Make and Model Recognition of Cars

Sparta Cheung
Department of Electrical and Computer Engineering
University of California, San Diego
9500 Gilman Dr., La Jolla, CA 92093
sparta@ucsd.edu

Alice Chu
Department of Computer Science
University of California, San Diego
9500 Gilman Dr., La Jolla, CA 92093
aachu@ucsd.edu

Abstract

Current work in car recognition generally relies on a single and specific view in order to identify the car make and model. This paper proposes a technique in performing make and model recognition given an image of an unidentified car viewed from an arbitrary angle. This work partially builds off of Louka Dlagnekov's previous work [1] on recognizing cars.

1 Introduction

Object detection and recognition are necessary in an artificially intelligent and autonomous system. Eventually, these systems are expected to venture to the outdoor environment. Thus, detection of common objects on the streets is necessary to provide input and feedback into the system. Pedestrian [9] and face [10] [11] recognition results have been accurate. Cars, however, proved to be a more difficult object for detection and recognition due to its varying structure from different perspectives of view of the same car, as well as varying between different makes and models.

The rest of this paper begins with a discussion of previous related work. In section 3, the overview of the proposed system is presented. Section 4 discusses methods of accomplishing subtasks of the system. In section 5, results of the system are shown. Section 6 is a discussion of what we learned from the project and future work associated with it. Finally, section 8 provides our references.

2 Related Work

Several different approaches of car detection and make and model recognition (MMR) have been proposed in the past. These approaches have used feature detection [8], 3D modeling [6], and Scale Invariant Feature Transforms (SIFT) [3] While each method of approach to this problem produced considerably accurate detection and classifications of cars, they are constrained to work well only with a specific set of data that is taken in a set condition, i.e. fixed camera position overlooking passing cars directly underneath.

3 System Overview

To determine the make and model of a unknown car, the image is passed to a Matlab script which utilizes various image processing techniques to find the best matching car model within our dataset.

The image of an unknown car model of a specific view (i.e. rearview, side view, ¾ view, etc.) is matched against a dataset of portable gray map (PGM) images taken from the existing database of cars arranged by Dlagnekov [1] along with images taken from websites of various car companies with the same point of view of the car as the query image. The images taken from Dlagnekov's dataset are primarily composed of rear views of cars since his research focused on license plate recognition. The images taken from websites are available in four profile angles: front, rear, side and ¾ views. Car images with a clean background are ideal in order to minimize the number of irrelevant interest points detected.

We designed an MMR system that consists of three primary steps in the image processing algorithm. The first step consists of detecting interesting features, or interest points, on the image of the query car and one of the images in the dataset. The interest points represent interesting features of objects in the image. We assume that the image includes a car object and be of the same size in pixels to eliminate the need for a car detection scheme and to improve feature matching, respectively. Features that are detected will be from both the car and the background scene.

In the second step of the algorithm, interest points of the query image are compared to sets of interest points of each of the images in the database based on appearance. An interest point matcher is used to find the closest matches between the two sets of interest points. The interest point matcher returns two sets of coordinates of paired and matched interest point locations.

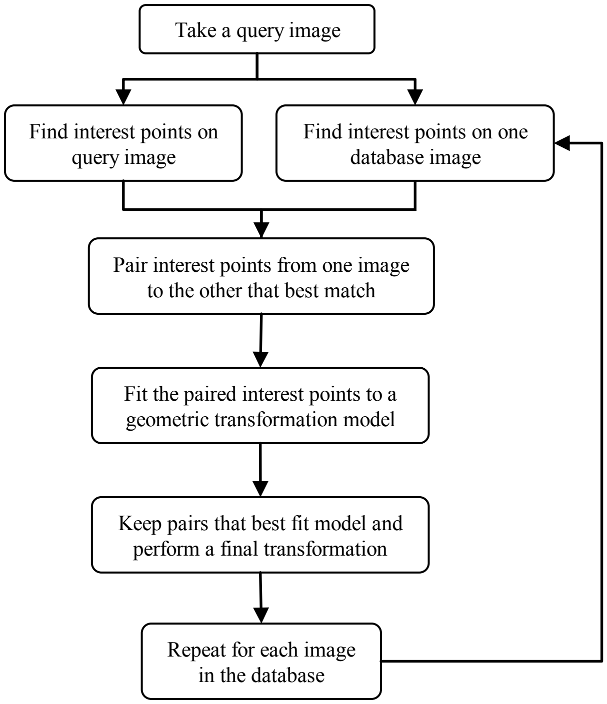


Fig. 1. Algorithm for processing query and database images

In the final step, the paired interest points are used to find a subset of inliers which fit best to a given geometric transformation model. Matched pairs of points between the images that do not fit this model are subsequently not considered for the final transformation to fit the locations of the interest points on the query image to the interest points of an image in the database. The result is a geometrically transformed query image with interest point locations closely corresponding to the locations of the interest points of an image in the database.

This algorithm, shown in Fig. 1, is repeated for N times, where N is the number of images in the database, and the same query image is compared against an image in the database for each iteration. The image in the database with the highest inlier count will be labeled as being the best match to the query image, and that image's make and model will be used to label the query image.

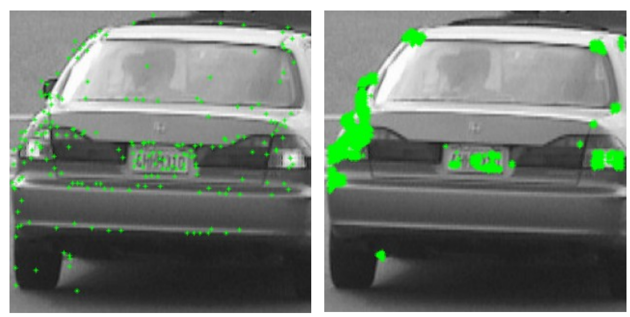
In order to accomplish MMR of query images representing other views of the car, the database can be expanded to include images of cars with the same point of view as that of the query image.

4 Methods

4.1 Interest Point Detection

For step one of the processing algorithm, we looked into two interest point detecting techniques: Scale-Invariant Feature Transform (SIFT) [3] and Harris corner detection [4]. SIFT defines interest points as minima and maxima of the difference of Gaussians that occur at multiple scales. These interest points are used in a nearest-neighbor to find an object model using the Hough transform and lease-squares fit. This technique of feature detection is invariant to image translation, scaling, rotation, and partially illumination changes, affine or 3D projection, allowing a consistent detection of features on cars. Since the goal of the project is to differentiate car makes and models from different views, rotation, size, and illumination differences are expected. In order to overcome this problem, feature detectors must be invariant to these differences, which is accomplished by SIFT.

Harris corner detection defines its interest points as sharp changes in gradient direction, which is calculated by using the sum of square difference. Early forms of this corner detector analyzed edges found by edge detection to find these rapid changes in direction, which became referred to as corners and eliminated the need to use edge detection. Improvements made to this detector used image gradients instead of direct pixel values to add invariance to illumination. Fig. 2 shows the comparison between SIFT features and Harris detected corner.



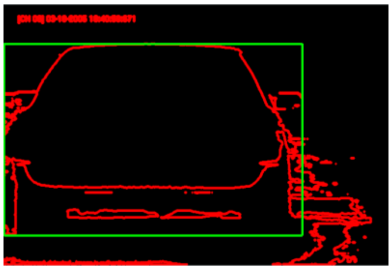
(a) SIFT features (b) Harris corners

Fig. 2. Interest points detected by different feature detectors

4.2 Interest Point Matching

We investigate two different methods for matching interest points. The first was Lowe's implementation of a SIFT feature matcher, which is available as a software package available from the author. The matcher binary executable did not allow configuration of interest point matching thresholds and constants, limiting our ability to control the sensitivity of the matches.

A second option for interest point matching is a technique we call Fast Normalized Cross Correlation. This technique finds the correlation of gradient values between regions of the two images, resulting in a gamma value in the range [-1, 1], where a value of -1 represents the correlation of two perfectly anti-match regions and 1 represents the correlation of two perfectly matched regions. We chose to find the correlation of gradient

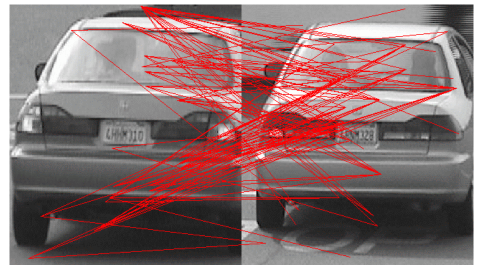


(a) Box drawn around largest contour in the image

(b) Location of car detected using contour segmentation

Fig. 3. Contour Segmentation

(a) Matches found by SIFT feature matcher



(b) Matches found by Fast Normalized Cross Correlation

Fig. 4. Different feature matching techniques

values rather than image pixel values to minimize the effect of illumination difference between regions of the two images. The invariance in illumination changes is important because images of cars are generally taken outdoors where lighting condition changes dramatically due to the time of the day and condition in which the pictures were taken.

4.3 Inlier Extraction

Images of unknown cars are assumed to be taken on the streets or in a parking lot. This presents the problem of having a background scene in the image that can greatly affect the relevance of interest points that are detected. A method that we investigated to eliminate outliers, or interest points not associated with the car, is to use contour segmentation. This method consists of using an edge detector and finding all contours in the image. After all the contours in the image are found, the largest contour by pixel area is assumed to be associated with the car, assuming that the input images are focused on the car. The contour is then enclosed in a rectangular region, and any interest points detected outside of this region will be disregarded as they do not correspond to the car.

Another method that was explored was mirror symmetry. Exploiting the fact that most today cars are symmetrical across the y-axis, interest points from one half image of a car are compared to the interest points on the other half image. This forces matching SIFT interest point matching to be matched to a similar feature located on the opposite side of the car.

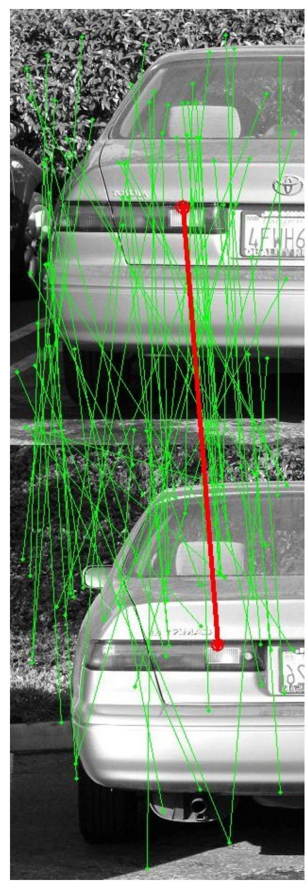


Fig. 5. Matching interest points found by mirror symmetry

After pairing interest points between the two half-images of the same car, outliers caused by the background scene or noise can be identified by finding a subset of matches that fit a model where the relative location of the pairs are similarly displaced in the two half-image relative locations. Fig. 5 shows the result of mirror symmetry interest point comparison.

Though mirror symmetry helps in identifying and removing outliers, it is not an exhaustive algorithm for the needs of system. If only the left or right half of two car images were compared to each other, dents, deformities, or other derivations in appearance unique to a specific car (i. e. bumper stickers, etc) can cause interest points to mismatch with interest points that would have matched otherwise, affecting the accuracy of MMR.

This led us to research RANdom SAmpling Consensus (RANSAC) [2]. RANSAC is a method in which inliers and outliers can be identified given a mathematical model to which to fit the data onto. We used an affine transformation model to extract inliers and discard outliers in our case of MMR. RANSAC is applied to a pair of images to be compared randomly choosing three pairs of matched interest points, required by the affine transform. These three pairs of interest points are fit to an affine transformation, and the transformation is applied to the rest of the interest points of the two images. Inliers are identified by selecting interest points in one image that are within a threshold distance away from its matched interest point of the other image, with both images adjusted by displacing interest points of each image by the centroid of each image to remove differences in translation of the interest points.

The process of selecting three random matched interest points to perform a test transformation is repeated until the inlier count exceeds a threshold value. We have found empirically for this threshold value to be at 80% of the total number of matched interest points to produce a good affine transformation from the query image to an image in the database.

Using RANSAC in addition to SIFT ensures that the interest points found by SIFT are located in the same relative location for a higher rate of accurate matching. We chose to use an affine transform over a homography transform or others transforms because it takes into account slight angle differences between the test car image and the dataset car image while being easier to implement than other transformations. Also, images of cars in our database vary in angle such that an affine transformation is sufficient.

The algorithm begins by subtracting the calculated centroid value, specific to the current set of matched interest points, from the x- and y-axis of the points to create matrices of normalized points.

(a) Interest points detected in image 1. Three yellow circles indicate randomly chosen points for affine transformation

(b) Interest points detected in image 2 in red. Three yellow points matching chosen points in image 1 above are chosen in this image.

(c) Displacement vectors shown between image 1 and image 2

Fig. 6. RANSAC and affine transform

Three random sets of points are chosen to be the initial coordinate space. Using these three sets of points, a transform matrix is calculated and applied to the rest of the points. The rest of the points are then eliminated if they are found to be outliers (distance between the two points is farther than a set threshold value). If the fraction of the remaining inliers to number of total matches is above the set threshold, then the two images are considered a good match. If the fraction is too low, another random set of three points are chosen to form a new transform matrix. These steps are repeated until a transform matrix yielding enough inliers is found, or until a certain number of iterations are run. The top three best matched models are determined at the end of the algorithm. Fig. 6 shows the result of using RANSAC with an affine transformation model on the car database images.

Warping the best matched dataset images using the transform matrix will result in a visual representation of the two car images having interest points at the same location if the two images were placed on top of each other with a common centroid center. The MMR of an unknown car can be determined as long as an image of the same or similar make and model of a similar profile view exists in the database for that car.

Fig. 7. Images of six unique Honda Accords 1998-2002 model

5 Results

With the variety of methods available for use, we performed multiple tests of different combinations of techniques on six images of similar but unique Honda Accord model from 1998-2002, shown in Fig. 7, striving to find a system with the highest rate of matching correct matches.

Implementing SIFT interest point detection with the SIFT interest point matcher was a natural choice due to its invariance over image translation, scaling, and rotation. SIFT interest points are also known to be highly distinctive and easy to extract. However, for the hundreds of interest points detected by SIFT per image that were passed to the SIFT interest point matcher, only at most approximately 30 pairs of matched interest points were found. In most cases, only four matches were found, which is the minimum number of matched pairs needed by RANSAC in order to calculate an affine transform matrix. With our script exiting before it even reached the RANSAC loop most of the time, we knew we had to look into using a different interest point detector, different interest point matcher, or both.

Next we tried a combination of Harris corner detection with Fast Normalized Cross Correlation for matching interest points. Detected corners of the Harris corner detector on multiple car images of the same make and model were located in the same relative area, suggesting similar matching results as SIFT interest points. The problem we found with the Harris corner detection technique was dozens of interest points were detected in clumps. This did not give us many distinct interest points to work with.

A problem with the SIFT interest point matcher was too strict and was not able to match features that appeared similar to an observer and was located in the same relative location of the car. Therefore, Fast Normalized Cross Correlation was used to match the detected interest points.

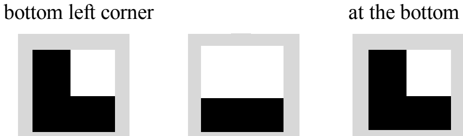
The largest problem we ran across when trying to implement Fast Normalized Cross Correlation in Matlab was in dealing with special cases when the cropped region, needed for the computation of correlation values, on an image would lie beyond the border of the image. This occurs when the interest point lies very close to the edge of the image. To solve this problem, we define a mask for each template, also having the size n x n. This mask holds "1" value in elements of the matrix that lie on the image and "0" value in indices that went off the border. We then perform element-by-element multiplication of the two masks to form the final mask for the query image and the image of in the database to obtain the region of which both image templates has valid gradient data. The template sizes needed to be the same size in both images in order to calculate correlating gradient values. Fig. 8 shows an example of a special case to be handled in this way.

The results of the Fast Normalized Cross Correlation matcher were not any better than the SIFT matcher. Only a handful of matched pairs would pass the gamma threshold of 0.5. This indicated to us that Harris Corner Detector/Fast Normalized Cross Correlation was not the best pair of methods to use.

Since the interest points found by the Harris corner detector were not highly distinctive, we decided to revert to using the SIFT interest point detector, which gave us a widespread set of points on the car image to use. Since the SIFT interest point matcher was too strict and we were unable to modify the parameters of detection in the executable binary, we decided to use the Fast Normalized

(a) image 1 with test crop on bottom left corner

(b) image 2 with test crop at the bottom



(c) image 1 crop mask

(d) image 2 crop mask

(e) mask1 & mask 2

Fig. 8. The white section of mask 1 (c) represents the region with data in image 1. Mask 2corresponds to image 2. The white section of mask 1& mask 2 represents the region of which both images have valid data.

Cross Correlation method to compare and match the interest points. The result of this algorithm is our final system:

Step 1: detect interest points using SIFT interest point detector

Step 2: find matching interest points based on appearance using Fast Normalized Cross Correlation

Step 3: use set of matched pairs of interest points from Step 2 and remove pairs that do not match based on RANSAC using the affine transformation as the model

A remaining problem consists of the location of matched pairs. As shown in Fig. 9, the final set of matched pairs are all located on a single restricted area of the car, even when features are detected in all regions of the car. Though we have proposed a technique to implement a database that will have the ability to accomplish MMR given an image of a car from an arbitrary angle of view, we have not reached that point of the project given the duration of the quarter.

Fig. 9. Matched interest points based on final system.

6 Conclusion

This paper presents a technique in accomplishing MMR for images of cars with an arbitrary angle of view. Car make and model recognition is a fairly unexplored field in machine vision, but some progress has been made in several areas that can help further the studies in this field. Our work in MMR should provide some insight for those looking into this field in the future on the different techniques that are available and applicable to MMR.

Future work on this project should focus on improving the interest point matching results and the building of the database image reader and comparator to a query image to find the best match.

In retrospect, many of the topics in machine vision that we learned for this project can be used for many other CSE190a projects, such as that of 3D Photography [13] from previous quarters. As such, we have progressed much in learning how to accomplish many other machine vision tasks.

7 Acknowledgement

Thanks to Professor Serge Belongie and the Winter 2008 CSE190a for the valuable advices and help we have received throughout this project. Much appreciation also goes to Louka Dlagnekov and Nadav Ben-Haim for generously providing their source code and help.

8 References

[1] L. Dlagnekov. Video-based Car Surveillance: License Plate, Make, and Model Recognition. *Master's Thesis*, University of California, San Diego, 2005.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM*, 24(6):381-395, 1981.

[3] David G. Lowe, Object recognition from local scale-invariant features, *International Conference on Computer Vision,* Corfu, Greece, pp 1150-1157, Sep. 1999.

[4] C. Harris and M.J. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, pp 147-152, 1988.

[5] C. P. Papageorgiou and T. Poggio. A Trainable Object Detection System: Car Detection in Static Images. *CBCL Paper #180/AI Memo #1673*, Massachusetts Institute of Technology, Cambridge, MA, April 1999.

[6] G. D. Sullivan, K. D. Baker, A. D. Worrall, C. I. Attwood, and P. R. Remagnino. Model-Based Vehicle Detection and Classification Using Orthographic Approximations. *Proc of 7th British Machine Vision Conf.*, (2), pp 695-704, 1996.

[7] B. Leung. Component-based Car Detection in Street Scene Images. *Master's Thesis*, Massachusetts Institute of Technology, 2004.

[8] B. Morris and M. M. Trivedi, Robust Classification and Tracking of Vehicles in Traffic Video Streams, *IEEE International Intelligent Transportation Systems Conference,* Sep. 2006.

[9] A. Mohan, Object detection in images by components, A. I. Memo / C. B. C. L Paper 1664, Center for Biological and Computational Learning, MIT, Cambridge, MA, 1999.

[10] B. Heisele, T. Poggio, and M. Pontil. Face detection in still gray images. A. I. Memo / C. B. C. L Paper 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA, 2000.

[11] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* volume 1, pages 657-662, Hawaii, 2001.

[12] D. A. Torres. More Local Structure Information for Make-Model Recognition. CSE252C Project, UCSD. Fall 2005.

[13] K. Kho. 3D Photography. CSE190a Project, UCSD. Winter 2007.