International Journal of Research in Advent Technology, Vol.2, No.5, May 2014
E-ISSN: 2321-9637

3D Model and Part Fusion for Vehicle Retrieval

M.Nagarasan[1], T.N.Chitradevi[2], S.Senthilnathan[3]
Department of computer science and engineering[1,2, 3]
Aditya institute of technology, Coimbatore.[1, 3],Sri Ramakrishna Engineering College, Coimbatore[2]
nagarasancs@gmail.com[1], chitradevi.04@gmail.com[2],senthil.sn1985@gmail.com[3]

Abstract-In the fast emerging trend, Content-based and attribute-based vehicle retrieval plays an important role in surveillance system. Traditional vehicle retrieval is extremely challenging because of large variations in viewing angle/position, illumination, occlusion, and background noise. This work presents a general framework for solve this problem that provide 3D models attempting to improve the vehicle retrieval performance and searching vehicles based on enlightening parts such as grille, lamp, wheel, mirror, and front window. By fitting the 3-D vehicle models to a 2-D image to extract those parts using active shape model. Then compare different 3D model fitting approaches and verify that the impact of part rectification. For improving vehicle retrieval performance using LSH (Locality Sensitive Hashing) and inverted index.

Index Terms-3-D model construction, 3-D model fitting, vehicle Retrieval.

1. INTRODUCTION

In video surveillance, the most important objects are people and vehicles. This work is focus on vehicles. There are more and more surveillance video datasets are available. However, it is impossible for human deal with. Therefore, effective vehicle retrieval is increasing significantly.Object matching and recognition [4] remain an important and long-term task with continuing interest from computer vision and various applications in security, surveillance, and robotics. Many types of representations have been exploited to match and recognize objects by a set of low-dimensional parameters, such as shape, texture, structure, and other specific feature patterns. However, when it comes to unconstrained conditions such as highly varying pose and severely changing illumination, the problem becomes extremely challenging.Appearance-based methods are well applied on vehicles, with no difference with other objects

In the last few years, a number of studies have been undertaken to classify vehicles according to their type, make [8], model, logo or color. However, the evaluation of each of these classification methods [7] has been performed on in-house datasets. Since each dataset involves its own camera viewpoints, lighting conditions and resolutions, there has been no way to compare the relative merits of these methods. These are satisfy by the part rectification based vehicle retrieval using 3D models. Vehicle retrieval and recognition are very challenging because of surveillance videos with time information and surveillance images have several problems such as background noise, occlusion, illumination, and shape variations.So it is hard to retrieve the same type of vehicles without constructing correspondence. In people search domain, using 2D models to extract the parts of a people within bounding box generally fails due to shape variations.

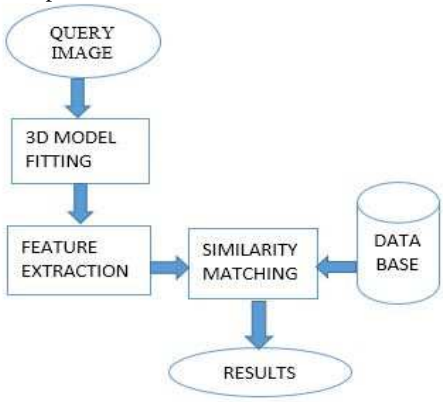

FIG 1 OVERVIEW OF OBJECT RETRIEVAL

Fig.1describes the object retrieval system. Vehicle detection algorithms are challenging for classification the vehicles and recognition the vehicle. So most of the people applying 3D model [6] and extract the features from vehicle. In retrieval, bag of the features and visual word construction methods are used. But this type of features had some noises. It is slow to get processed for retrieve the vehicle. In past days retrieval system used CBIR (content based image retrieval technique) for efficient retrieval of objects. It is suitable for only without noise images then it is hard to retrieval.

In this work, apply part based vehicle retrieval by 3D model fitting and part fusion. First, construct 3D model using Active Shape Model, Principal component analysis, and Generalized Procrustes analysis. Second, fitting 3D models to 2D images using Point set registration [9], and Jacobian system then extract the parts from 3D fitting methods using weighted Jacobian system. Finally, vehicle retrieval achieved by Locality Sensitive Hashing and inverted index methods. Objective and scope of the work is to improve the reliability of 2-D model, improve speed of retrieval with large data set and efficiently analyze the contents of the image and retrieve similar images from the database when metadata is insufficient.

2. LITERATURE SURVEY

Most of the system and approach have large variations in angle, shape, and illumination on vehicle representation and classification.so people use 3D models to improve their work. M.Zeeshan Zia [1] represent object class model by 3D geometry method, and crowded scenarios the occlusion is common. It does not make any assumptions about the nature of the occluder. It presents two layer model, consisting of a robust, but coarse 2D object detector, followed by a detailed 3D model of pose and shape.

3D representations for object recognition and modeling [2] also challenging.3D representation using CAD models for model fitting approaches but 3D model fitting has not accuracy. In addition, unlike people or face recognition, using 2-D models to extract parts of a vehicle within the bounding box generally fails due to dramatic variations in viewing angles. Employing 3-D models is more suitable for this work. In fact, using 3-D vehicle model is one of the major line of research in the fields of vehicle detection, pose estimation, classification, etc.

In classification [3], presented deformable 3D vehicle model that more accurately describes the shape and appearance of vehicles.The model is aligned to images by predicting and matching image intensity edges. Novel algorithms are presented for fitting models to multiple still images and simultaneous tracking while estimating shape in video.CAD model is useless and a waste of computational resources while construct on 3D model.so simply choose the deformable 3D vehicle model for this work. The classification using voting algorithm for a multiclassVehicle type recognition system based on Oriented-Contour Points. This the method is robust to a partial occlusions of the patterns

In object detection [4], 3D shape model and voting procedure is suitable for highly accurate post estimation. But illumination and background clutter is possible. An unsupervised learning approach for view-invariant vehicle detection in traffic surveillance videos, which learns a large number of view-specific detectors during the training phase and given an unseen viewpoint exploits scene geometry and vehicle motion patterns to select a particular view-specific detector for object detection. The key advantage of this approach is that it enables utilization of fast and simple view-specific object detectors for accurate view-invariant object detection. Vehicle make and model recognition or content based vehicle retrieval is a relatively new research problem. The basic idea is to extract suitable features from the images of a vehicle, which can be used to not only retrieve vehicle images having similar appearances but also retrieve its make and model.

Most of the previous researches on sketch-based image retrieval mainly focus on the study of extracting effective features to better match a rough sketch to natural images. M. Eitz et al. [6] survey many different sketch features and compare their performances. However, the memory storage issue is not well discussed. Y. Cao et al. [10] propose the Mind Finder system to index large-scale image dataset. They use every edge pixel in the dataset images as an index point and consider an edge pixel locating on the boundary as a hit. In order to endure slight translation, they divide the image into six directions and find the hit points within a predefined tolerance. Their method preserves the spatial information and uses an inverted index in large-scale image dataset for efficient retrieval. Y. J. Lee et al. [7] propose the Shadow Draw system that guides user to draw better sketches. They extract BICE binary descriptors from images and use min-hash function to randomly permute the descriptors.Each descriptor is translated to n min-hash values of size k and indexed by an inverted look-up table. However, those prior works are usually based on the inverted index structure, which grows rapidly in data size. Therefore, they usually put the inverted index on the server side.Motion-based video retrieval is an active researcharea, where object's motion trajectory is used as an importantfeature for video retrieval [15].

The objective of content-based image retrieval is to efficiently analyze the contents of the image and retrieve similar images from the database when metadata such as keywords and tags are insufficient. To bridge the semantic gaps, how to efficiently use available features such as color, texture, interest points of images and spatial information is the key. Searching for people in surveillance videos, Feris et al. [5] build a surveillance system capable of vehicle retrieval based on semantic attributes such as facial hair, eyewear, clothing color, etc. This provide

efficient search on people. But dataset is huge, due to the presence of multiple frames from each person. So practical issues arise.

Designing attributes usually involves manually picking a set of words that are descriptive for the images under consideration, either heuristically [9] or through knowledge bases provided by domain specialists [11]. After deciding the set of attributes, additional human efforts are needed to label the attributes, in order to train attribute classifiers. The required human supervision hinders scaling up the process to develop a large number of attributes. More importantly, a manually defined set of attributes (andthe corresponding attribute classifiers) may be intuitive but not discriminative for the visual recognition task.

Searching for vehicles in surveillance videos based on semantic attributes. At the interface, the user specifies a set of vehicle characteristics such as color, direction of travel, speed, length, height, etc. and the system automatically retrieves video events that match the provided description [14]. Multiview detection approach relies on a novel and robust vehicle detection method, followed by attribute extraction, transformation of measurements into world coordinates, and database ingestion/search. Different from this system based on attributes and applied on surveillance videos, this work focuses on extracting informative parts from fitted vehicle models and part fusion for improving part-based vehicle image retrieval.

3. RELATED WORKS

Attribute-based retrieval framework has gained much attention recently. Previous work mainly focus on people [5] search, salient attributes or parts have beenutilized to identify targets. However, most of approachesusually use 2D models to extract parts within the boundingbox and generally fail under large variations of viewing angles. For vehicles, most approaches or systems either detectvehicles from background or classify vehicle types such ascars, buses, trucks. Some researches further use 3D models to improve the performance.

Given the bounding box [12] of an observedvehicle in image, it can be denote as BBh(l; r; t; b), where(l; r; t; b) are the left, right, top and bottom coordinates ofthe bounding box respectively. Similarly, the bounding boxof the 2D projected model can be also obtained and denotedas BBm (l; r; t; b). The aspect ratio difference between BBHand BBM in x and y directions are computed as

$$S_x = \frac{r_h - l_h}{r_m - l_m} \quad (1)$$

$$S_y = \frac{t_h - b_h}{t_m - b_m} \quad (2)$$

Utilizing the backward projection technique again, find the 3D vector vx on the ground plane whose 2D projection aligns with the horizontal axis of the image, and similarly, the 3D vector vy for the vertical axis of the image. Since these vectors are on the ground plane, their Y components are dropped [13]. Assuming all the vectors are unit vectors, the scales in the length SL and width SW dimensions of the vehicle model can be estimated as following:

$$\begin{cases} S_L = \alpha_L(S_x\|Vo.Vx\| + |Sy\|Vo.Vy\|), \\ S_W = \alpha_W(S_x\|V_o^T.Vx\| + |Sy\|V_o^T.Vy\|) \end{cases} \quad (3)$$

$\alpha_L$and α_W are the normalization factors.Since the width and height of a vehicle usually are correlated, the scaling factor of model's height is defined as thesame as SW The fitting problem can be formulated as a Jacobian system (JS)

$$\mathbf{I}\Delta_q = \mathbf{f} \quad (4)$$

wheref is the vector of signed errors, Δq is the vector of parameter displacement updated at each iteration, and I isthe Jacobian matrix with current parameters. The solution is derived by a least square method and iteratively optimizing the parameters until convergence.

Most of the model-fitting algorithms find corresponding points by local search of the projected edges depending only on some low-level features, such as edge intensity and edge orientation, which are likely to fail and converge to local maxima in common cases due to cluttered background or complex edges on the surface of vehicles. In this paper interested in whether it is possible to improve the fitting algorithm with some prior knowledge of parts (e.g., grille, lamp, and wheel). This can give different weights to different correspondences and lead to better fitting results. To validate through assumption, generate synthetic weight maps of parts by using annotated ground truth data, and formulate this problem into a weighted Jacobian system (WJS)

$$\mathbf{VI}\Delta_q = \mathbf{V_f} \quad (5)$$

where V is a diagonal weight matrix with each diagonal element V_{ii} representing the weight of each correspondence. Then take two important weights into consideration, distance weight V_{dist} and part weight V_{part}. $V_{ii}$is computed by a linear combination with λ

$$\mathbf{V_{ii} = \lambda \cdot V_{dist} + (1 - \lambda) \cdot V_{part}.} \quad (6)$$

The distance weight w_{dist} is based on the Beaton–Tukeybiweight. For each projected point, the edges far from the point will not be taken into computation.

The part weight w_{part} is determined by the value of the location of observed edge point in the part weight map. Higher weight values in the part weight map imply where the part is with higher probability. In other words, the projected point belongs to which part in the 3-D model, and the part weight will be higher if the observed edge point belongs to the correct part or near the location of the correct part according to the part weight map. In this experiments show that the 3-D model fitting precision is improved with the aid of the prior weight map and part rectification is calculated by s

$$s = \alpha \cdot x + \beta \cdot y + \gamma \cdot z \quad (7)$$

where $\alpha + \beta + \gamma = 1$. If get the barycentric coordinates of a, b, and c, it is able to calculate α , β and γ. Hence, by bilinear interpolation and inverted mapping, each point in the projected view can find the corresponding point in the original image and get the mapped pixel value. Furthermore, can remove background pixels which are out of projected triangles.By this way, can obtain rectified semantic parts.

4. PROPOSED SYSTEM

This work is focusing on improving vehicle retrieval based on content and attributes. In the offlineProcess, use 3-D vehicle models to build an active shape model (ASM) and apply 3-D model fitting on the input vehicle image. Then crop and rectify informative parts into the same reference view. After feature extraction, conduct part-based vehicle retrieval on NetCarShow300.

The overall framework illustrated in Fig 2. In surveillance work 3D model construction ischallenging. Because shape variations in vehicles. Therefore use CAD model for 3D construction using ASM model and applying different 3D model fitting using different state-of-art methods. Second, rectifying the enlighten parts using image warping and extract features from those rectified parts. Finally, apply LSH and inverted index for retrieval performance.

4.1. 3D Model Construction

Build a deformable 3-D vehicle model and ASM model before 3-D model fitting process for shape variations. To make sure the correspondence of the same physical shape, manually select 128 points for a half vehicle template model. The other half can be obtained by mirroring. Then adjust locations of points from the template model to corresponding locations in each training models. It is inevitable that there exists difference on rotation, translation, and scale factors between 3-D vehicle models. To eliminate the influences of these factors and only analyze shape variation by principal component analysis (PCA), then conduct generalized Procrustes analysis (GPA).

GPA is an approach to align shapes of each instance. The algorithm is outlined as following:

Step 1: Choose a reference shape or compute mean shape from all instances

Step 2: Superimpose all instances to current reference shape

Step 3: Compute mean shape of these superimposed shapes

Step 4: Compute Procrustes distance between the mean shape and the reference. If the distance is above a threshold, set reference to mean shape and continue to step 2.

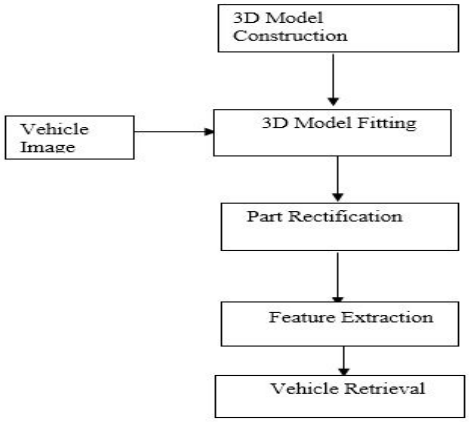

Fig 2 Architecture of the system

4.2. 3D Model Fitting

3-D vehicle model fitting procedure describes in Fig 3. Given the initial pose and shape, then generate the edge hypotheses by projecting the 3-D model into a 2-D imageand remove hidden lines by using depth map rendered from the 3-D mesh.For each projected edge point, the corresponding points are found along thenormal direction of the projected edges. Then, a 3-D model fitting method isperformed to optimized pose and shape parameters. The above procedure isrepeated several times until convergence.

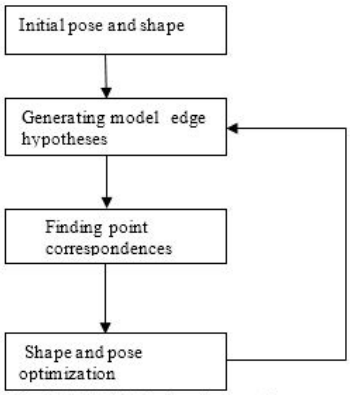

Fig 3 3D Model Fitting Approach

5. EXPERIMENTAL SETUP AND RESULTS

In the following conduct several experiments on a challenging dataset to show the performance for content based Vehicle retrieval approach. Also do the comparison between different 3-D model fitting methods. It is obvious that fitting precision may influence the part extraction. To The experiments are performed on Intel® Core™ i3 2330M @ 2.2GHz, 2200MHz, 2 core(s), 4 logical processors and 2 GB RAM with windows 7 32bit operating system and the experiments are implemented in Matlab R2012b.

To investigate this approach on retrieving vehicle images under different conditions and in various viewpoints so collect 300 images from NetCarShow.com[1] the NetCarShow300[2] dataset, where the size is comparable to commonly used vehicle type recognition datasets, which are composed of only frontal cropped grayscale vehicle images. NetCarShow300 dataset comprises 30 vehicle instances, such as Acura ZDX, Honda Odyssey, Honda Pilot, Opel Corsa, and Volvo V70. Each instance has 10 images, respectively. All images are 800 × 600 color images. Each image contains one main vehicle of which the frontal part is visible. The vehicles are presented in different environments, including noisy background, little occlusion, different illumination, and shadows. In Table 1 shows the results of vehicle retrieval. The part fusion of 70%+PHOG+L1 accuracy is 75.08%. Table 2 refers to compare the differences between model fitting approaches, generate a testing data with noisy initial positionby adding random noises to ground truth. Then measure average pixel distance (APD) and standard deviation (STD) of visible vertices between fitted models and ground truth. Here compare different 3D model fitting methods for improve the accuracy. Table 3 explore the Part fusion accuracy is 75.84%.

Descriptor +Distance Measure	**SIFT+L1**	**SIFT+COS**	**PHOG+L1**	**PHOG+COS**
Original Body Original Side	18.77%	18.21%	10.44%	9.30%
Rectified 50% Front Same Side	29.27%	26.48%	54.84%	45.05%
Rectified Grille Same Side	31.85%	27.17%	45.13%	34.53%
Rectified Lamp Same Side	13.17%	12.24%	47.31%	42.21%
Rectified Wheel Same Side	13.78%	11.87%	14.00%	12.13%
Rectified Mirror Same Side	15.18%	14.44%	19.21%	18.31%
Rectified Front Window Side	14.52%	13.57%	18.65%	17.21%
Fusion of 70% Grille, Lamp, Wheel, Mirror and Front Window	44.26%	40.23%	75.08%	53.79%

Table 1. Vehicle retrieval results

Method	**APD**	**STD**
Initial location	45.15	6.16
PR	21.35	5.11
JS	34.19	6.31
WJS	18.98	4.71

Table 2. 3D model fitting results

Angle(Degrees)	**Part Fusion**	**MAP (%)**
30,-30	Grille, Lamp, and Wheel	68.54
30,60	Grille, Lamp, Wheel, Mirror, and Front Window	75.84

Table 3. Part fusion results

6. PERFORMANCE ANALYSIS

In this chapter shows the performance of vehicle retrieval using different descriptor and distance measure, fitting precision using different fitting methods, and part fusion using different vehicle parts

Fig.4.Performance of vehicle retrieval

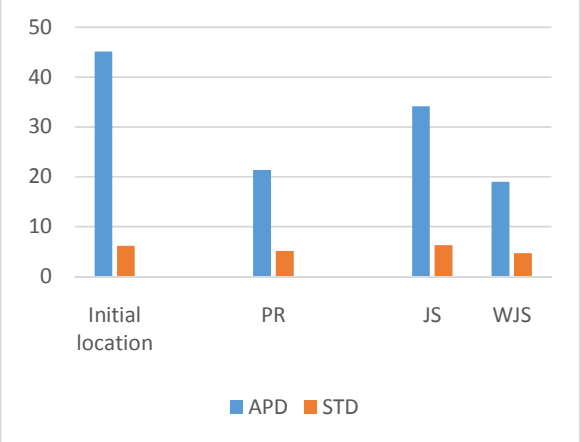

Fig.5.Performance of different model fitting approaches

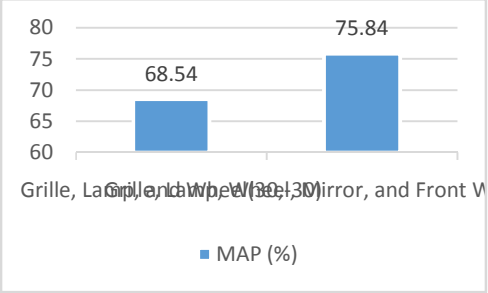

Fig.6.Performance of part fusion

7. CONCLUSION

In this proposed framework useful to retrieve the vehicle image in surveillance environment applications (parking system and theft control system). In this work effectively utilize 3D model fitting approaches to extract the parts. In vehicle retrieval system, LSH and inverted index approaches accurately retrieved vehicle images based on descriptors and feature extraction. The advantages of this system are easily retrieve the vehicle image based on vehicle model and manage the restraints (background clutter and illumination) efficiently. The computational cost of this system depends on 3D model fitting and object retrieval. In future work build a structural framework on more vehicle images without human annotation and extend the parts to effectively improve the performance.

REFERENCES

[1]. Cao, Yang, Changhu Wang, Liqing Zhang, and Lei Zhang. "Edgel index for large-scale sketch-based image search." In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 761-768. IEEE, 2011.

[2]. D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, “Attribute-based people search in surveillance environments,” in Proc. IEEE Workshop Appl. Comput. Vision, Dec. 2009, pp. 1–8.

[3]. Dyana, A., and Sukhendu Das. "MST-CSS (Multi-Spectro-Temporal Curvature Scale Space), a novel spatio-temporal representation for content-based video retrieval." Circuits and Systems for Video Technology, IEEE Transactions on 20, no. 8 (2010): 1080-1094.

[4]. Eitz, Mathias, Kristian Hildebrand, TamyBoubekeur, and Marc Alexa. "An evaluation of descriptors for large-scale image retrieval from sketched feature lines." Computers & Graphics 34, no. 5 (2010): 482-498.

[5]. Farhadi, Ali, Ian Endres, Derek Hoiem, and David Forsyth. "Describing objects by their attributes." In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 1778-1785. IEEE, 2009.

[6]. Feris, Rogerio, BehjatSiddiquie, Yun Zhai, James Petterson, Lisa Brown, and SharathPankanti. "Attribute-based vehicle search in crowded surveillance videos." In Proceedings of the 1st ACM International Conference on Multimedia Retrieval, p. 18. ACM, 2011.

[7]. Kuo, Yin-Hsi, Kuan-Ting Chen, Chien-Hsing Chiang, and Winston H. Hsu. "Query expansion for hash-based image object retrieval." In Proceedings of the 17th ACM international conference on Multimedia, pp. 65-74. ACM, 2009.

[8]. Leotta, Matthew J., and Joseph L. Mundy. "Vehicle surveillance with a generic, adaptive, 3d vehicle model." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33, no. 7 (2011): 1457-1469.

[9]. M. Arie-Nachimson and R. Basri. “Constructing Implicit 3D Shape Models for Pose Estimation” Proc. IEEE Int',l Conf. Computer Vision, 2009.

[10]. M. Stark, M. Goesele, and B. Schiele, “Back to the future: Learning shape models from 3D CAD data,” in Proc. BMVC, 2010, pp. 106.1–106.11.
[11]. Myronenko, Andriy, and Xubo Song. "Point set registration: Coherent point drift." Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, no. 12 (2010): 2262-2275.
[12]. S. M. Khan, H. Cheng, D. Matthies, and H. S. Sawhney, “3D model based vehicle classification in aerial imagery,” in Proc. IEEE Conf.Comput. Vision Pattern Recognit., Jun. 2010, pp. 1681–1687.
[13]. T. Ahon, J. Matas, C. He, and M. Pietikainen, “Rotation invariant image description with local binary pattern histogram fourier features,” in Proc.16th Scandinavian Conf. Image Anal., 2009, pp. 61–70.
[14]. Zafar, Iffat, Eran A. Edirisinghe, and B. SerpilAcar. "Localized contourlet features in vehicle make and model recognition." In IS&T/SPIE Electronic Imaging, pp. 725105-725105. International Society for Optics and Photonics, 2009.
[15]. Zia, M., Michael Stark, BerntSchiele, and Konrad Schindler. "Detailed 3d representations for object recognition and modeling." (2013) In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 3326-3333. IEEE, 2013.