XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6-11, 2009, Lisbon, Portugal

SIFT-BASED MEASUREMENTS FOR VEHICLE MODEL RECOGNITION

A. Psyllos[1], C. N. Anagnostopoulos[2], E. Kayafas[3]

[1]Electrical and Computer Engineering School, National Technical University of Athens, Athens , Greece, psyllos@central.ntua.gr

[2]Cultural Technology & Communication Dpt., University of the Aegean, Mytilene, Greece, canag@ct.aegean.gr

[3] Electrical and Computer Engineering School, National Technical University of Athens, Athens, Greece, kayafas@cs.ntua.gr

Abstract – A SIFT-based Vehicle Manufacturer and Model Recognition (VMMR) method was utilized to tackle the problem of vehicle security. Distinctive parts of the vehicle frontal view such as the headlights, grill and logo area were segmented. A series of experiments were conducted in a variety of outdoor conditions, where a query image that was rotated, scaled, shifted or set in different lighting conditions, matched against a database of model images. In this work, is shown that image processing functions based on Scale Invariant Feature Transform (SIFT) measurements can be used to obtain high performance object features recognition, creating a key-point fingerprint (pattern) for each image class. In the majority of the cases, SIFT method performs very well, in terms of efficiency and robustness .

Keywords : vehicle, recognition, measurement, SIFT

1. INTRODUCTION

Image matching is a fundamental problem in computer vision which occurs in many computer vision applications from a variety of fields including image retrieval for security enforcement and robot navigation. Content Based Image Retrieval (CBIR) addresses matching and retrieval of images sharing similar visual content from a database of images. A common approach to accurate image matching is known as “keypoint” or “interesting point” extraction from the images for comparison. It involves identifying points that can be reliably extracted from different images of the same object or the same category of objects.

Earlier research into invariant keypoints focused on invariance to rotation and translation, Siggelkow [1], Schultz-Mirbach [2]. Scale Invariant Feature Transforms (SIFT) were introduced by Lowe [3], [4], [5] and they are invariant to rotation, translation and scale variation between images and partially invariant to affine distortion, illumination variance and noise. Research related to fully invariant features, published by Brown and Lowe [6], Mikolajczyk and Schmidt [7].

Vehicle classification in general categories is a task that has been adequately addressed in the literature Weber [8], Kato [9], Lai [10], [11]. Approaches related with vehicle model identification have been published previously with encouraging results. Dlangenkov and Belongie [12] utilized Scale Invariant Feature Transform (SIFT) features making them suitable for Vehicle Recognition, using a vehicle database of rear-view vehicle images. Petrovic and Cootes [13] presented an interesting approach for vehicle model recognition and verification that displays the best results in respective tasks. Merler [14] presents a car detection system based on color segmentation and labeling, which performs color recognition. Čonos [15] deals with a vehicle type recognition problem from frontal view images. He proposed a SIFT-based descriptor for feature extraction but his method is computational expensive -in some cases takes more than 12 hours to be accomplished.

In this work, a novel method is proposed, whose aim is to obtain reliable recognition for a vehicle manufacturer and vehicle model, (eg. Alfa Romeo 156), from a frontal view image and using an image database of models. This effort was assisted by a previously developed license plate recognition module Anagnostopoulos [16], C. Anagnostopoulos, I. Anagnostopoulos, Loumos and Kayafas [17] and a special image processing technique, called phase congruency Covesi [18], Psyllos, Anagnostopoulos, Loumos and Kayafas [19], Psyllos, Anagnostopoulos and Kayafas [20].

The recognition method consists mainly of six modules: 1) Vehicle License Plate Recognition, 2) Vehicle Frontal View Segmentation, 3) Vehicle Mask Segmentation, 4) SIFT Matching, 5) Vehicle Manufacturer Recognition and 6) Vehicle Model Recognition as depicted in Figure 1.

2. VEHICLE FRONTAL VIEW AND MASK SEGMENTATION

Vehicle front view image from photo camera or framed video camera sequence, is first converted to greyscale and scaled to a fixed size. The Licence Place Recognition (LPR) module [16], [17] was applied in order to locate the position

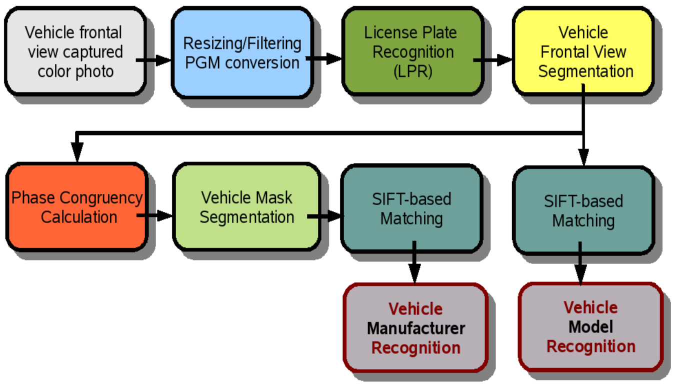

Fig. 1. VMMR System Architecture.

of vehicle license plates and the first segmentation of vehicle image yields the vehicle "mask" which is defined as the frontal view of the vehicle including headlights, grill and manufacturer logo area, see Fig. 2.

Fig. 2. Definition of the vehicle "mask" (RoI) based on license plate geometry.

This mask was further segmented so as to identify and isolate the manufacturer logo. To accomplish this task, we have implemented a method that is based on phase congruency calculation, Psyllos [19], [20] which is a dimensionless measure to assess the existence of significant features. We have used the code provided by Kovesi [18], including the default values proposed in his study and measured a characteristic feature curve for every image, as an "image signature" of the vehicle, which is unique and representative for each of the samples used, see Fig. 3.

3. VEHICLE MANUFACTURER AND VEHICLE MODEL RECOGNITION

SIFT is the state-of-the-art in the field of image recognition and the method of choice for a wide range of applications. It is based on the idea of representing images by a set of descriptors based on gradient orientation histograms. The procedure in brief is as follows: The points of interest, that here will be called keypoints, are located as local peaks in the scale-space [21] of the images and filtered to preserve only those that are likely to remain stable over transformations. The local image gradients are measured at a

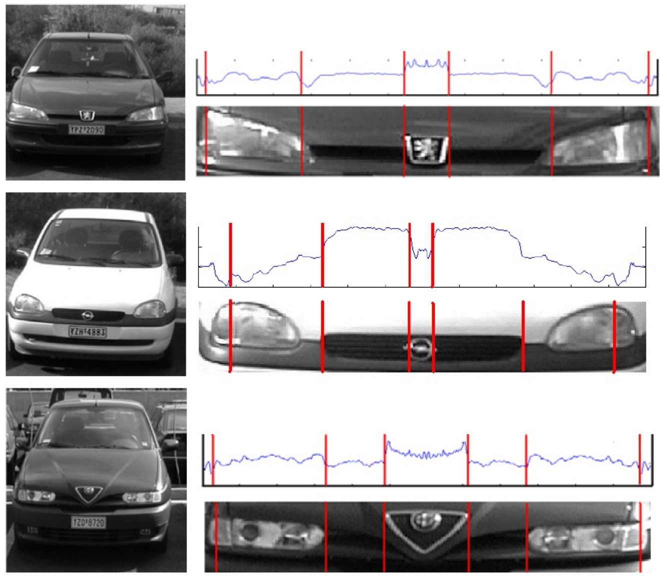

Fig. 3. Typical examples of vehicle mask detection, calculation of image signature for mask segmentation.

neighborhood region (patch) centered about the keypoint location. This patch has been previously rotated on the basis of its dominant keypoint orientation and scaled according to the scale of the detected feature. The keypoint descriptor is created by sampling the magnitudes and orientations of the image gradient in the patch around the keypoint, obtaining an array of histograms, which in the typical SIFT case is a 128-dimensional vector. This descriptor captures the rough spatial structure of the patch and weighted by using a Gaussian window where the nearest sub-regions of the patch are more important than the remote ones.

This descriptor is orientation invariant since it is calculated relatively to the main orientation and scale invariant since it retains the information about the scale of the located keypoint. In order to achieve invariance to illumination changes, the descriptor vector is normalized to unit length. Finally, the descriptor is thresholded so as to remove elements with small values and thus to increase robustness at noise. It is also resilient to small perspective deformations which increase it's robustness for vehicle recognition in non-controlled conditions.

The main contribution of this work, is a modified SIFT method, which omits the step of dominant gradient orientation calculation and the relative rotation of the image patch. We will refer, throughout this text to this method, as V-SIFT. Since most of the vehicle images are with fixed upright (vertical) orientation this modification is expected to be faster without any significant loss of recognition efficiency.

For each keypoint i from the query image, the descriptor is used to find its nearest-neighbor (NN) matches among all stored keypoints from all images in the database. The nearest neighbors were selected to have their Euclidean metric distance smaller than a appropriate threshold. That is, the number of database NN for each keypoint depends on the selected threshold. A KD-Tree data structure was utilised since it has low search time complexity.

The best database image match is further validated using a similarity-based Generalised Hough Transform (GHT) clustering technique, see Ballard [22], Lowe [4], followed by RANSAC [23] for homography calculation and geometric verification.

Another contribution of this work, is the keypoint database model enrichment process,for creating the database and after each successful recognition. Keypoints belonging to common parts of the images were selected and the keypoint descriptors are re-assigned to the position, scale and orientation of the respective database image match. By this technique we substantially increase the number of keypoints for every model, thus making the recognition process more robust in outdoor conditions (low-contrast, partially lighting, reflections, cloudy weather, rain, etc.).

4. EXPERIMENTAL RESULTS

In order to recognise the vehicle model from a query image, first the manufacturer of the vehicle is recognised via logo query matching against a logo database. Then the query image mask, is matched against a vehicle model database for the manufacturer recognised in the previous step. By using this technique, the problem of matching any vehicle model from any manufacturer, is reduced into a set of smaller problems, increasing the speed of recognition by an order of magnitude. The method for keypoint detection and description is the modified SIFT with is non-rotational invariance (V-SIFT).

Manufacturer Recognition

<u>Database set</u>

Sample images have been downloaded from the Medialab LPR Database [24] and contain frontal views from moving and non-moving vehicles captured by a Nikon Coolpix 885 adjusted to acquire 1024 x 768 pixel images. The distance between the camera and the vehicle varied from 2 up to 6 meters at a height of 1.6-1.8 meters from the ground. For the creation of the testing set, 400 manually cropped logo images have been selected from this database, some examples are shown in Fig. 4.

These images correspond to 10 classes, ten classes of selected vehicle manufacturers. Each class contains an equal number of samples (40) and for each of samples the keypoints were detected and descriptors were created. Then from those 40 images, one image was selected by an expert as a reference view and the rest of the 39 sample views were registered according to that reference view using the homography calculated by RANSAC algorithm. Only areas belonging to the common parts of the images were selected and the keypoint descriptors were re-referenced to the new position, scale and orientation. In this way, the number of keypoints for every manufacturer logo is substantially increased by merging the keypoints, thus making the recognition process more robust in illumination conditions. We will call this merging procedure as M-SIFT. Finally, a set of databases, each one per vehicle manufacturer, containing merged keypoints, was created.

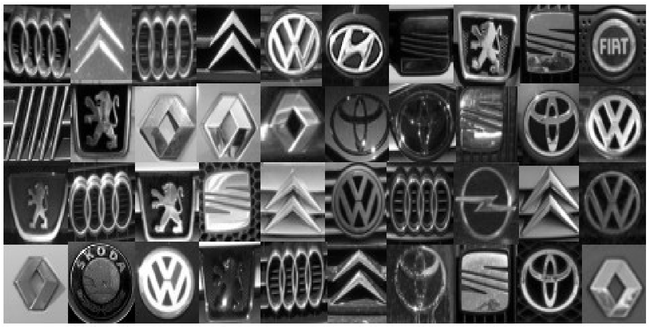

Fig. 4. Examples of segmented logos in the database set.

<u>Query set</u>

Sample query images have been downloaded also from the Medialab LPR Database [24] and with the same characteristics as the database set. We have tested 800 query original images which have been automatically cropped to obtain vehicle masks and manufacturer logos, by successively applying LPR detection and PC calculation, as described in section II. For every image a number of keypoints were detected and their respective descriptors were calculated. Then the query image descriptors were matched against the database ones, using a threshold criterion and finally, the database manufacturer logo belonging to thae biggest cluster of the GHT is considered to be the most similar with the query image and checked for geometric consistency, using RANSAC.

Runs performed in Suse 11.1 Linux on a dual-core Pentium IV, with 2.40 GHz and 3 GB memory. The SIFT procedure is followed as described in the previous sections and we have arrived at the following results, that are shown in Table 1. Examples of the logo matching process are shown in Fig. 5.

Table 1. Vehicle Manufacturer Recognition Rate.

Manufacturer	True	False	No Match
Alfa Romeo	76	4	0
Audi	70	8	2
Bmw	77	3	0
Citroen	80	0	0
Fiat	68	6	6
Peugeot	75	5	0
Renault	76	4	0
Seat	73	5	2
Toyota	78	1	1
Volkswagen	79	1	0
Total	**752**	**37**	**11**
Average (%)	**94**	**5**	**1**

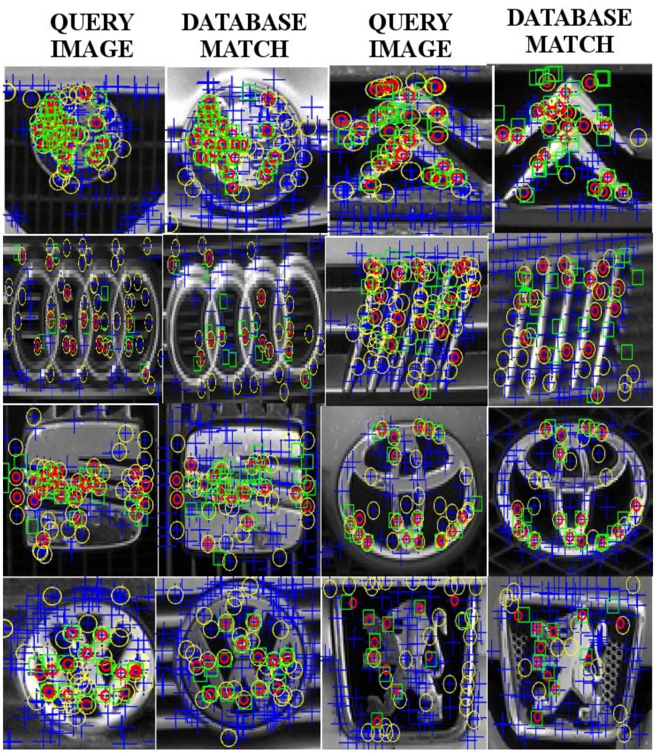

Fig. 5. Examples of successfully matched logos.

Model Recognition

Database set

Sample images are the same as those used for logo recognition [24], with the only difference that they include the mask of the vehicle, instead of the logo. For the creation of the testing set, 400 manually cropped mask images have been selected from this database. These images correspond to 10 classes, ten classes of the selected vehicle manufacturers. Each class contains 40 mask images separated in 10 vehicle models-this means that we have 4 images for each vehicle model. For each of these samples the keypoints were detected and descriptors were created. Then a reference image view was selected by an expert for every model and the rest of the 3 sample views were registered according to that reference view using a homography calculated by a RANSAC algorithm, exactly with same procedure followed in manufacturer recognition section. Finally, a database containing keypoints from 10 vehicle models for every manufacturer was created.

Query set

Sample query images have been downloaded also from the Medialab LPR Database [24] and with the same characteristics as the database set. We have tested 800 query original images which have been automatically cropped to obtain vehicle masks, by successively applying LPR detection, as described in section II. As is the manufacturer recognition, for every image a number of keypoints were detected and their respective descriptors were calculated. Then the query image descriptors were matched against the database ones, using a threshold criterion and finally, the database vehicle model belonging to the biggest cluster of the GHT is considered to be the most similar with the query image and checked for geometric consistency, using RANSAC. Examples of the vehicle model matching process are shown in Fig. 6.

Table 2. Vehicle Model Recognition Rate.

Models per Manufacturer	True	False	No Match
Alfa Romeo	74	4	2
Audi	70	7	3
Bmw	75	3	2
Citroen	78	2	0
Fiat	69	8	3
Peugeot	77	2	1
Renault	74	3	3
Seat	73	5	2
Toyota	74	4	2
Volkswagen	75	3	2
Total	**739**	**41**	**20**
Average (%)	**92**	**5**	**3**

From the Tables 1 and 2 we deduce that the Total Recognition Rate = Vehicle Manufacturer Recognition Rate x Vehicle Model Recognition Rate = 94% x 92% ∼ 87 % which is adequate for practical applications.

The speed of the recognition is shown for comparison with standard SIFT, M-SIFT, V-SIFT and M-SIFT+V-SIFT in parallel, in Table 3.

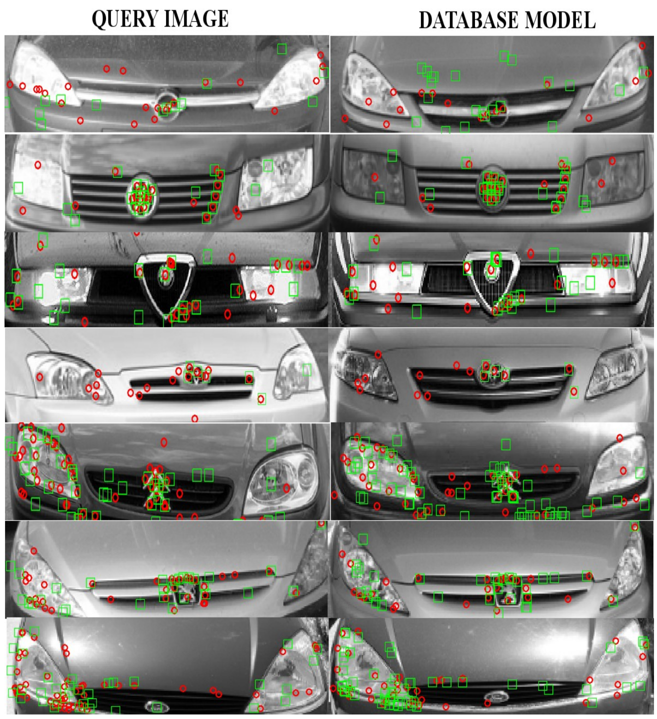

Fig. 6. Query image and database model matching examples. Red circles: keypoints matched, green squares: geometrical keypoint correspondences.

Table 3. Vehicle Logo / Model Recognition Speed.

Methodology	Total Recognition Time (ms)
SIFT	850
M-SIFT	1020
V-SIFT	630
M-SIFT + V-SIFT	913

From Table 3 we see that application of M-SIFT increase the total recognition time, since it increases the keypoint database size and consequently the search time for matching. The V-SIFT is about 26% faster than SIFT and the combination of both M-SIFT and V-SIFT is nearly the same as SIFT.

In order to compare the efficiency of these SIFT-based methodologies we use the independent variables, '*Recall*' and '*1-Rrecision*' defined by eq. (1), (2), as introduced by [25].

$$\text{Recall}=\frac{\text{correct matches}}{\text{total correspondences}} \quad (1)$$

$$1-\text{Precision}=\frac{\text{false matches}}{\text{total matches}} \quad (2)$$

Total correspondences are the real (ground truth) matches which are usually more than the total matches (true + false) found. Ground truth matches were calculated using an estimated homography between query and database images. Keypoints from one image were mapped to the other image and real correspondences occur when the respective keypoints are close to each other in space and scale. Using 1-Rrecision versus Recall, the total recognition performance was evaluated, for various query images and varying the threshold value for NN matching.

The results for standard SIFT features, V-SIFT, M-SIFT and combined V-SIFT+M-SIFT, are shown in Fig. 7.

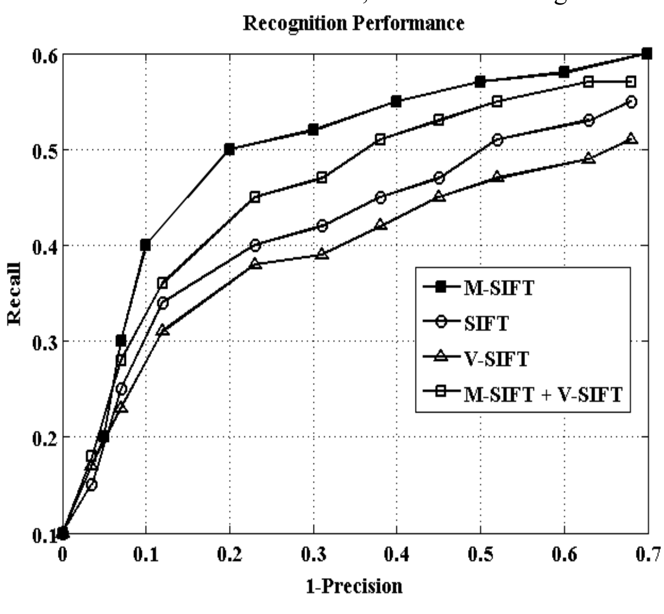

Fig. 7. Recognition performance for SIFT, M-SIFT, V-SIFT and M-SIFT+V-SIFT, respectively.

5. CONCLUSIONS

The vehicle recognition system performed well when applied to a set of vehicle image databases, across with a license plate recognition and congruency calculation modules assisted for segmentation. The vehicle mask and the logo section were successfully segmented and there was a very good rate for vehicle manufacturer identification, as well as for vehicle model recognition, resulting in a 87% total recognition performance. The recognition speed is rather fast (less than a second) which is suitable for a real-time application.

To further boost performance and robustness, we need to extend the system to deal with a wider range of viewpoints or 3-D recognition, as well as recognition on many-vehicle complex scenes and under a greater variety of illumination conditions.

REFERENCES

[1] S. Siggelkow, “*Feature Histograms for Content-Based Image Retrieval”*, PhD Thesis, Albert-Ludwigs-University Frieiburg , December 2002.

[2] H. Schulz-Mirbach, “Invariant Features for Gray Scale Images”, *DAGM Symposium*, I, pp. 1-14, 1995.

[3] D. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 2, nº 60, pp. 91–110, 2004.

[4] D. Lowe, “Object recognition from local scale-invariant features”, *International Conference on Computer Vision*, Corfu, Greece , pp. 1150-1157, September 1999.

[5] D. Lowe, “Local feature view clustering for 3D object recognition”, *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, vol. 1, pp. 682-688, 2001.

[6] M. Brown and D. Lowe, “Invariant features from interest point groups”, *British Machine Vision Conference (BMVC)*, Cardiff, Wales, pp. 656-665, September 2002.

[7] K. Mikolajczyk K. and C. Schmid, “An Affine Invariant Interest Point Detector”, *European Conference on Computer Vision*, vol. 1, pp. 128 - 142, 2002.

[8] M. Weber, M. Welling and P. Perona, “Unsupervised Learning of Models for Recognition”, *Lecture Notes in Computer Science 1842*, Springer-Verlag, pp. 18–32, 2000.

[9] T. Kato, Y. Ninomiya and I. Masaki, “Preceding vehicle recognition based on learning from sample images”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, nº 4, pp. 252–260, 2002.

[10] A. H. S. Lai, N. H. C. Yung, “Vehicle-type identification through automated virtual loop assignment and block-based direction-biased motion estimation”, *IEEE Trans. Intelligent Transportation Systems*, vol. 1, nº 2, pp. 86-97, 2000.

[11] A. H. S. Lai, G. S. K. Fung and N. H. C. Yung, "Vehicle Type Classification from Visual-Based Dimension Estimation", *IEEE Intelligent Transportation Systems Conference*, Oakland (CA), USA, pp. 201-206, 2001.

[12] L. Dlagnekov, S. Belongie, "*Recognizing Cars*", University of California San Diego, Tech. Rep. CS2005-0833, 2005.

[13] V. S. Petrovic and T. F. Cootes, "Analysis of Features for Rigid Structure Vehicle Type Recognition", *British Machine Vision Conference*, vol. 2, pp. 587-596, 2004.

[14] M. Merler, "*Car Color and Logo* Recognition", CSE 190A Projects in Vision and Learning, University of California, 2006.

[15] M. Čonos, "*Recognition of vehicle make from a frontal view*", Diploma Thesis, Faculty of Electrical Engineering, Czech Technical University, 2006.

[16] C. N. Anagnostopoulos, "*Artificial Vision and Computational Intelligence techniques for industrial applications and quality control*", PhD Thesis, Electrical and Computer Engineering Dpt., National Technical University of Athens, 2002.

[17] C. N. Anagnostopoulos, I. Anagnostopoulos, V. Loumos and E. Kayafas, "A license plate recognition algorithm for intelligent transportation system applications", *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, nº 3, pp. 377-392, 2006.

[18] P. D. Kovesi, "Image Features From Phase Congruency", *Videre: A Journal of Computer Vision Research*, MIT Press, vol. 1, nº 3, pp. 1-27, 1999.

[19] A. Psyllos, C. N. Anagnostopoulos, V. Loumos and E. Kayafas, "Image Processing & Artificial Neural Networks for Vehicle Make and Model Recognition", *10th International Conference on Applications of Advanced Technologies in Transportation*, Athens, Greece, May 2008.

[20] A. Psyllos, C. N. Anagnostopoulos and E. Kayafas, "Vehicle Authentication from Digital Image Measurements", *16th IMEKO TC4 Symposium & 13th Workshop on ADC Modeling and Testing* Florence, Italy, September 2008.

[21] T. Lindeberg, "Scale-space theory : A basic tool for analysing structures at different scales", Journal of Applied Statistics, 21(2), pp. 224-270, 1994.

[22] D. Ballard, "Generalizing the Hough transform to detect arbitrary patterns", *Pattern Recognition*, vol. 3, nº 2, pp. 111-122, 1981.

[23] M. Fischler and R. Bolles, "Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *ACM Communications*, vol. 24, pp. 381–395, 1981.

[24] Images Database: http://www.medialab.ntua.gr/ /research/LPRdatabase.html last accessed : 04/2009

[25] Y. Ke, R., Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors" , Conference on Computer Vision and Pattern Recognition, vol. 2 pp. 506-513, 2004.