# Investigating 3-D Model and Part Information for Improving Content-Based Vehicle Retrieval

**M. R. Hire[1], S. D. Pable[2]**

**[1,2]Dept. of Electronics and Telecommunication**
**Matoshri college of Engineering and Research Centre, Nashik, India**

## Abstract

Content based vehicle retrieval, a technique which uses visual contents to search images from the large scale image databases, is an active area of research for the past decade. Due to large variation in viewing angle position, illumination, occlusion, background, traditional vehicle retrieval is extremely challenging. This approaches problem in a different way by converting 2D image database into 3D database. For training we have used 77 images in database. First we convert 2D query image into its 3D view and then comparing with 3D database. Then proposed a model fitting approach with weighted jacobian system which leverage the prior knowledge of part information and shows promising improvements. For comparison similarity metrics are used. It contains the analysis done after the application of similarity measure named Euclidian distance, Minkowski Distance, cosine etc. Support vector machine is used which is statistical classification algorithm that classifies data by separating two classes with the help of a functional hyper plane and it helps to improve performance of our system. We have evaluate precision/recall show that this method is very effective.

***Keywords:*** *3D database construction,3D model fitting, Feature extraction, Similarity measures,SVM*

## 1. Introduction

The image retrieval methodology is an extremely sensitive and complex task that relies upon information gathered through techniques that is two dimensional. The use of vehicles in many cities has increased rapidly, especially in recent years, due to urbanization and modernization, and thus, traffic congestion in cities has become a major issue. Therefore, control of vehicles and identification of traffic violators to maintain discipline is becoming a necessary task in many cities. Current system is 2d system which faces problems such as illumination, occlusion, discontinuity in the signals and background noise. Shadow or illumination possibly makes edge detectors or feature detectors incur more noises. Occluded parts cause discontinuity and incompleteness in shape. Background noise and shape variations are potential difficulties when detecting or segmenting vehicles. So, there is a need to develop a system which overcomes the limitations cause due to 2D technique. Hence, the proposed system uses 2D to 3D conversion for image retrieval by using similarity metrics and SVM classifier.

The conventional method in content based image retrieval for surveillance image data classification is done by two dimensional methodologies. This may result in misclassification of data. Sometime these types of problem identification are impractical for large amounts of data as well as for noisy data. A noisy data may be produced due to some technical fault or by human errors and can lead misclassification of 2D image data. Accurate and fast automatic detection of correct vehicle from large database is an important task now days. To overcome this problem, there are various learning methods. One of the method is to classify data from support vector machine classifier and to overcome a drawback of illumination occlusion of 2D system is to use rigid dataset. These methods are good at solving real-world ambiguities. Hence systems which harness power of 3D is developed to overcome the problem of 2D system.

Vehicle make and model recognition (MMR) or content based vehicle retrieval is a relatively new research problem. The basic objective 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.

The objective of content-based image retrieval (CBIR) 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.

## 2. Related Work

“Person attribute search for large area video surveillance”[2] provide a useful way for security personnel to shift through large quantities of video data to find particular person of interest ,given a recent observation report describing that person.SVM classifier is used which shows effective performance.

"Content-Based Image Retrieval Using Shape and Depth from an Engineering Database" presented conversion of an image and uses the shape information in an image along with its 3D information. A linear approximation procedure that can capture the depth information using the idea of shape from shading has used.

"Effect of Similarity Measures for CBIR Using Bins Approach"presented suitable similarity measure for content based image retrieval. It contains the analysis done after the application of similarity measure named Minkowski Distance from order first to fifth.

"Back to the future"[3] of Learning shape models from 3D CAD data formulated transition between the 3D geometry of objects and 2D representations. In this area, they go back to the ideas from the early days of computer vision, by using 3D object models as the only source of information for building a multi-view object class detector and the techniques based on a multi-view detection data set

## 3. Proposed System

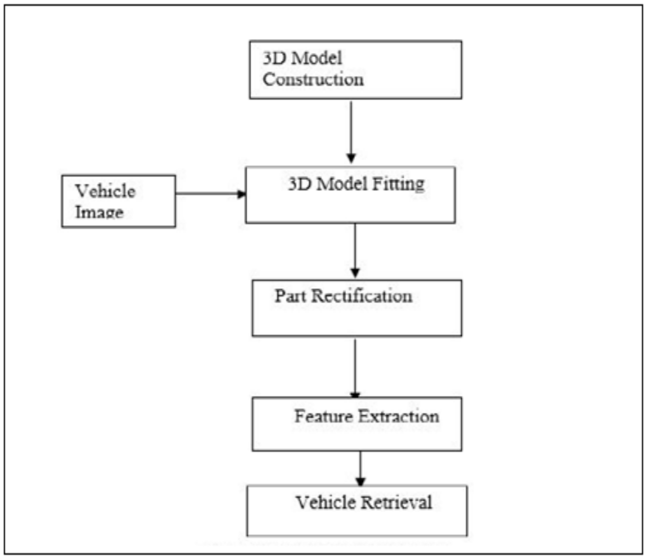


Fig.3.1 Proposed System Development Design Flow

Fig 3.1 shows actual design flow of proposed system. First 2D image dataset is used and then that dataset converted into 3D which forms a new databset.As the query image is given, it then compare with new 3D dataset. Fitting algorithms are used to perform this process. Next stage is of part rectification where car parts are rectify and final features are extracted.

3.1 2D Image to 3D Conversion:

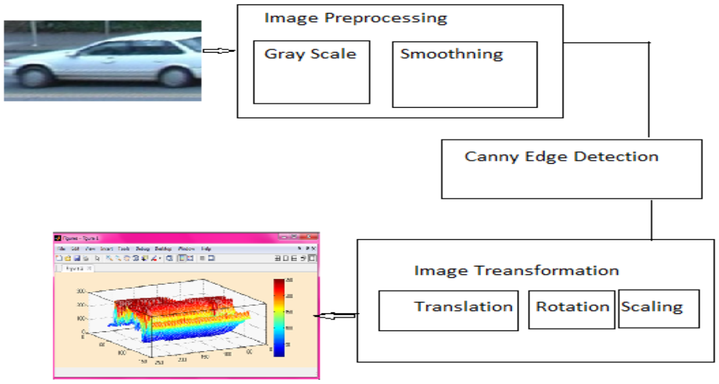
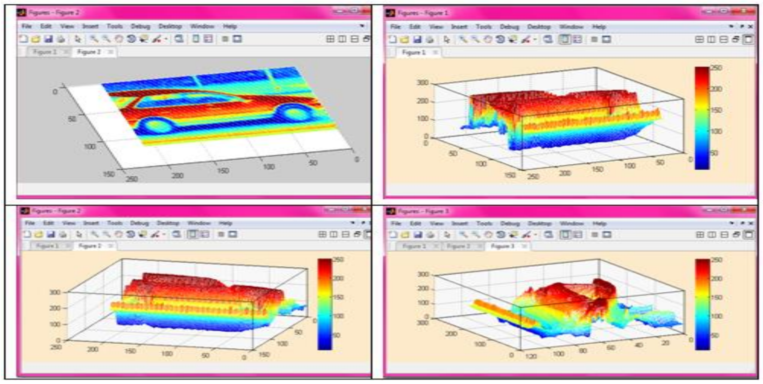
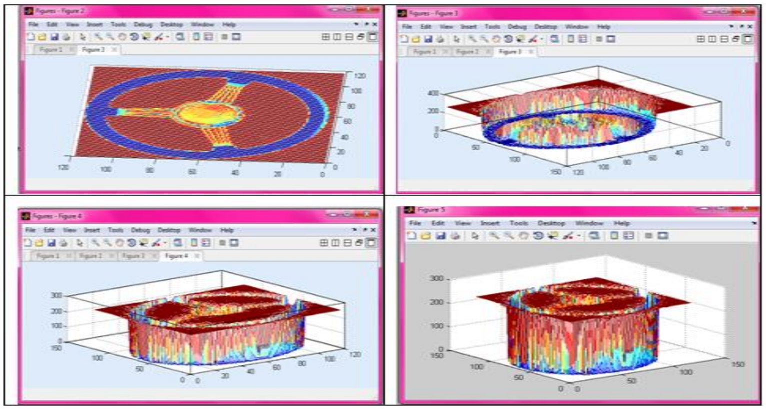


Figure 3,2 2D to 3D Conversion [11]





3.2 3D Vehicle Model Fitting Approaches:

In the model fitting step, we assume that initial position and pose of a vehicle in an image can be estimated. In order to extract parts of vehicles, model fitting is essential.

In the model fitting step, we assume that initial position and pose of a vehicle in an image can be estimated. It is reasonable since detecting the direction and location of the moving vehicles or some multi view object detection can be derived by many promising approaches. Content-based vehicle retrieval is based on the information about the target vehicle. That is, we must detect and estimate the pose of our target before we deal with this object.

In order to extract parts of vehicles (fig 3.1), model fitting is essential. Therefore, we investigate and compare two different state of-the-art approaches in [10] and [11]. One depends on PR and the other solves a weighted Jacobean system. The two approaches have not been compared before this paper, so we try to implement them and do several sensitivity tests to see their capabilities in different configurations. Moreover, we propose to leverage the prior knowledge of semantic parts (e.g., grille, lamp, and wheel) and further improve the challenging 3-D alignment problem.

3-D vehicle model fitting procedure describes in Fig 3.2 Given the initial pose and shape, then generate the edge hypotheses by projecting the 3-D model into a 2-D image and remove hidden lines by using depth map rendered from the 3-D mesh. For each projected edge point, the corresponding points are found along the normal direction of the projected edges. Then, a 3- D model fitting method is performed to optimized pose and shape parameters. The above procedure is repeated several times until convergence.

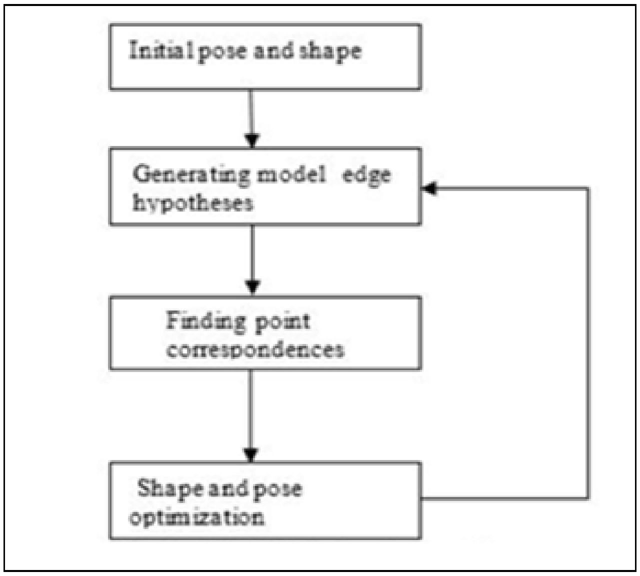

Initial pose and shape
Generating model edge hypotheses
Finding point correspondences
Shape and pose optimization

Fig 3.3 Model fitting approach design flow

Table 1 Observation Table

| **METHOD** | **APD** | **STD** |
|---|---|---|
| Initial Location | 47.15 | 6.06 |
| PR(KC) | 39.26 | 9.90 |
| PR(Rigid CPD) | 29.59 | 6.91 |
| PR(Nonrigid CPD) | 26.53 | 6.84 |
| JS | 34.19 | 6.31 |
| **WTD JS (ours)** | **14.46** | **13.27** |

### 4. Similarity Measurement:

Retrieval result is not a single image but a list of images rank by their similarity with query image.since CBVR is not based on exact matching.for a model shape indexed by FD features $fm=[fm^1,fm^2,\ldots\ldots\ldots fm^N]$ and database index by FD feature $fd=[fd^1,fd^2,\ldots\ldots\ldots fd^N]$ the euclidean distance between two feature vectors can be then used as a simillarity measurement d is

$$\sqrt{\sum_{i=0}^{N-1} |\mathrm{fim}-\mathrm{fid}|2} \qquad (4.1)$$

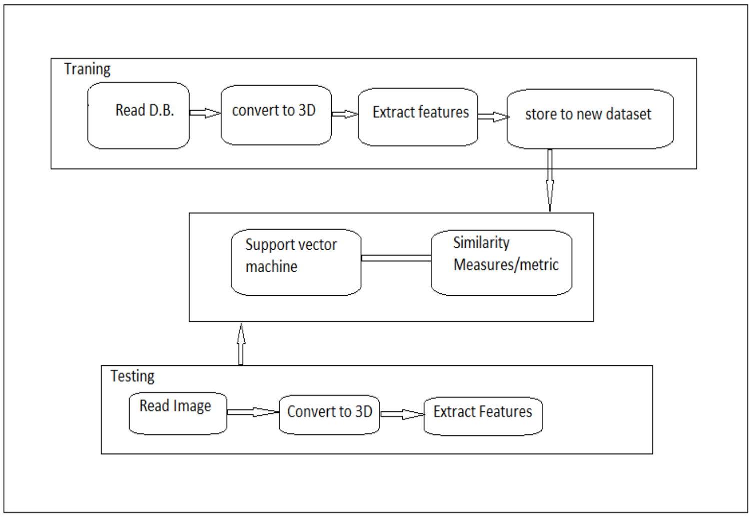


Fig 4.1 Shows Use Of Similarity Metrics between Training and Testing

Fig (4.1) shows that once the feature vector databases are ready we can fire the desired query to retrieve the similar images from the database.to facillitate this,retrieval system has to perform the important task of applying the similarity measures so that distance between query image and database image will be calculated and images having less distance will be retrieve in the final set.

### 4.1 L1 (.Manhattan distance)

The Manhattan distance computes the sum of difference in each dimension of two vectors in n dimensional vector space. It is the sum of the absolute differences of their corresponding components. Manhattan distance is also called the $L_1$ distance. If $u = ( x_1 , x_2 ....x_n )$ and $v = ( y_1 , y_2 .....y_n )$ are two vectors in n dimensional hyper plane, then the Manhattan Distance $MD(u, v)$ between two vectors u, v is given by the equation 4.2.

$$MD\ (u,\ v)\ =\ |\ x1\ -\ y_1\ |\ +|\ x_2\ -\ y_2\ |+....\ +|\ x_n\ -\ y_n\ | \quad (4.2)$$

Now for two RGB scale images of size pxq,I1(a,b,c) and I2(a,b,c) where a=1,2……p,b=1,2…….p and c =1,2,3 where c represent color intensity values Red , Green , Blue respectively.Manhattan distance is measured using equation 4.3

$$MD(I_1, I_2) = \sum_{a=1}^{p}\sum_{b=1}^{q}\sum_{c=1}^{3}\left|I_1(a,b,c) - I_2(a,b,c)\right| \quad (4.3)$$

As the number of pixels, n which falls in skin region varies with varying size of the image, so rather than taking the absolute distance further the distance is being normalized using equation 4.4

$$MD_1(I_1, I_2) = \frac{MD(I_1, I_2)}{n} \quad (4.4)$$

where n= number of pixels considered

### 4.2 L2 (Euclidean Distance)

It is also called the L2 distance. For the same two vectors in n dimensional hyper plane
$u = ( x_1 , x_2 ....x_n )$ and $v = ( y_1 , y_2 .....y_n )$ the Euclidean Distance $ED(u,v)$ is defined as euation 4.5

$$ED(u,v) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + .. + (x_n - y_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4.5)$$

And for the same two RGB images $I1(a, b, c)$, $I\ 2\ (a, b, c)$ , Euclidean Distance is measured using equation 4.6

$$ED(I_1, I_2) = \sum_{a=1}^{p}\sum_{b=1}^{q}\sqrt{\sum_{c=1}^{3}(I_1(a,b,c) - I_2(a,b,c))^2} \quad (4.6)$$

Further the Euclidean distance is normalized using equation 4.7

$$ED_1(I_1, I_2) = \frac{ED_1(I_1, I_2)}{n} \qquad d = \sqrt{\sum |\ x_i - y_i\ |^2} \quad (4.7)$$

### 4.3 Cosine Correlation

The cosine correlation distances function can be expressed as follows

$$\frac{(D(n))\bullet(Q(n))}{\sqrt{\left[|D(n)|^2\,|Q(n)|^2\right]}} \quad (4.8)$$

Where D(n) and Q(n) are database and query feature vectors respectively.Correlation measures in general are invariant to scale transformations and tend to give the similarity measure for those feature vectors whose values are linearly related. In Figure 4.2. Cosine Correlation distance is compared with the Euclidean distance. We can clearly notice that Euclidean distance ed2 > ed1 between query image QI with two database image features DI1and DI2 respectively for QI. At the same time we can see that θ1 > θ2 i.e distance L6 for DI1 and DI2 respectively for QI.

If we scaled the query feature vector by simply constant factor k it becomes k.QI ; now if we calculate the ED for DI1 and DI2 with query k.QI we got ed1' and ed2' now the relation they have is ed1' > ed2' which is exactly opposite to what we had for QI. But if we see the cosine correlation distance; it will not change even though we have scaled up the query feature vector to k.QI. It clearly states that Euclidean distance varies with variation in the scale of the feature vector but cosine correlation distance is invariant to this scale transformation.

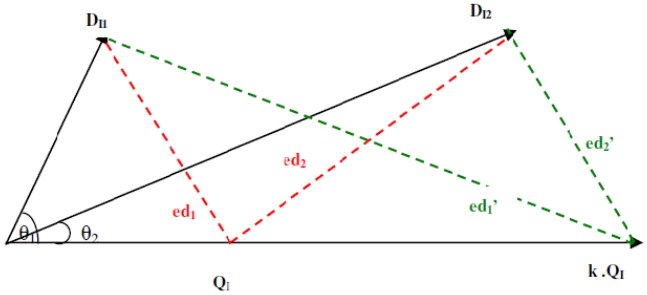


Fig 4.2 Comparison of Euclidean and Cosine Correlation Distance

### 4.4 chebyshev

In chebyshev similarity measure the minimum distance calculated as follow

$$d = \max_i |\ x_i - y_i\ | \quad (4.9)$$

### 4.5 minkowski

$$L_p\ (p \geq 1 where\ p = 1,2,3.....\infty)$$

$$d = \sqrt[p]{\sum_{i=1}^{n}|\ x_i - y_i\ |^p} \quad (4.10)$$

Within this family very few have been used in image retrieval and they are Euclidean L2, City block L1(taxicab norm, Manhattan) and Chebyshev L∞ dissimilarity formulas.

### 4.6 Cityblock

In cityblock similarity measure the minimum distance calculated as follow

$$d = \sum_{i=1}^{n} | x_i - y_i | \qquad (4.11)$$

## 5. Simulation Results:

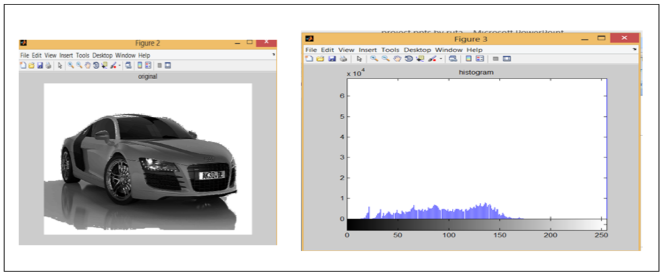


Figure 2
File Edit View Insert Tools Desktop Window Help
original
Figure 3
File Edit View Insert Tools Desktop Window Help
histogram
x 10^4
6
5
4
3
2
1
0
0
50
100
150
200
250

Figure 5.1 Histogram plot for original image

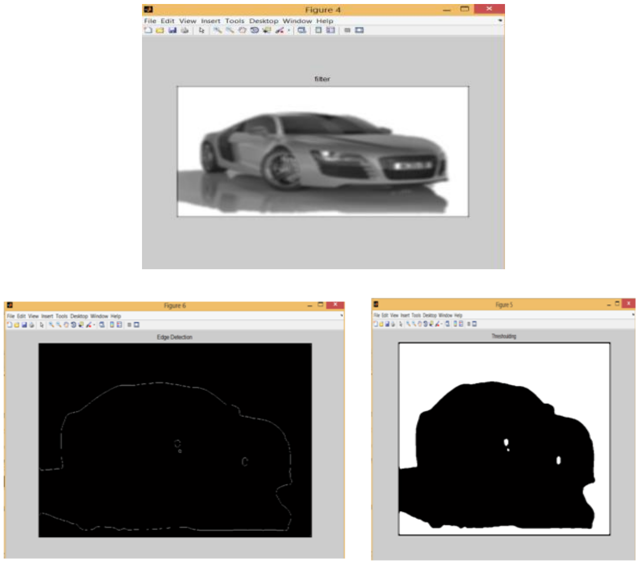


Figure 5.2 Filtered image, Thresholding and canny edge detection output

### 5.1 Expected Values And GUI Result:

In this experiment, we apply several descriptors on extracted parts which are resized to the same number of pixels while keeping the ratio between height and width. Those retrieved vehicle instances which have the same label as the query instance are correct. We compare the mean average precision(MAP) performance on different sources including a whole vehicle image and three parts, grilles, lamps, and the most visible wheel. Then, we do sensitivity tests to select the late fusion weights and obtain the best parameters. We test on three state-of-the-art feature descriptors. Firstly, Difference of Gaussian (DOG) detector and SIFT descriptor are used[25]. For constructing the visual word vocabulary, we start by applying SIFT descriptor on the images of three informative parts. Each descriptor contains 4 × 4 cells with 8 orientation bins, resulting in 128 dimensional feature vectors. Table III Performance (in MAP) for Vehicle Retrieval Experiments**.**

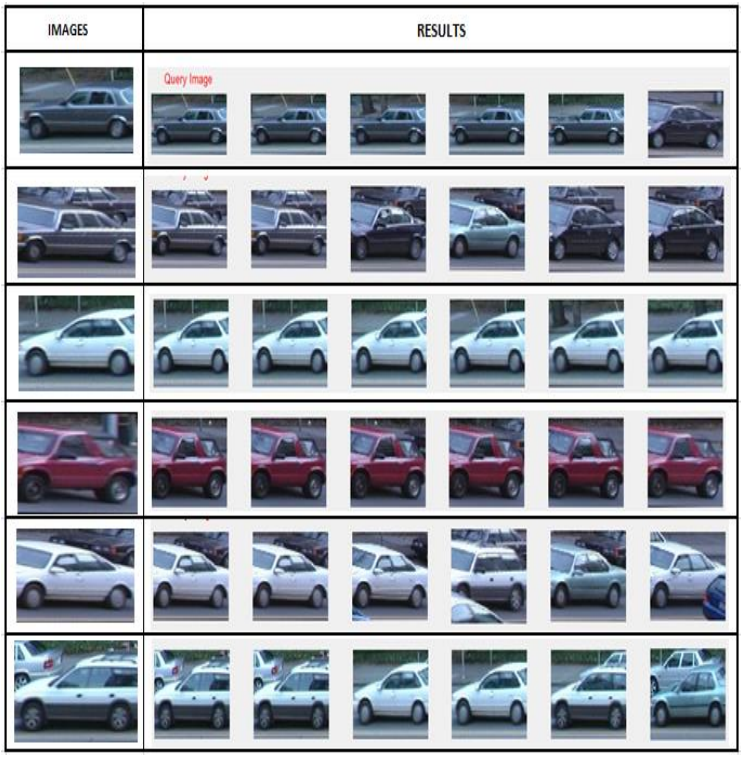
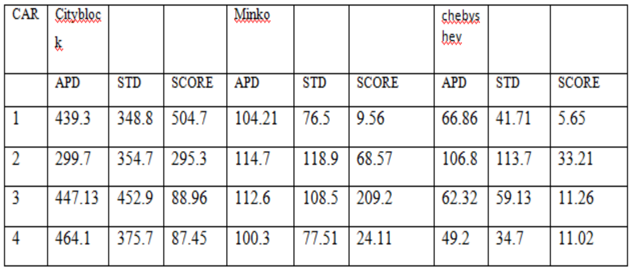
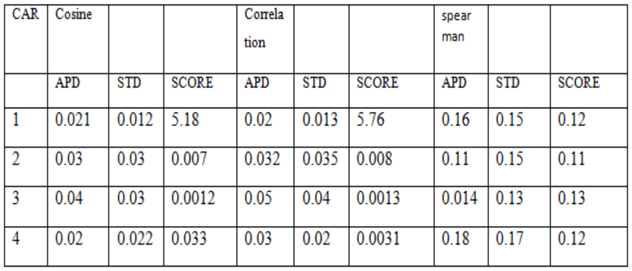


Table 1: Result of Images Other Than Dataset:

| CA<br>R | L1 | | | L2 | | | Nr.<br>L2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCOR<br>E |
| 1 | 16.56 | 15,63 | 7.29 | 104.2 | 76.53 | 9.56 | 24.5 | 18.06 | 2.25 |
| 2 | 11.73 | 13.21 | 11.28 | 114.7 | 128.9 | 68.57 | 21.4 | 28.08 | 12.81 |
| 3 | 14.46 | 13.27 | 7.5 | 112.6 | 108.7 | 20.29 | 22.1 | 21.30 | 3.98 |
| 4 | 21.11 | 20.13 | 7.38 | 100.3 | 77.51 | 21.11 | 22.8 | 17.67 | 5.89 |



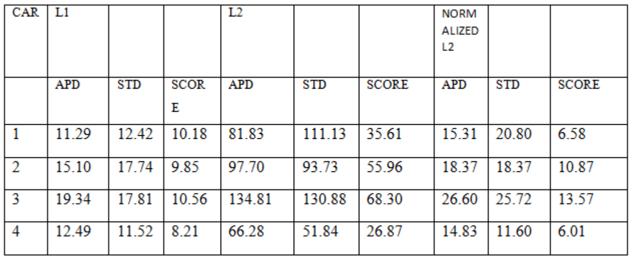
| CAR | Cityblock | | | Minko | | | chebyshev | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCORE |
| 1 | 439.3 | 348.8 | 504.7 | 104.21 | 76.5 | 9.56 | 66.86 | 41.71 | 5.65 |
| 2 | 299.7 | 354.7 | 295.3 | 114.7 | 118.9 | 68.57 | 106.8 | 113.7 | 33.21 |
| 3 | 447.13 | 452.9 | 88.96 | 112.6 | 108.5 | 209.2 | 62.32 | 59.13 | 11.26 |
| 4 | 464.1 | 375.7 | 87.45 | 100.3 | 77.51 | 24.11 | 49.2 | 34.7 | 11.02 |



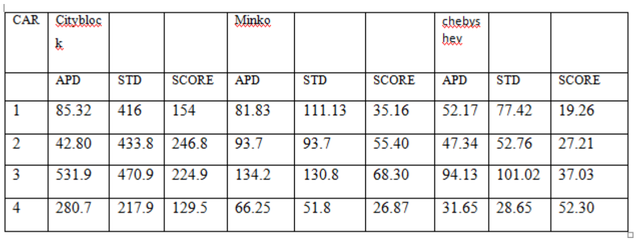
| CAR | Cosine | | | Correlation | | | spearman | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCORE |
| 1 | 0.021 | 0.012 | 5.18 | 0.02 | 0.013 | 5.76 | 0.16 | 0.15 | 0.12 |
| 2 | 0.03 | 0.03 | 0.007 | 0.032 | 0.035 | 0.008 | 0.11 | 0.15 | 0.11 |
| 3 | 0.04 | 0.03 | 0.0012 | 0.05 | 0.04 | 0.0013 | 0.014 | 0.13 | 0.13 |
| 4 | 0.02 | 0.022 | 0.033 | 0.03 | 0.02 | 0.0031 | 0.18 | 0.17 | 0.12 |

For testing four images are used other than dataset whose APD,STD,Score values are being calculated. average pixel distance and standard deviation is different for didderent image.

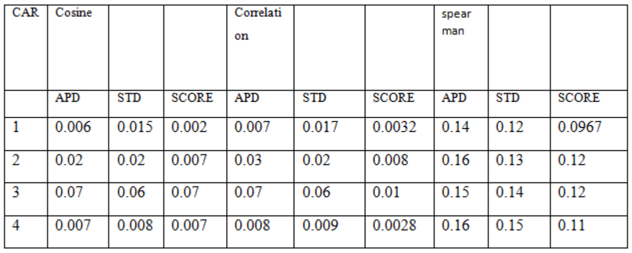
Table 2: Result of Dataset Images



| CAR | L1 | | | L2 | | | NORMALIZED L2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCORE |
| 1 | 11.29 | 12.42 | 10.18 | 81.83 | 111.13 | 35.61 | 15.31 | 20.80 | 6.58 |
| 2 | 15.10 | 17.74 | 9.85 | 97.70 | 93.73 | 55.96 | 18.37 | 18.37 | 10.87 |
| 3 | 19.34 | 17.81 | 10.56 | 134.81 | 130.88 | 68.30 | 26.60 | 25.72 | 13.57 |
| 4 | 12.49 | 11.52 | 8.21 | 66.28 | 51.84 | 26.87 | 14.83 | 11.60 | 6.01 |



| CAR | Cityblock | | | Minko | | | chebyshev | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCORE |
| 1 | 85.32 | 416 | 154 | 81.83 | 111.13 | 35.16 | 52.17 | 77.42 | 19.26 |
| 2 | 42.80 | 433.8 | 246.8 | 93.7 | 93.7 | 55.40 | 47.34 | 52.76 | 27.21 |
| 3 | 531.9 | 470.9 | 224.9 | 134.2 | 130.8 | 68.30 | 94.13 | 101.02 | 37.03 |
| 4 | 280.7 | 217.9 | 129.5 | 66.25 | 51.8 | 26.87 | 31.65 | 28.65 | 52.30 |



| CAR | Cosine | | | Correlation | | | spearman | | |
|---|---|---|---|---|---|---|---|---|---|
| | APD | STD | SCORE | APD | STD | SCORE | APD | STD | SCORE |
| 1 | 0.006 | 0.015 | 0.002 | 0.007 | 0.017 | 0.0032 | 0.14 | 0.12 | 0.0967 |
| 2 | 0.02 | 0.02 | 0.007 | 0.03 | 0.02 | 0.008 | 0.16 | 0.13 | 0.12 |
| 3 | 0.07 | 0.06 | 0.07 | 0.07 | 0.06 | 0.01 | 0.15 | 0.14 | 0.12 |
| 4 | 0.007 | 0.008 | 0.007 | 0.008 | 0.009 | 0.0028 | 0.16 | 0.15 | 0.11 |

## 5.2 PERFORMANCE EVALUATION:

**Precision and Recall:**

| Dataset images | Precision | Recall |
|---|---|---|
| 77 | 0.77 | 0.87 |

Results obtained here are interpreted in the terms of PRCP: Precision Recall Cross over Point. This parameter is designed using the conventional parameters precision and recall defined in following equation. According to this once the distance is calculated between the query image and database images, these distances are sorted in ascending order. According to PRCP logic we are selecting first 77 images from sorted distances and among these we have to count the images which are relevant to query; this is what called PRCP value for that query because we have total 77 images of each class in our database. Precision: Precision is the fraction of the relevant images which has been retrieved (from all retrieved) Recall: Recall is the fraction of the relevant images which has been retrieved (from all relevant).

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in datset}}$$

### 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] N. Kumar, P. Belhumeur, and S. Nayar,"Facetracer: A search enginefor large collections of images with faces," in Proc.10th Eur. Conf.Comput Vision: Part IV, 2008 pp. 340–353.

[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] M. Stark, M. Goesele, and B. Schiele, “Back to the future: Learning shape models from 3Dd CAD data,” in Proc. BMVC, 2010, pp. 106.1– 106.11.

[4] M. Arie-Nachmison and R. Basri, “Constructing implicit 3d shape models for pose estimation,” in Proc. ICCV, 2009, pp. 1341–1348

[5] J. Liebelt, C. Schmid, and K. Schertler, “Viewpoint-independent object class detection using 3d feature maps,” in Proc. IEEE Conf. Comput.Vision Pattern Recognit., Jun. 2008, pp. 1–8.

[6] Y. Guo, Y. Shan, H. S. Sawhney, and R. Kumar, “PEET: Prototype embedding and embedding transition for matching vehicles over disparate viewpoints,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., Jun. 2007, pp. 1–8.

[7] J. M. Ferryman, A. Worrall, G. D. Sullivan, and K. Baker, “A generic deformable model for vehicle recognition,” in Proc. 1995 British Machine Vision Conf.(Vol. 1), 1995, pp. 127–136.

[8] D. Koller, K. Danilidis, and H. H. Nagel, “Model-based object tracking in monocular image sequences of road traffic scenes,” Int. J. Comput. Vision, vol. 10, no. 3, 1993, pp. 257–281.

[9] S. M. Khan, H. Cheng, D. Matthies, and H. S. Sawhney, “3D model based vehicle classification in aerial imagery,” in Proc. IEEE

[10] M. J. Leotta and J. L. Mundy, “Predicting high resolution image edges with a generic, adaptive, 3-D vehicle model,” in Proc.IEEE Conf. Comput. Vision Pattern Recognit., Jun. 2009, pp. 1311–1318.

[11] Y. Tsin, Y. Genc, and V. Ramesh, “Explicit 3d modeling for vehicle monitoring in non-overlapping cameras,” in Proc. 2009 Sixth IEEE Int.Conf. Adv. Video Signal Based Surv., Sep. 2009, pp. 110–115.

[12] J. R. Smith and S.-F. Chang, “Visual SEEK: A fully automated content based image query system,” in Proc. Fourth ACM Int. Conf. Multimedia, 1996, pp. 87–98.

[13] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image segmentation using expectation-maximization and its application to image querying,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 8, pp. 1026–1038, Aug. 2002.

[14] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” Int. J. Comput.Vision, vol. 60, no. 1, pp. 63–86, 2004.

[15] R. Feris, B. Siddiquie, Y. Zhai, J. Petterson, L. Brown, and S. Pankanti,“Attribute-based vehicle search in crowded surveillance videos,” in Proc.1st ACM Int. Conf. Multimedia Retrieval, 2011, pp. 18:1–18:8.

[16] V. Petrovic and T. F. Cootes, “Analysis of features for rigid structure vehicle type recognition,” in Proc. British Mach. Vision Conf., 2004, pp. 587–596.

[17] P. Negri, X. Clady, M. Milgram, U. Pierre, and M. Curie-paris, “An oriented- contour point based voting algorithm for vehicle type classification,”inProc. Int. Conf. Pattern Recognit., 2006, pp. 574–577.

[18] F. M. Kazemi, S. Samadi, H. R. Poorreza, and M.-R.Akbarzadeh-T,“Vehicle recognition based on fourier, wavelet and curvelet transforms a comparative study,” in Proc. Third Int. Conf. Inform. Technol.: NewGenerations, pp. 939–940, 2007.

[19] S. Rahati, R. Moravejian, E. M. Kazemi, and F. M. Kazemi, “Vehicle recognition usingcontourlet transform and svm,” in Proc. Fifth Int. Conf.Inform. Technol.: New Generations, 2008, pp. 894–898.

[20] I.Zafar, E. A. Edirisinghe, and B. S. Acar, “Localized contourlet features in vehicle make and model recognition,” in Proc. SPIE, 2009, pp.725 105–725 105–9.

[21] J. Gower, “Generalized procrustes analysis,” Psychometrika, vol. 40, no.1, pp 33–51, 1975.

[22] Y. Tsin and T. Kanade, “A correlation-based approach to robust point set registration,” in Proc. ECCV, 2004, pp. 558–569.

[23] A. Myronenko and X. B. Song, “Point set registration coherent point drift,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 12, pp. 2262–2275, May 2009.

[24] A. E. Beaton and J. W. Tukey, “The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data,” in Proc. Technometrics, 1974, pp. 147–185.

[25] J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” in Proc. ICCV, 2003, pp. 1470–1477.

[26] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in Proc.CVPR, 2006, pp. 475–426.

[27] 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.

[28] J. Yang, Y.-G.Jiang, A. G. Hauptmann, and C.-W. Ngo, “Evaluating bag-of-visual-words representations in scene classification,” in Proc. Int. Workshop Multimedia Inform. Retrieval, 2007, pp. 197–206.