefficients as follows

$$||y_k - m_k||^2 = s^2 \cdot ||R(\phi) \cdot \bar{b}_k||^2 + 2sR(\phi)\bar{b}_k e(\phi)^T v_k + ||v_k||^2$$

The following new sensor model is obtained based on above equation:

$$\begin{aligned} 0 = h(x_k, v_k) =& s^2 \cdot ||R(\phi) \cdot \bar{b}_k||^2 \\ &+ 2sR(\phi)\bar{b}_k e(\phi)^T v_k + ||v_k||^2 - ||y_k - m_k||^2 \end{aligned}$$

where the state, the measurement noise, the scaled factor and the measurement are mapped to a pseudo measurement 0. An Unscented Kalman filter (UKF) is utilized to update the state of the Fourier coefficients. More details of the UKF filtering can be found in [13].

Fig. 4 illustrates the shape estimation result based on the RHM with a Bayesian estimator. As we can see, the Fourier coefficients represent most details of the shape information based on the measurements. The SVM is utilized to consider the corresponding parameters for vehicle and non-vehicle classification.
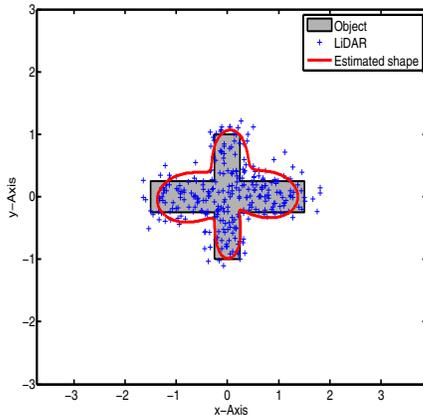
Fig. 4.   Example for shape estimation

B. Vehicle Classification

Vehicle classification is achieved using support vector machine, which is based on the statistical theory of learning, developed by Vapnik [14]. It provides a set of principles to be followed in order to obtain classifiers with good generalization, defined as its ability to predict correctly the class of new data where the learning occurred.

SVM has been extensively used as a classification tool and has found many successes in a wide range of applications. In this paper, shape parameter (13 Fourier descriptors) is utilized to classify vehicle and non-vehicle objects. Compared to spatial and temporal domain features, the Fourier coefficients extract the whole shape information for the purpose of classification. More details about the SVM can be found in [14].

IV. EXPERIMENT EVALUATION

We will first explain the experimental setup and then present our evaluation results.

In order to apply the shape parameter for hypothesis verification, it is necessary to train the system first. We take advantage of the autonomous driving platform Annieway [15] for the evaluation. The platform consists of 5000 training images with 1893 objects are manually labeled to train the classifier and 2000 test images for the evaluation. In RHM estimation process, the scaled factor s is assumed to be Gaussian distributed with mean 0.5 and variance 0.02, while the initialize of the state and covariance is based on the empirical parameters $[1, 0, ..., 0]^T$ and $\mathrm{diag}[0.5, ..., 0.5]$, $\mathbf{Q}, \mathbf{R}$ is represented as $\mathrm{diag}[0.1, ..., 0.1]$ and $\mathrm{diag}[0.3, ..., 0.3]$. In SVM training process, we directly extracted the estimated shape parameters with the labeled objects (here we only considered vehicle and non-vehicle objects) into the SVM to calculate the support vector. The success rate for vehicle and non-vehicle classification is 91%, calculated by SVM with the training data. It illustrates that the shape parameter is a reliable vector for the SVM to classify objects.

The evaluation is processed by the test dataset. The results are presented in Fig. 5. The first row is the original image

	The proposed approach	Baseline
Detection rate	91.5%	91.5%
False rate	2.7%	10.8%

from the camera and the second row is the detected result in hypothesis generation phase. The third row is the fusion result in hypothesis verification phase. As we can see, the potential vehicles are detected in hypothesis generation phase. However, after hypothesis verification phase, the false objects are eliminated. This highlights how the proposed approach eliminates the false hypothesis during the whole process.

Table I expresses the overall performance of the proposed approach. From table I we can see that both the detection rates are same in hypothesis generation phase. However, in hypothesis verification phase, the proposed approach greatly eliminates the false detections by utilizing the shape parameters. The detection quality is improved after the false hypothesis elimination.

The contributions of the proposed approach can be concluded as follows:

First, a sensor fusion methodology is proposed which provides more complete information in comparison to each single sensor. The camera provides features for object detection in the hypothesis generation phase, whereas LiDAR provides range information for object classification in hypothesis verification phase.

Second, using shape parameters for object classification is first proposed. In this paper, the RHM model is utilized to extract object's shape parameter to classification. The experiment verifies the feasibility of the proposed method.

Third, false detection rate is reduced. The experiment indicates that the proposed approach achieves high reliability for vehicle detection with a lower false alarm rate in urban environments.

V. CONCLUSION

It is a challenge to detect vehicles robustly under various scenarios. Pedestrians, trucks and bicycles cause the results to deviate from the real status. In previous works, LiDAR is often employed to detect objects and generate target hypotheses, whereas a vision based classifier is responsible for object validation or final classification. However, in this paper, the scheme of hypothesis generation phase and hypothesis verification phase is based on the stereo camera and LiDAR, respectively. In comparison to others, the benefit of the proposed approach is to utilize the shape information to assist vision based techniques for vehicle detection and classification. The evaluation results illustrate that the proposed approach achieves a lower false alarm rate in urban environments, which may be helpful in future autonomous navigation systems.

Future work will concentrate on modeling the contour parameters from 2D models to 3D models.

Fig. 5.   Classification result using SVM

REFERENCES

[1] L. Leon and R. Hirata, "Vehicle detection using mixture of deformable parts models: Static and dynamic camera," in *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, 2012, pp. 237–244.

[2] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laurgeau, "Vehicle detection combining gradient analysis and adaboost classification," in *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, 2005, pp. 66–71.

[3] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary gabor filter optimization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 125–137, 2005.

[4] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 3, pp. 266–280, 2000.

[5] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: features and classifiers," *Image Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 2019–2034, 2006.

[6] L. Huang and M. Barth, "Tightly-coupled lidar and computer vision integration for vehicle detection," in *Intelligent Vehicles Symposium, 2009 IEEE*, 2009, pp. 604–609.

[7] R. Dominguez, E. Onieva, J. Alonso, J. Villagra, and C. Gonzalez, "Lidar based perception solution for autonomous vehicles," in *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, 2011, pp. 790–795.

[8] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, "Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner," *Autonomous Robots*, vol. 19, pp. 117 –140, 2005.

[9] S. Rodrguez, V. Frmont, P. Bonnifait, and V. Cherfaoui, "Multi-modal object detection and localization for high integrity driving assistance," *Machine Vision and Applications*, vol. 14, pp. 1 –16, 2011.

[10] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A lidar and vision-based approach for pedestrian and vehicle detection and tracking," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, 2007, pp. 1044–1049.

[11] F. Nashashibi, A. Khammari, and C. Laurgeau, "Vehicle recognition and tracking using a generic multisensor and multialgorithm fusion approach," *International Journal of Vehicle Autonomous Systems*, vol. 6, pp. 134–154, 2008.

[12] L. Spinello and R. Siegwart, "Human detection using multimodal and multidimensional features," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 3264–3269.

[13] M. Baum and U. Hanebeck, "Shape tracking of extended objects and group targets with star-convex rhms," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, 2011, pp. 1–8.

[14] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[16] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "A toolbox for automatic calibration of range and camera sensors using a single shot," in *International Conference on Robotics and Automation (ICRA)*, 2012.

[17] S. Sivaraman and M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior

analysis," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–23, 2013.

[18] Q. Yuan, A. Thangali, V. Ablavsky, and S. Sclaroff, "Learning a family of detectors via multiplicative kernels," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 3, pp. 514–530, 2011.

[19] W.-C. Chang and C.-W. Cho, "Online boosting for vehicle detection," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 3, pp. 892–902, 2010.

[20] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 267–276, 2010.

[21] B.-F. Lin, Y.-M. Chan, L.-C. Fu, P.-Y. Hsiao, L.-A. Chuang, S.-S. Huang, and M.-F. Lo, "Integrating appearance and edge features for sedan vehicle detection in the blind-spot area," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 2, pp. 737–747, 2012.

[22] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 583–595, 2011.

[23] A. Barth and U. Franke, "Estimating the driving state of oncoming vehicles from a moving platform using stereo vision," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 4, pp. 560–571, 2009.

[24] M. Perrollaz, J.-D. Yoder, A. Negre, A. Spalanzani, and C. Laugier, "A visibility-based approach for occupancy grid computation in disparity space," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 3, pp. 1383–1393, 2012.

[25] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, June 2005, pp. 807–814 vol. 2.

[26] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, pp. 137–154, 2004.

[27] D. Zhang and G. Lu, "Study and evaluation of different fourier methods for image retrieval," *Image and Vision Computing*, vol. 23, pp. 33–49, 2005.