ViCoMoR

IROS Workshop on Visual Control of Mobile Robots (ViCoMoR)

Half Day Workshop

September 26th, 2011, San Francisco, California,
in conjunction with the IEEE/RSJ International
Conference on Intelligent Robots and Systems

http://vicomor.unizar.es

Organizers

Youcef Mezouar
LASMEA - Université Blaise Pascal, France

Gonzalo López-Nicolás
I3A - Universidad de Zaragoza, Spain

Contents

Web: http://vicomor.unizar.es
IROS Workshop on Visual Control of Mobile Robots (ViCoMoR)
September 26th, 2011, San Francisco, California, in conjunction with the
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

The organization of this workshop was partly supported by Ministerio de Ciencia e Innovación/Unión Europea, DPI2009-08126, the ANR R-discover project and grant I09200 from Gyeonggi Technology Development Program funded by Gyeonggi Province.

IROS Workshop on Visual Control of Mobile Robots (ViCoMoR)
September 26, 2011, San Francisco, California, USA

Visual Road Recognition Using Artificial Neural Networks and Stereo Vision

Patrick Y. Shinzato and Fernando S. Osorio and Denis F. Wolf
Mobile Robotic Laboratory,
Institute of Mathematics and Computer Science,
University of Sao Paulo - ICMC-USP,
Sao Carlos, SP Brazil
Email: shinzato@icmc.usp.br, fosorio@icmc.usp.br, denis@icmc.usp.br

***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 propose a novel stereo vision based on artificial neural network that can identify the road using color, texture and disparity information from images. Several features are used as inputs of the ANN such as: average, entropy, energy and variance from different color channels (RGB, HSV, YCrCb). As a result, our system is able to estimate the road location and the confidence factor for each part of detected environment. Furthermore, our system presents a good generalization capacity. Experimental tests have been performed in several situations in order to validate the proposed approach.**

I. Introduction

Visual road recognition, also known as “lane detection”, “road detection” and “road following”, 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] [3]. Details about these and others works can be found in several surveys [4] [5] [6] [7].

Most work developed before the last decade was based on certain assumptions about specific features of the road, such as lane markings [8] [9], geometric models [10] and road boundaries [11]. 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 [12] [13] [14] [15]. We can highlight developed systems for the DARPA Grand Challenge [16] like [17] [18] [19] focusing on desert roads.

One of the most representative works in this area is the NAVLAB project [3]. Systems known as SCARF [20], UNSCARF [10], YARF [21], ALVINN [22], MANIAC [23] and RALPH [24] 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 [25] and the authors mention that when is necessary a retraining then this is a shortcoming [24]. According to [23], 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) [23] 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 ALVINN network is added to the system, MANIAC must be retrained. Both systems trained properly, ALVINN and MANIAC, can handle nonhomogeneous roads in various lighting conditions. However, this approach only works on straight or slightly curved roads [12].

Other group that developed visual road recognition based on ANN was the Intelligent Systems Division of the National Institute of Standards and Technology [26] [27]. 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 nonhomogeneous road appearance if the nonhomogeneity 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 nonroad, 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 a lot of influence in the presence of shadows and lighting changes in the environment. A later work [28] 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.

In this work, we present a visual road detection system that use ANN with stereo images that contains depth information. Beyond the depth information, our ANN received several different image features as input. Features like averages, entropy, energy and variance from differents color channels (RGB, HSV, YCrCb) from sub-images. Other detail about this classification is that it provides confidence factor for each sub-image classification that can be used by control algorithm. Unlike [26], our system does not need to be retrained all the time because the generalization capacity of our system is more powerful than theirs. Therefore, our system does not require make assumption about location of road.

II. System Description

The system's goal is to identify the road region on a image obtained by a stereo camera attached to a vehicle. To accomplish this task, our system calculates the disparity of pixels using images left and right from camera. Immediately after, our system transforms the color image into set of sub-images and generates image features for each one. These features and disparity are used by ANN in order to classify if a sub-image belongs to a road class or not. A control algorithm uses the results in order to control the vehicle autonomously. After executing an action, the system captures another pair of images from environment and returns to first stage. The Fig. 1 show how the system works.

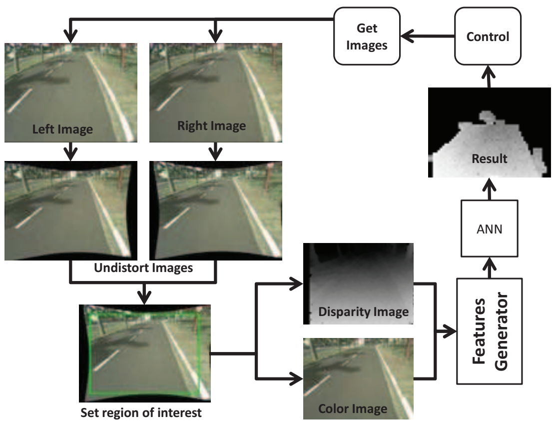


Fig. 1. (a) The System Architecture: Given a pair of images, the disparity is calculated. After that, disparity and color image are transformed into a set of sub-images that will be classified by ANN.

A. Stereo Vision

A stereo camera has two lenses to always capture a pair of images. These images have a shift between parts of the image proportional to the distance of the lens. Due to of this, it is possible to determine the depth of a point, estimating the difference of its position within the two images. This method is similar to the functioning of human vision.

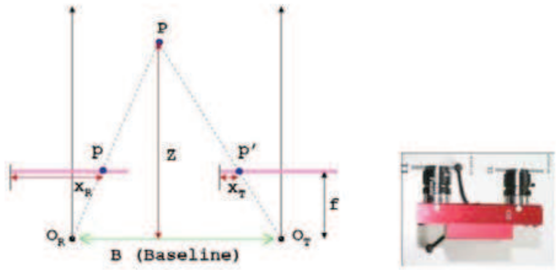


Fig. 2. Canonical system of a camera with two lenses. f is focal length, B is the distance between the lens.

Disparity of a point p is the distance on the X-axis with corresponding point p' in another image. Fig. 2 shows canonical system of a camera with two lenses. Match 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, ie, the search only happens in the neighbors of the same line. However, this does not happen in reality because of the camera lenses have distortions and are not perfectly aligned. Due to this, it is necessary calibrate camera in order to rectify the images and limit the search space. The method used to calibrate our camera and calculate the disparity is described in [29].

B. Generating Of Sub-images With Features

This stage transforms an color image into set a of sub-images and generates image-features for each them. More specifically, a image resolution ($M \times N$) pixels was decomposed in many sub-images with ($K \times K$) pixels, as show 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 collumn 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 ($(j * K) \leq n < ((j * K) + K)$). For each group, many image features are generated. These features will be used as input ANN that determine whether the sub-image belongs to a road class or not. If the sub-image is classified as belonging to road 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.

Several statistical measures like mean, entropy and variance were used as image features. For each image, all measures were calculated with each color channel - we used RGB, HSV, YCrCb and normalized RGB. Also, the average of disparity of all pixels inside each sub-image was calculated. Thus, we generated a group of 49 features to be used as inputs by ANN. However, how this is a large number of inputs, we decided

(a) (b) (c)

Fig. 3. In features generation stage, the image (a) is transformed into set of sub-images that represents each square from the image (b). 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.

to use a selection method “CFS”. This method selected the disparity-feature and all features shown in Table I. Finally, our system uses only 14 features to classify sub-images between road or non-road class.

TABLE I
FEATURES CALCULATED BY OUR SYSTEM. NOTE THAT RN, GN, BN ARE RGB CHANNELS NORMALIZED.

Measure	Channels from several color spaces											
	R	G	B	H	S	V	Y	Cr	Cb	RN	GN	BN
Mean	×	×	×	×	×	×	×	×	×	×		×
Entropy				×								
Variance												
Energy								×				

C. Artificial Neural Network

We used a multilayer perceptron (MLP) [30], which is a feedforward neural network model that maps sets of input data onto specific outputs. We use the back propagation technique [31], which estimates the weights based on the amount of error in the output compared to the expected results.

The ANN topology consists in, basically, two layers, where the hidden layer has seven neurons and the output layer has two neurons, as shows the Fig. 4, one neuron for road class and other for non-road class. The input layer has 14 inputs, and they are all normalized.

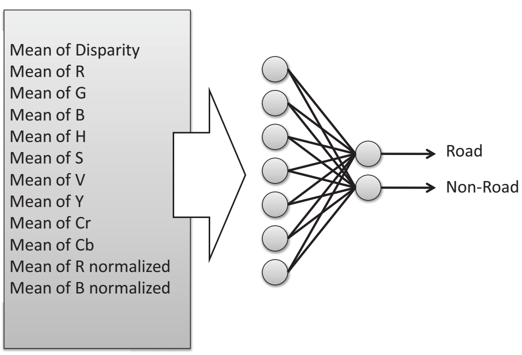


Fig. 4. ANN topology: The ANN uses some features, not all, to classify the sub-image between belonging to a road class or not.

Regarding ANN convergence evaluation, two metrics are frequently used: “MSE” and “Hit Rate”. The MSE is “Mean-Square Error” and usually the training step stops when the “MSE” converges to zero or some acceptable value. However, a small mean-square error does not necessarily imply good generalization [32]. Also this metric does not provide how many patterns are missclassified, i.e., if the error is higher in some patterns or if the error is evenly spread in all patterns. Other way of evaluating the convergence is checking how many patterns were classified correctly, or “Hit Rate”. In this case, the problem is to define a good precision to interpret the ANN output (i.e. given a ANN output, determine whether the output is equal to expected output or not), since the output may not be exactly the value expected. Seeking for a more adequate assessment to the proposed problem, we used a method known as AUC (area under an ROC curve).

According Fawcett [33], 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.

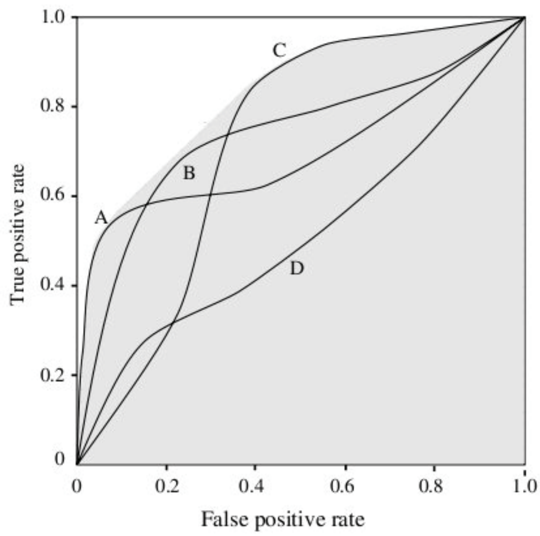


Fig. 5. ROC Curve sample for 4 classifiers. Image adapted from [33]

D. Visual Navigation Map (VNMap)

After classify all sub-images from an image with ANN, our system generate a VNMap filtering the resulting image with a growth algorithm. Thanks to it, the sub-images belonging to non-road class are painted pure black. Fig. 6 shows a sample of an image classified, where the Fig. 6(b) shows the VNMap in gray-scale - black represents non-road class, white represents road class and the gray represents the intermediate values. Is important to note that when the network does not achieve a good classification then the filter can consider the whole region as not navigable.

(a) (b)

Fig. 6. (a) is a color image. (b) is results from a classification sample from our system. Black represents non-road class, white represents road class and the gray represents the intermediate values.

III. Experiments and Results

In order to validate the proposed system, several experiments have been perfomed. Our setup for the experiments was a eletric car equipped with (Videre STOC Color 15cm) camera. The image resolution was (640×480) pixels. The car and camera were used only for data collection. In order to execute the experiments with ANNs, we used the OpenCV library [29] that has also been used in the image acquisition and to visualize the processed results from system. The sub-image size used was $K = 10$, so each image has 3072 sub-images. We use the *semi-global block matching* method to calculate the disparity of pixels. The ANN has been trained until reaching 500 cycles.

Several paths traversed by the vehicle have been recorded using stereo camera. These paths are composed by road, sidewalks, parking, buildings, and vegetation. Also, some stretchs presents adverse conditions such as dirt (Fig. 7). We selected 128 pairs of images to compose our database. We randomly selected $\frac{3}{4}$ of images to use in training step and the last $\frac{1}{4}$ for evaluation step. The method 8-fold cross validation [32] was used in training step. The Table II shows results on evaluation set, ie, the AUC of each instance of ANN. Each line represents some instance of ANN trained with $\frac{7}{8}$ of training set and tested with last $\frac{1}{8}$ represented by each column. The last line represents the average of all instances evaluated. In general, our system achive AUC of 0.96 on evaluation set, which is a sactisfatory result.

The Fig. 8 shows some samples of classifications. It is possible to see that our system (ANN + filter) has achieved a high degree of certainty about the sub-images belonging to road class in all cases. In general, our system has been able to distinguish the road from the sidewalk and other items as cars and trees of the scenes evaluated. The small errors obtained are related to traffic lanes that have very different colors of asphalt and were manually interpreted as belonging to road class in this experiment. In addition, the system accumulates error with the loss of accuracy at the edges, since the method of generating attributes averaging a portions of the image.

(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

Fig. 7. Samples of scenarios used in this work.

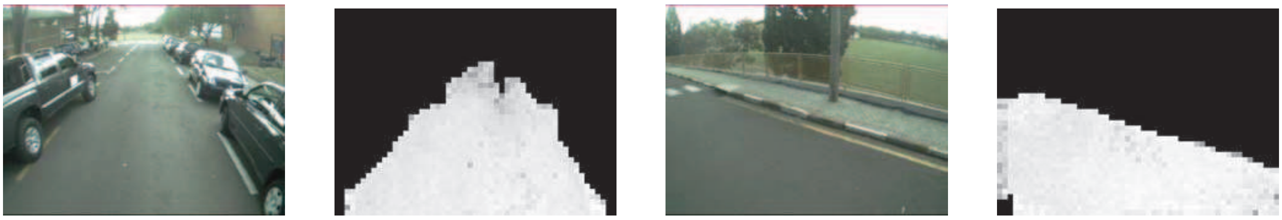
TABLE II
Table of some AUC obtained on the set of evaluation

some instances	T1	T2	T3	T4	T5	T6	T7	T8
1	0.98	0.89	0.96	0.97	0.95	0.98	0.97	0.96
2	0.98	0.89	0.96	0.97	0.96	0.98	0.98	0.96
3	0.98	0.89	0.96	0.97	0.95	0.98	0.98	0.96
4	0.98	0.89	0.97	0.97	0.96	0.98	0.98	0.96
5	0.98	0.89	0.96	0.97	0.96	0.98	0.98	0.97
6	0.98	0.89	0.97	0.98	0.96	0.98	0.98	0.96
7	0.98	0.88	0.96	0.97	0.95	0.98	0.98	0.97
8	0.98	0.89	0.97	0.97	0.95	0.98	0.98	0.97
9	0.98	0.89	0.96	0.97	0.96	0.98	0.98	0.96
10	0.98	0.89	0.97	0.97	0.96	0.98	0.98	0.96
average	0.98	0.89	0.96	0.97	0.96	0.98	0.98	0.96

An important observation is that the disparity helped eliminate the misclassification of near obstacles with similar colors to the road - see Fig.8(e)-(f). Also, we may be concluded that the farther away is the observed point, less significant is the value of disparity. Because the sub-images near the vanishing point of the straight roads were not correctly classified - see Fig.8(c)-(d) and Fig.8(k)-(l).

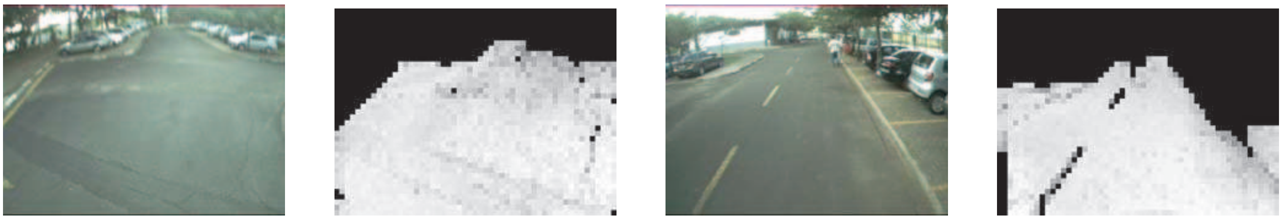
Based on the experiments, we concluded that results are satisfatory and can be used by some control algorithm. Fig.9 shows in green the sector chosen by the algorithm based on direction of goal using a GPS and the classification from scene obtained by stereo camera. The other sectors colored with blue are other options avaiable in accordance with road identifier classification. When the sector is not colored so the sector not achieved the minimum threshold from polar histogram, which means that the vehicle should not go in those directions. This implementation only shows the direction that our vehicle must follow, we do not implement the control of the vehicle yet.

(a) (b) (c) (d)



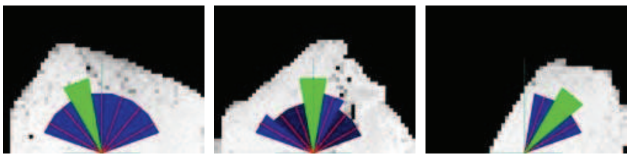
(e) (f) (g) (h)

(i) (j) (k) (l)



(m) (n) (o) (p)

Fig. 8. Samples of results from our system.



(a) (b) (c)

Fig. 9. Results from control algorithm, the image (a) is result when the goal is left of vehicle. (b) when goal is in the front of vehicle. And finally, (c) when the goal is the right of vehicle.

IV. Conclusion

Visual road recognition is one of the desirable skills to improve autonomous vehicles systems. We presented a visual road detection system that use ANN with depth information obtained by a stereo camera. Our ANN is capable to learn colors, textures and disparity of any sub-image instead of totally road appearance. Also, our training evaluation method is a more adequate assessment to the proposed problem, since many results with many low degrees of certainty lead to low scores. Finally, the system classification provides confidence factor for each pixel-group classification of image that can be used by a control algorithm.

In general, the results obtained in the experiments were relevant, since the system reached good classification results when the training step obtains good score. 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 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 without making assumptions about the image. Finally, as the system classifies each block independently, we intend to improve the processing efficience using a GPU.

Acknowledgment

The authors acknowledge the support granted by CNPq and FAPESP to the INCT-SEC (National Institute of Science and Technology - Critical Embedded Systems - Brazil), processes 573963/2008-9 and 08/57870-9.

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*, 2000, pp. 52 –57.

[3] 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.

[4] 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.

[5] 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, Feb. 2002.

[6] 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.

[7] E. D. Dickmanns, "Vehicles capable of dynamic vision: a new breed of technical beings?" *Artif. Intell.*, vol. 103, pp. 49–76, August 1998.

[8] 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*, 1999, pp. 577 –580.

[9] 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, Jan. 1998.

[10] 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*, vol. 3, April 1991, pp. 2496 – 2501.

[11] A. Broggi and S. Bert, "Vision-based road detection in automotive systems: A real-time expectation-driven approach," *Journal of Artificial Intelligence Research*, vol. 3, pp. 325–348, 1995.

[12] 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*, September 2006, pp. 939 –944.

[13] 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*, September 2010, pp. 1135 –1140.

[14] J. Alvarez and A. Lopez, "Novel index for objective evaluation of road detection algorithms," in *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, October 2008, pp. 815 –820.

[15] M. Manz, F. von Hundelshausen, and H.-J. Wuensche, "A hybrid estimation approach for autonomous dirt road following using multiple clothoid segments," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, May 2010, pp. 2410 –2415.

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

[17] 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.

[18] D. Lieb, A. Lookingbill, and S. Thrun, "Adaptive road following using self-supervised learning and reverse optical flow," in *Proceedings of Robotics: Science and Systems*, Cambridge, USA, June 2005.

[19] J. Lee, C. Crane, S. Kim, and J. Kim, "Road following in an unstructured desert environment using monocular color vision as applied to the darpa grand challenge," in *International Conference on Control, Automation, and Systems (ICCAS)*, 2005.

[20] J. Crisman and C. Thorpe, "Scarf: A color vision system that tracks roads and intersections," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 1, pp. 49 – 58, February 1993.

[21] K. Kluge and C. Thorpe, "The yarf system for vision-based road following," *Mathematical and Computer Modelling*, vol. 22, no. 4-7, pp. 213 – 233, 1995.

[22] D. A. Pomerleau, "Neural network vision for robot driving," in *The Handbook of Brain Theory and Neural Networks*. University Press, 1996, pp. 161–181.

[23] 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.

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

[25] D. Pomerleau, "Efficient training of artificial neural networks for autonomous navigation," *Neural Computation*, vol. 3, no. 1, pp. 88–97, 1991.

[26] 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*, October 2004, pp. 167 – 172.

[27] 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*, 2003, pp. 157 – 160.

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

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

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

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

[32] S. Haykin, *Neural Networks: A Comprehensive Foundation (2nd Edition)*, 2nd ed. Prentice Hall, July 1998.

[33] T. Fawcett, "An introduction to roc analysis," *Pattern Recogn. Lett.*, vol. 27, pp. 861–874, June 2006. [Online]. Available: http://portal.acm.org/citation.cfm?id=1159473.1159475