X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará

Performance Evaluation of Different Machine Learning Techniques With Stereo Vision Used to Road Detection Task

Patrick Y. Shinzato and and Caio Mendes and Fernando Osorio and Denis F. Wolf
University of Sao Paulo - USP
Institute of Mathematics and Computer Science - ICMC
Mobile Robotics Lab - LRM
{shinzato,caiom,fosorio,denis}@icmc.usp.br

Resumo – Reconhecimento de estrada usando informação visual é uma habilidade importante para a navegação autônoma em ambientes urbanos. Devido à isso, um grande número de abordagens de reconhecimento visual de estrada foram propostos na literatura nas útimas três décadas. Este artigo propõe um método de reconhecimento de estrada utilizando informação de profundidade, obtida por uma câmera estéreo, junto com outros atributos baseados em cor. Vários atributos foram avaliados com ajuda de métodos de seleção e os subconjuntos resultantes foram uilizados para avaliarmos diversas técnicas de aprendizado de máquina na tarefa de classificação. Nós utilizamos média, entropia e energia de RGB, HSV, YCbCr e RGB normalizado além da média da disparidade como atributos de entrada dos métodos de aprendizado de máquina. Testes experimentais foram executados em várias situações a fim de validar a abordagem proposta.

Palavras-chave – Visão Computacional, Seleção de Atributos, Aprendizado de Máquina, Naive Bayes, Árvore de Decisão, Redes Neurais Artificiais

Abstract – Road recognition using visual information is an important capability for autonomous navigation in urban environments. Over the last three decades, a large number of visual road recognition approaches have been appeared in the literature. This paper proposes the use of depth information obtained from stereo camera along with other features based on color to detect road. Several features were evaluated using selection methods and its derived subsets were used to test some machine learning techniques in classification task. We used averages, entropy and energy from RGB, HSV, YCbCr and normalized RGB and mean of disparity as input features. Experimental tests have been performed in several situations in order to validate the proposed approach.

Keywords – Computer Vision, Attribute Selection, Machine Learning, Naive Bayes, Decision Tree, Artificial Neural Network

1 INTRODUCTION

Visual road recognition is one of the desirable skills to improve autonomous vehicles systems. As a result, visual road recognition systems have been developed by many research groups since the early 1980s, such as [1, 2]. Details about these and others works can be found in several surveys [3–5].

In this work, we present an analysis of the use of some machine learning(ML) techniques to road detection based on stereo images that contains depth information. Beyond the depth information, the ML techniques received several different image-features as input. Features like averages, entropy and energy from different color channels (RGB, HSV, YCbCr and normalized RGB) were used. Also, our approach transforms an image into a set of objects, where each object represents a part of image. This way, to detect road within an image, you must first classify all objects from the image and combine it based on location of object into image. We tested three types of ML techniques: Naive Bayes(NB), Decision Tree(DT) and Artificial Neural Network(ANN). To compare these methods, we used the value Area Under ROC Curve(AUC). The results showed that using the depth information, along with other features based on color, improves the performance of the classifier. Moreover, the experiments showed that ANN are better than the other techniques evaluated fo this purpose. Based on these results, we can develop an system more robust than others from literature. Also, we do not need make assumptions about road location, orientation of vehicle and the road, and finally traffic situations.

2 RELATED WORK

Most of the work developed before the last decade was based on certain assumptions about specific features of the road, such as lane markings [6] [7], geometric models [8] and road boundaries [9]. These systems have limitations and in most cases they showed satisfactory results only in autonomous driving on paved, structured and well-maintained roads. Furthermore they required favorable conditions of weather and traffic. Autonomous driving on unpaved or unstructured roads, and adverse conditions have also been well-studied in the last decade [10] [11]. We can highlight developed systems for the DARPA Grand Challenge [12] like [13] focusing on desert roads.

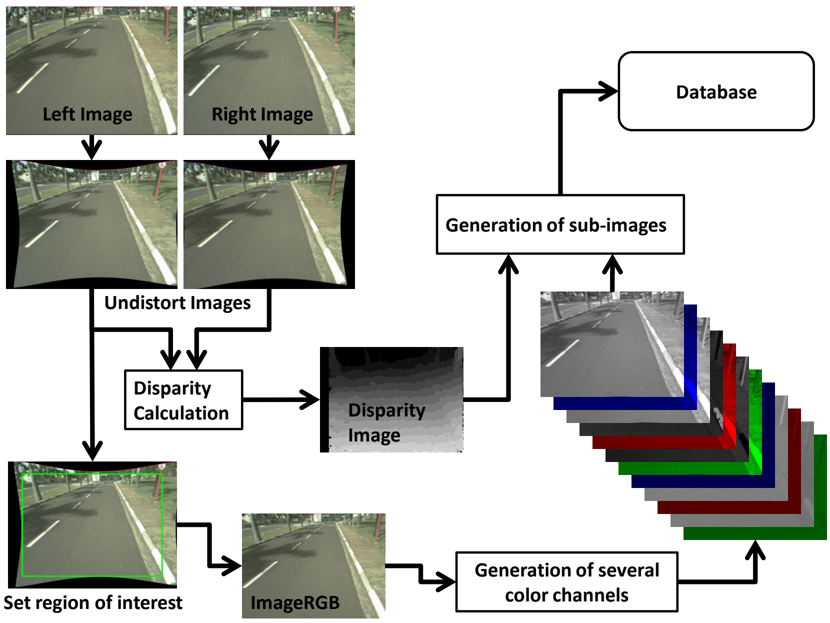


Figura 1: Flowchart of objects generation for each pair of images: Given a pair of images captured by a stereo camera, these images are undistorted and after the disparity is calculated. Also, RGB-image corresponding to the disparity-image is transformed in several color channels. After that, disparity-image and these color channels are used to generate a set of sub-images that will be classified by ML techniques. Its important to note that each sub-image is one object from database.

One of the most representative works in this area is the NAVLAB project [14]. Systems known as SCARF , UNSCARF , YARF , ALVINN [15], MANIAC [16] and RALPH [17] were also developed by the same research group. Among these systems, the most relevant reference for this paper are ALVINN and MANIAC because they are also based on artificial neural networks (ANN) for road recognition. The idea of ALVINN consists of monitoring a human driver in order to learn the steering of wheels while driving on roads on varying conditions. This system, after several upgrades, was able to travel on single-lane paved and unpaved roads and multi-lane lined and unlined roads at speeds of up to 55 mph. However, it is important to emphasize that this system was designed and tested to drive on well-maintained roads like highways under favorable traffic conditions. Beyond those limitations, the learning step takes a few minutes [18] and the authors mention that when is necessary a retraining then this is a shortcoming [17]. According to [16], the major problem of ALVINN is the lack of ability to learn features which would allow the system to drive on road types other than that on which it was trained. In order to improve the autonomous control, MANIAC (Multiple ALVINN Networks In Autonomous Control) [16] has been developed. In this system, several ALVINN networks must be trained separately on their respective roads types that are expected to be encountered during driving. Then the MANIAC system must be trained using stored exemplars from the ALVINN training runs. If a new network is added to the system, MANIAC must be retrained. Both systems trained properly, ALVINN and MANIAC, can handle non-homogeneous roads in various lighting conditions. However, this approach only works on straight or slightly curved roads [10].

Other group that developed visual road recognition based on ANN was the Intelligent Systems Division of the National Institute of Standards and Technology [19] [20]. They developed a system that make use of a dynamically trained ANN to distinguish between areas of road and nonroad. This approach is capable of dealing with non-homogeneous road appearance if the non-homogeneity is accurately represented in the training data. In order to generate training data, three regions from image were labeled as road and three others regions as non-road, i.e., the authors made assumptions about the location of the road in the image, which causes problems in certain traffic situations. Additionally, this system works with the RGB color channel that suffers from the influence in the presence of shadows and lighting changes in the environment. A later work [21] proposed dynamic location of regions labeled as road in order to avoid these problems. However, under shadows situations, the new system becomes less accurate than the previous one because the dynamic location does not incorporate the road with shadow information in the training database.

3 GENERATION OF OBJECTS

The overall method to generate the objects that will be classified consists of three major process. These are: (3.1) Disparity calculation, (3.2) Generation of several color channels from RGB-image, and (3.3) Generation of sub-images, where each sub-image is some object from database. Fig. 1 shows the flowchart of objects generation.

3.1 DISPARITY CALCULATION

A stereo camera has two lenses to capture a pair of images, as show Fig. 2(a). This camera tries to be similar to the functioning of human vision. These pair of images have a shift between parts of the image proportional to the distance of the lens. Due to

(a) Stereo Camera (b) Disparity Image (c) Canonical system

Figura 2: (a)Stereo Camera; (b) Image of disparity calculation of an image; (c) Canonical system of a perfectly undistorted and aligned stereo camera, where **f** is focal length, **B** is the horizontal distance between the lens, and **Z** is depth.

of this, it is possible to determine the depth of a point calculating the difference of its position within the two images. This difference is called of disparity and the result can be see in Fig. 2(b). This disparity-image is used together with various color channels to create the objects from the database.

Disparity of a point **p**, in other words, is the distance on the X-axis with corresponding point **p'** in another image. Fig. 2(c) shows canonical system of a perfectly undistorted and aligned camera with two lenses. Matching algorithms are used to calculate disparity and it has a high computational cost, therefore one should minimize the search space. In a canonical perfect system, such a search could be limited horizontally, i.e., the search only happens in the neighbors of the same line. However, this does not happen in real situations because of the camera lenses have distortions and are not perfectly aligned. Due to this, it is necessary to calibrate camera to rectify and undistorted the images in order to limit the search space. The method used to calibrate our camera and calculate the disparity is described in [22].

3.2 GENERATION OF SEVERAL COLOR CHANNELS

After transforming captured RGB-image into an undistorted RGB-image, the corresponding region to disparity image is selected in this undistorted RGB-image. This image is converted into others spaces color like HSV, YCbCr, and normalized RGB. These color channels are used together with disparity-image to create the objects from the database.

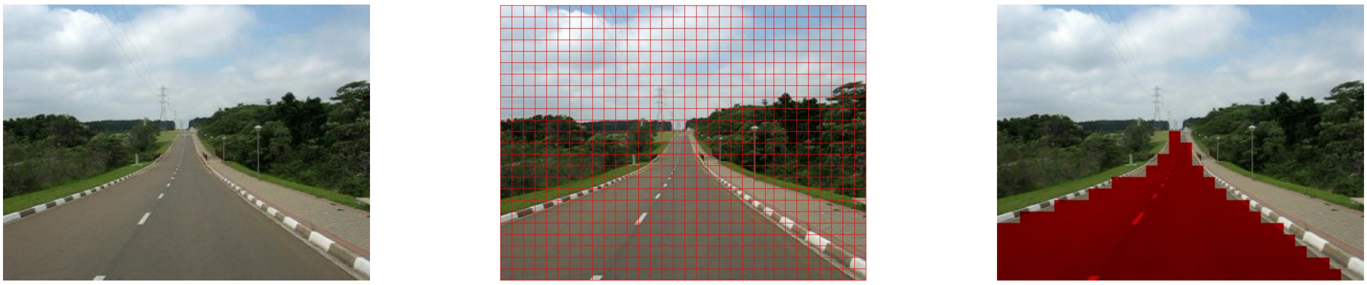
A color space is mathematical representation that describes colors using tuples of numbers. RGB color space is a space where each color can be defined by the quantities variation of R (*red*), G (*green*) and B (*blue*) components. HSV color space is a space that contains H (*hue*), S (*saturation*) and V (*value*) (or brightness). YCbCr color space is composed by Y (*luminance*), Cb (*blue-difference*) and Cr (*red-difference*). Finally, normalized RGB is calculated as follows:

$RN = 255 * \frac{R}{R+G+B}$, $GN = 255 * \frac{G}{R+G+B}$, $BN = 255 * \frac{B}{R+G+B}$.

3.3 GENERATION OF SUB-IMAGES

In this process, all color channels and disparity-image is transformed into a set of objects, where each object represents a sub-image. More specifically, an image with resolution $(M \times N)$ pixels is decomposed in many sub-images with $(K \times K)$ pixels, as shows Fig. 3(a) which is transformed in Fig. 3(b). Mathematically, it can be defined as follows: suppose an image represented by a matrix I of size $(M \times N)$. The element $I(m, n)$ corresponds to the pixel in row m and column n of image, where $(0 \leq m < M)$ and $(0 \leq n < N)$. Therefore, sub-image(i, j) is represented by group $G(i, j)$ that contains all the pixels $I(m, n)$ such that ($(i * K) \leq m < ((i * K) + K)$) and ($(j * K) \leq n < ((j * K) + K)$). If the sub-image is classified as navigable class, then all pixels from group are considered as belonging to this class, Fig.3(c) shows sub-images belonging to road class painted red. This strategy has been used to reduce the amount of data, allowing faster processing and obtaining information like texture from sub-images.

For each sub-image, several features are generated. These features are statistical measures like mean, entropy and energy of pixels belonging to sub-image. Each measure is calculated for all color channels created in the previous process. Also, for each sub-image is calculated only the mean of disparity from its pixels. Thus, we generated a group of 37 features to be used as inputs by several ML techniques that determine whether the sub-image belongs to "road class" or not. Is important to note that all features were normalized because these features have different scales. For training and evaluation purposes, we manually generated sub-image's labels by selecting parts of image to define sub-images that should be labeled as road and non-road, as shows Fig. 4. This Figure shows pixels with classification road painted blue and pixels with classification non-road painted red.



(a) (b) (c)

Figura 3: Generation of sub-images process, the image (a) is transformed into set of sub-images that represents each square from the image (b). Each sub-image is represented by one object into database used by ML techniques. After the classification, we can obtained results like (c), where all pixels from a square receive the same classification. Red squares were classified as belonging to road class.

(a) (b) (c) (d) (e) (f)

Figura 4: Samples of manual classification to generate the learning database. Images (a)(c)(e) are original images, images (b)(d)(f) are images manually labeled. The pixels painted of red represent pixels that classifier must returns non-road class. Pixels painted of blue represents pixels that classifier must return road class.

4 MACHINE LEARNING TECHNIQUES

In this work, we evaluated three types of ML techniques: a probabilistic method called Naive Bayes [23], a decision tree technique C4.5 [24] and an artificial neural network(ANN) from multilayer perceptron (MLP) [25] type. Also, we use a selection method CFS [26] in order to reduce number of input features used in classification.

According to [23], Naive Bayes classifier provides a simple approach to learning probabilistic knowledge. This method was designed for use in supervised learning problems, in which the goal is predict the class of test objects based on training on training set that includes class information. The decision tree technique [24] uses the divide-and-conquist in order to solve a decision problem. C4.5 belongs to a succession of learners that trace their origins back to the work of Hunt and others. The ANN was used with back propagation technique [27], which estimates the weights based on the amount of error in the output compared to the expected results. The ANN topology consists on two layers, where the hidden layer has ten neurons and the output layer has two neurons. The input layer depends on features chosen for input.

Seeking for a more adequate comparison method to the proposed problem, we compared the ML techniques using AUC (area under an ROC curve) obtained from several images from evaluation set. According the Fawcett [28], a receiver operating characteristics (ROC) graph is a technique for visualizing, organizing and selecting classifiers based on their performance. ROC graps are two-dimensional graphs in which true positive rate is plotted on the Y-axis and false positive rate is plotted on the X-axis, as show Fig. 5. Each point in "ROC curve" is produced by different thresholds. To evaluate a classifier, the area under the ROC curve is calculated. This value will always be between 0 and 1.0. Is important to note that AUC values close to or below 0.5 indicate classifiers with poor performance. The closer to 1.0 the better the performance of the classifier.

5 EXPERIMENTS AND RESULTS

In order to validate the proposed system, several experiments have been performed. Several paths traversed by the vehicle CaRINA ("Car Robotic Intelligent of Navigation Autonomous") have been recorded using stereo camera (Videre STOC Color 15cm). These paths are composed by road, sidewalks, parking, buildings, and vegetation. Also, some stretches presents adverse conditions such as dirt (Fig. 6). The vehicle and camera were used only for data collection. The image captured by the camera has a resolution (640×480) pixels. After the rectification and selection of region of interest, the image resolution is reduced to (530×390) pixels. The OpenCV library [22] has been used in the image acquisition and to visualize the processed results from system. We use the *semi-global block matching* method to calculate the disparity of pixels. The sub-image size used was $K = 10$, so each pair of image has over 2000 sub-images.

We selected 128 images from these data and randomly selected $\frac{3}{4}$ of images to use in training step and the $\frac{1}{4}$ remaining for

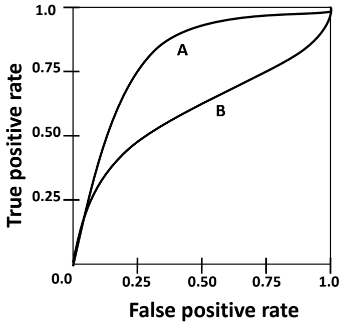


Figura 5: ROC Curve sample for 2 classifiers. In this example, **A** is better than **B**.

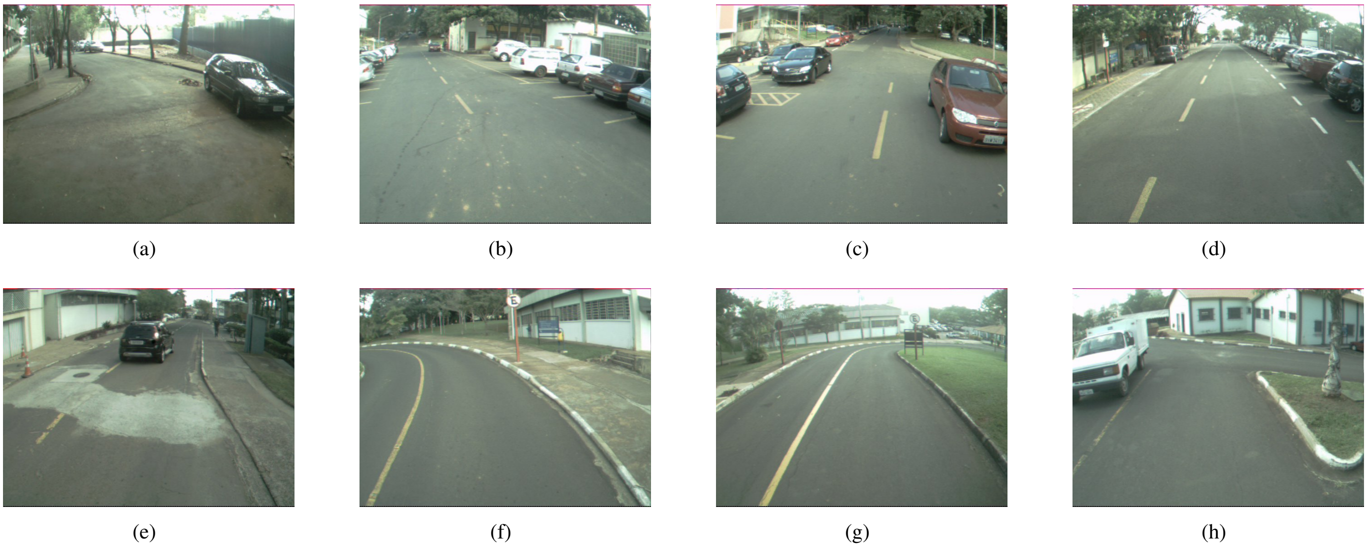


Figura 6: Samples of scenarios used in this work.

evaluation step. In order to evaluate several ML techniques, we extracted all features from selected images to compose a database formatted in ARFF files. This file type can be used by WEKA software [29] that was used to select features and evaluate ML techniques.

5.1 FEATURES SELECTION

To evaluate the contribution of the feature disparity in the classification task, we executed the method on two different databases. The difference between these databases is that one of them contains disparity and the other do not. We used the CFS evaluator with search method BestFirst in cross validation mode with several different seeds.

Tabela 1: Features selected by CFS depending of database source.

	features selected	number of features
database with disparity (Set1)	Mean of S and GN Entropy of B, Cr, Cb, RN and GN Energy of B Disparity	9
database without disparity (Set2)	Mean of S, Cr and GN Entropy of B, Cr, Cb, RN and GN Energy of B, Cb, RN and GN	12

Table 1 shows only features that were chosen by the selection method in all executions. The set of features selected from the database with disparity was called "Set1". The set of features selected from the database without disparity was called "Set2". As expected, the disparity was chosen by selection algorithm on "Set1" database. Also, we can see that the features from set "Set1", except by disparity feature, also appear into "Set2". From this selection, we created two variations of sets. We created the "Set1m" removing disparity feature from "Set1" and we created the "Set2m" adding disparity feature into "Set2". We decided use these four sets to evaluate ML techniques.

6 MACHINE LEARNING TECHNIQUE EVALUATION

The three ML techniques were trained using cross validation method on the training data for each one of the variations of input features. Thus, each ML technique has one evaluation for each input set. For evaluation purposes, we created one database for each image belongs to evaluation set. Each database contains all objects from corresponding image. How exist different types of roads, straight road and curves, different widths and still various possible positions of vehicle in relation to road, these databases for each image are not balanced. Because of this, we decided use AUC value to compare ML techniques.

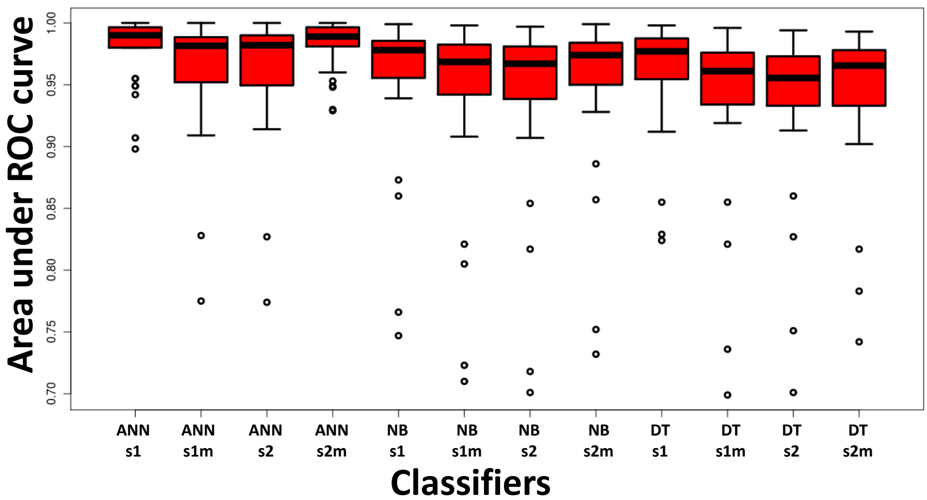


Figura 7: Evaluation in boxplot format of ANN, NaiveBayes and Decision Tree for all input set evaluated.

Since the evaluation set has 32 images, each technique generated 32 AUC values. Fig. 7 shows each evaluation in boxplot format, where each column represents boxplot generated by 32 AUC values of some technique with some input set. Overall, all ML techniques reached good performances. However, in a more careful analysis, we observed that the classifiers NB and DT had more occurrences with low AUC than ANN, suggesting that the ANN achieved a better generalization. Also, checking the first quartile, we see the performance achieved in 75% of the images. In this criterion, we can see that the networks trained on the database with disparity - "s1" and "s2m" - achieved better performances than other techniques with same set of input. Because of this, we consider the ANN are better than others. The Fig. 8 shows some results.

Tabela 2: ANN

ANN	training w/o disparity	training w/ disparity
selecting w/o disparity	0.950	0.981
selecting w/ disparity	0.955	0.980

In order to quantify how the disparity contributes to the classification, we compare the first quartile of the four ANN. This values are show in Table 2. We can see that the ANN trained on the database with disparity achieved better performances than when trained without disparity. A similar behavior can be seen in the graphic from Fig. 7 with the other techniques. Comparing the minimum value and outliers from "ANN_s1" and "ANN_s2m", we can see that the input set called "Set1" is the best choice to use in this task of road detection with stereo camera.

7 CONCLUSION

Visual road recognition is one of the desirable skills to improve autonomous vehicles systems. We presented an analysis about the use of three types of ML techniques: Naive Bayes, Decision Tree and Artificial Neural Network to detect road based on stereo images that contains depth information. Beyond the depth information, the ML techniques received several different image-features as input. Our approach is capable to learn colors, textures and disparity of any sub-image instead of totally road appearance. For compare these methods, we used the value area under ROC curve.

In general, the results showed that using the depth information, along with other features based on color, the performance of the classifier is improved. In addition, the experiments showed that ANN obtained better results than the other techniques evaluated. Based on these results, we can develop a system more robust than others from literature, where we do not need to make assumptions about road location, orientation of vehicle and the road, and finally traffic situations. As future work, we plan to integrate it with other visual systems like *lane detection* in order to improve the system in urban scenarios. We intend to integrate our road detection system with some control algorithm like a adaptation of VFH and control the vehicle. We also plan to integrate our approach with laser mapping in order to make conditions to retrain the ANN without human intervention and

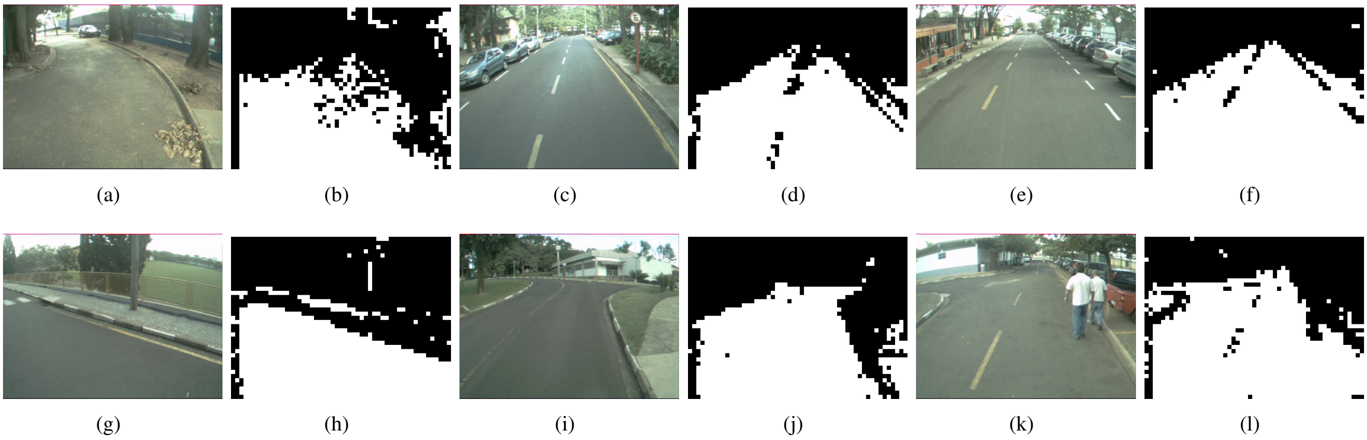


Figura 8: Images shows of results from ANN.

without making assumptions about the image. Finally, as the system classifies each block independently, we intend to improve the processing efficience using a GPU.

REFERENCES

[1] A. Broggi, M. Bertozzi and A. Fascioli. "ARGO and the MilleMiglia in Automatico Tour". *Intelligent Systems and their Applications, IEEE*, vol. 14, no. 1, pp. 55 –64, January 1999.

[2] R. Gregor, M. Lutzeler, M. Pellkofer, K. Siedersberger and E. Dickmanns. "EMS-Vision: a perceptual system for autonomous vehicles". In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, pp. 52 –57, 2000.

[3] F. Bonin-Font, A. Ortiz and G. Oliver. "Visual Navigation for Mobile Robots: A Survey". *Journal of Intelligent & Robotic Systems*, vol. 53, pp. 263–296, 2008. 10.1007/s10846-008-9235-4.

[4] G. Desouza and A. Kak. "Vision for mobile robot navigation: a survey". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 2, pp. 237 –267, February 2002.

[5] M. Bertozzi, A. Broggi and A. Fascioli. "Vision-based intelligent vehicles: State of the art and perspectives". *Robotics and Autonomous Systems*, vol. 32, no. 1, pp. 1–16, 2000.

[6] A. Takahashi, Y. Ninomiya, M. Ohta and K. Tange. "A robust lane detection using real-time voting processor". In *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, pp. 577 –580, 1999.

[7] M. Bertozzi and A. Broggi. "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection". *Image Processing, IEEE Transactions on*, vol. 7, no. 1, pp. 62 –81, January 1998.

[8] J. Crisman and C. Thorpe. "UNSCARF, A Color Vision System for the Detection of Unstructured Roads". In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 3, pp. 2496 – 2501, April 1991.

[9] A. Broggi and S. Bertè. "Vision-Based Road Detection in Automotive Systems: A Real-Time Expectation-Driven Approach". *Journal of Artificial Intelligence Research*, vol. 3, pp. 325–348, 1995.

[10] C. Tan, T. Hong, T. Chang and M. Shneier. "Color model-based real-time learning for road following". In *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pp. 939 –944, September 2006.

[11] F. Diego, J. Alvarez, J. Serrat and A. Lopez. "Vision-based road detection via on-line video registration". In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pp. 1135 –1140, September 2010.

[12] DARPA05. `http://www.darpa.mil/grandchallenge05/`, 2005. visited in February/2011.

[13] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun and G. Bradski. "Self-supervised Monocular Road Detection in Desert Terrain". In *Proceedings of Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[14] C. Thorpe, M. Hebert, T. Kanade and S. Shafer. "Vision and navigation for the Carnegie-Mellon Navlab". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, no. 3, pp. 362 –373, May 1988.

[15] D. A. Pomerleau. "Neural Network Vision for Robot Driving". In *The Handbook of Brain Theory and Neural Networks*, pp. 161–181. University Press, 1996.

[16] T. Jochem, D. Pomerleau and C. Thorpe. "MANIAC: A Next Generation Neurally Based Autonomous Road Follower". In *Proceedings of the International Conference on Intelligent Autonomous Systems*, February 1993.

[17] D. Pomerleau and T. Jochem. "Rapidly adapting machine vision for automated vehicle steering". *IEEE Expert*, vol. 11, no. 2, pp. 19 –27, April 1996.

[18] D. Pomerleau. "Efficient Training of Artificial Neural Networks for Autonomous Navigation". *Neural Computation*, vol. 3, no. 1, pp. 88–97, 1991.

[19] M. Foedisch and A. Takeuchi. "Adaptive real-time road detection using neural networks". In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pp. 167 – 172, October 2004.

[20] P. Conrad and M. Foedisch. "Performance evaluation of color based road detection using neural nets and support vector machines". In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd*, pp. 157 – 160, 2003.

[21] M. Foedisch and A. Takeuchi. "Adaptive road detection through continuous environment learning". In *Applied Imagery Pattern Recognition Workshop, 2004. Proceedings. 33rd*, pp. 16 – 21, October 2004.

[22] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA, 2008.

[23] G. H. John and P. Langley. "Estimating Continuous Distributions in Bayesian Classifiers". In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338–345, San Mateo, 1995. Morgan Kaufmann.

[24] R. Kohavi and R. Quinlan. "Decision Tree Discovery". In *IN HANDBOOK OF DATA MINING AND KNOWLEDGE DISCOVERY*, pp. 267–276. University Press, 1999.

[25] P. S. Churchland and T. J. Sejnowski. *The Computational Brain*. MIT Press, Cambridge, MA, USA, 1994.

[26] M. A. Hall. "Correlation-based Feature Selection for Machine Learning". Technical report, 1999.

[27] D. E. Rumelhart, G. E. Hinton and R. J. Williams. "Learning internal representations by error propagation". pp. 673–695, 1988.

[28] T. Fawcett. "An introduction to ROC analysis". *Pattern Recogn. Lett.*, vol. 27, pp. 861–874, June 2006.

[29] WEKA. `http://www.cs.waikato.ac.nz/ml/weka/`, 2011. visited in June/2011.