1998 World Congress on Computational Intelligence, vol IJCNN, Anchorage, Alaska, May 4-9, 1998

Better Road Following by Integrating Omni-View Images and Neural Nets

Zhigang Zhu*, Shiqiang Yang, Dingji Shi, Guangyou Xu

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Abstract

In this paper we present recent results of integrating omni-directional view image analysis and a set of adaptive backpropagation networks to understand the outdoor road scene by a mobile robot. Road is classified before orientation estimation so that the system can deal with road images with different types effectively and efficiently. A modified omni-view image sensor captures images with 360 degree view around the robot in real-time. New rotation-invariant image features are extracted by a series of image transformations, and serve as the inputs of a road classification network. Then the classification result (the road category) activates one of the road orientation networks to estimate the road orientation of the input image classified in that category. Experimental results with real scene images and some practical problems for actual applications are discussed.

1. Introduction

An autonomous mobile robot (vehicle) should have three basic functions in order to move safely in an outdoor road environment: road following, obstacle avoidance and landmark recognition. All of them need the comprehensive understanding of the natural road scene. In this paper we deal with the first problem and part of the last problems in an integrated manner. A robot moves along the road and makes decisions when it reaches some pre-defined points. It should obtain two kinds of information from the visual sensors: the robot heading (or the road orientation) and its location (in terms of road categories, e.g., straight road, intersection or T-junction).

Two problems in visual algorithms and systems may prevent them from being used in this real world application successfully. Firstly, most mobile robots can only view objects in front of them due to the narrow view angle of the commonly used TV cameras. As the result, robots may go astray or collide against objects from the side or behind. Secondly, most of the vision algorithms for outdoor road understanding only applied to the predefined environment. Once the environment changes, they may work improperly.

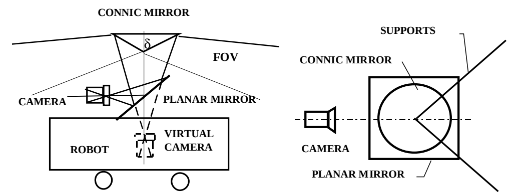
To solve the first problem, several researchers have studied the omni-directional vision system. Elkins and Hall [1] used fish eye lens for the visual navigation of an outdoor mobile robot. The mobile robot located itself by referring to the known targets in the environment. Yagi et al. [2] applied a conic mirror to acquire omnidirectional image for the indoor mobile robot. Vertical edges common in the indoor structured environment were used to detect obstacles while the robot carried out a constant linear motion on flat floor. Hong et al. [3] used a spherical mirror to capture the omnidirectional images and studied the image-based homing problem in the indoor environment. The 1D horizon circle of the omnidirectional image captured at a location was matched with those of a series of predefined "homing" locations and guided the robot to reach the nearest "home". Stein and Medioni[4] used a rotating camera to acquire the 360-degree panoramic images in the terrain. The omnidirectional curves of the horizon were used to find "drop off" location of the robot. The

* The author is currently on a leave in the Computer Science Department, University of Massachusetts at Amherst, MA 01003. Email: zhu@cs.umass.edu, or zhuzhg@mail.tsinghua.edu.cn. This work was supported by China Advanced Research Project during 1990-1996.

intuitive characteristics of the above approaches are that specific features in the omnidirectional images are used to solve the given problems in the predefined environments. Omnidirectional image methods applied to the outdoor road scenes still need investigations.

Artificial neural networks (ANN) are a reasonable solution for the second problem. In a large project, a full-sized self-driving van named ALVINN (Autonomous Land Vehicle In a Neural Net) equipped with video cameras and four onboard workstations has been developed and built at CMU [7]. ALVINN is a fully connected three-layered backpropagation network, whose input is 32x32 sub-sampled road image from a video camera and whose output is the vehicle heading (1 out of 45) required to make the van stay on the road. Nine hidden units are used in the system and the output is updated 15 times per second. Vehicle headings, while steered by the human driver, provide the feedback necessary for training. Although the ALVINN has successfully driven the Navlab vehicle on various types of the road in various weather conditions, the system is still not perfect. The images taken at different viewing directions by a commonly used TV camera are quite different from each other due to the perspective projection and quite different viewing zones, so the network will be too complicated for complex scene. In their recent work, Jochem and Pomerleau [10,11] proposed the so-called virtual camera method to handle the lane transition problem of highway driving. The basic idea is to find the suitable image sub-region and to transform it to an image captured by a "virtual camera" from the desired viewing point, so that the input requirements of the ALVINN are satisfied. Problem may also arise when the vehicle head for directions not included in the training set, or when the road scenes vary very much during the long driving.

The combination of omnidirectional imagery and the neural networks leads to a better solution for afore-mentioned problems[6,9]. The omnidirectional images provide rotation-invariant features to the neural networks, while the neural networks provide an adaptive way for image classification and identification. In this paper we present our current research on the neural-based omnidirectional image understanding. A modified omni-view imaging system was designed for the outdoor application so that the robot would not be troubled by the limited view angle of the camera. New rotation-invariant and rotation-dependence features are extracted by a series of image transformations without any image segmentation and feature extraction, and the separation greatly simplifies the road classification and the heading decision.



(a) Side view (b) Top view

Fig. 1. Geometry of the omni-view image sensor

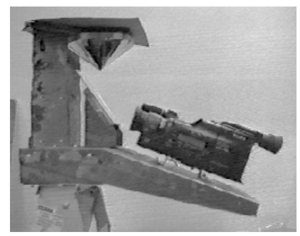


Fig 2. A prototype of the omni-view image (OVI) sensor.

2. Omni-view imaging sensor

In order to obtain in real-time the omni-view of the road and objects around the mobile robot with suitable resolution, we modified the conic projection sensor system COPIS proposed by Yagi et al. [2], aiming to deal with the situation of the outdoor road scene. The geometry and configurations of the omni-view image (OVI) sensor are shown in Fig. 1. A conic mirror (with a vertex angle δ=55°) is fixed on the roof of the robot by two very thin sheet steel supports (the thickness is 1 mm), whose intersection line coincides the vertical axis of the conic mirror. A planar mirror is placed beneath the conic mirror with tilt angle $\pi/4$. The camera is mounted horizontally on the roof of the robot. The optical axis of the camera, the vertical axis of the conic mirror, and the normal of the planar mirror lie in one vertical plane. The distance between the conic mirror and the roof of the robot must be large enough to avoid occlusion of the visual field by the robot itself. The position of the planar mirror and the camera should be carefully adjusted to ensure the coincidence of the optical axis of the "virtual camera" inside the planar mirror and the vertical axis of the conic mirror. We use a planar mirror and a horizontally placed camera instead of a camera pointing upward for sake of easy installation and adjustments, and protection of the camera lens. Moreover, the omni-view image after two mirror reflections gives un-inverse view of the scene, as opposite to the mirror image by the COPIS system [2]. The transparent tube used in the COPIS is replaced by two thin sheet steel supports for the conic mirror because we have found that the commonly available transparent tube is not completely transparent, and the reflection by the tube troubles image analysis.

Fig. 2 shows a prototype of the omni-view image (OVI) sensor used in our experiments. The camera is placed on a titled plane (about 15 degrees) instead of a horizontal plane in order to avoid the occlusion of the field of view by the large-size camcorder. Consequently the tilt angle of the planar mirror increases to about 60 degrees. In our prototype system, the diameter of the conic mirror is 110 mm and the nearest edge of each thin support sheet is 100 mm far away from the conic vertical axis, so each sheet occupies less than 0.6 degrees out of the 360-degree's field of view. The image taken by our OVI sensor represents a 360° view of the scene around the robot, ranging from about 5 meters to 30 meters on the ground. Omnidirectional image taken by a conic mirror is equivalent to the image taken by a tilting line scan camera rotating around a vertical axis, while it optical center is moving along a circle around the same axis. Fig. 3 shows two omni-view images captured when the robot just entered a T-junction and was in the center of it.

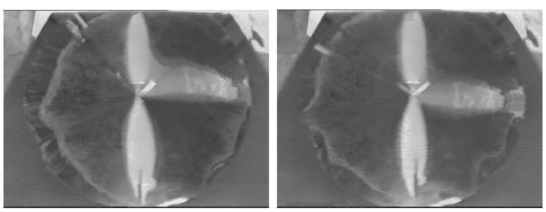


Fig.3. Omni-view images

3. Rotation-invariant features

3.1. Polar transform and projection transform

Suppose the origin of the OVI coordinate system xoy is in the center of the image where the conic vertex is projected, we transform the Cartesian coordinate image I(x,y) into a polar coordinate image I(r,θ) by

$$r = \sqrt{x^2 + y^2}\ ,\quad \theta = \tan^{-1}(y / x) \tag{1}$$

where r is the radius and θ is the orientation angle(0 - 2π). Ideally, for a 256x 256 original sensor image, the resolution of the angle in the polar image is about 1 degree, so the resolution of corresponding polar image is 128x360. The 2D polar image I(r, θ) is transformed to a 1D orientation projection u(θ) by accumulating (projecting) the image along radius r:

$$u(\theta) = \sum_r I(r,\theta),\quad \theta \in [0,2\pi) \tag{2}$$

3.2. Road orientation

In order to train and test the BP networks, we should know the ground truth of the road categories and orientations. The former can be easily provided by the human operators since the image sequence for a certain road category lasts for a long time period, while the later should be estimated by the computer since the road orientations change from frame to frame. We use a simple road finding and tracking algorithm[8] to determine the road orientations if the roads show intuitive peaks in the orientation projection u(θ):

Step 1. Find peaks in the orientation projection u(θ);

Step 2. Examine the geometric relations among the detected peaks.

Step 3. Track the road by a simple local correlation-based matching of the orientation projections.

3.3. Rotation invariant and dependent features

Suppose that the orientation projection u(θ) is sampled and then normalized into **U** = {u(n), n=0,...,N-1} satisfying zero mean and the unity standard deviation. The normalized procedure eliminates or at least reduces the influence of any illumination changes of images that are captured at different times. If rotation angle of the robot is

$$\phi = -\frac{2\pi n_0}{N} \tag{3}$$

where $\phi \in [-2\pi, 0]$, then the new orientation projection v(n) is the circular shift of u(n) by n_0, and can be denoted as

$$v(n) = u(n - n_0) \tag{4}$$

where u(n) is the orientation projection when the robot heads for the front road. The Fourier transforms of u(n) and v(n) can be expressed as

$$\begin{aligned} a(k) &= a_k \exp(j\psi_k) \\ b(k) &= b_k \exp(j\varphi_k) \end{aligned}\ ,\ \text{k=0,...,N-1} \tag{5}$$

where

$$b_k = a_k, k = 0,\ldots, N-1 \tag{6}$$

$$e^{j\psi_k} = e^{j(\varphi_k + k\phi)} \tag{7}$$

Equation (6) says that the Fourier amplitudes are invariant to the rotation of the omni-view images. Therefore they are appropriate features for road scene classifications. Equation (7) give the basic relation to estimate the orientation difference between two omni-view images. For real scene images, the equality can not validate exactly. Analysis and experimental results show that the Fourier phases are very sensitive to the noises, especially to the geometrical distortion of the orientation projections. So the orientation difference should be estimated by searching the minimum value of the following distance

$$d(\phi) = \sum_{k=0}^{N-1} (a_k e^{j\psi_k} - b_k e^{j(\varphi_k + k\phi)})^2 \qquad (8)$$

for each $\phi(n_0) = \phi = -\frac{2\pi n_0}{N}$, n_0=0,1,...,N-1. This is equivalent to find the maximum value of the circular cross-correlation function (CCF)

$$C(n_0) = \sum_{n=0}^{N-1} u(n)v((n+n_0) \text{modulo } N) \qquad (9)$$

which is O(N^2) computation. Since we have obtained the Fourier transform of u(n) and v(n), we hope to use them for orientation estimation. The correlation theorem states that the (circular) correlation of two real signal sequence u(n) and v(n) is equal to the inverse discrete Fourier transform (DFT) of the product of the DFT b(k) and the conjugation of DFT a(k) , i.e.,

$$C(n_0) = F^{-1}(a^*(k)b(k)) \qquad (10)$$

The alternative approach has only O(3N$\log_2$N) time complex. The advantage of the global correlation method is that no feature extraction is needed. However the orientation difference estimated by the method cannot indicate the exact road orientation if the distortions between the two images are not negligible. The CCF method, together with the peak finding and tracking method, is used in the data collection and training for the networks since the behavior of the robot can be controlled in the training stage. In the real operations, the orientation is estimated using the neural networks.

From equations (9) and (10) we can obtained another rotation-invariant sequence, the auto-correlation function (ACF) of u(n)

$$R(n_0) = \sum_{n=0}^{N-1} u(n)v((n+n_0) \text{modulo } N) \qquad (11)$$

The ACF of u(n) is equal to the inverse Fourier transform of the energy spectrum of u(n)

$$R(n_0) = F^{-1}(a^*(k)a(k)) = F^{-1}(a_k^2) \qquad (12)$$

4. Architecture

The basic model for Road Understanding Neural Networks (RUNN) is an adaptive combination of an image processing module (IPM) and several fully connected backpropagation (BP) networks --- a single three-layered Road Classification Network (RCN), a two-layered Road Orientation Network (RON) for each road category. The inputs of the RCN are P ($\le$N) rotation-invariant image data (e.g., Fourier amplitudes or ACF), and the outputs are M road categories. The inputs of each RON are Q ($\le$N) rotation-dependence image data (e.g., the orientation projection), and the outputs are L road orientations. The data representation of the inputs and the number of hidden units for each network are decided by the experiments of training and testing of the network using real images.

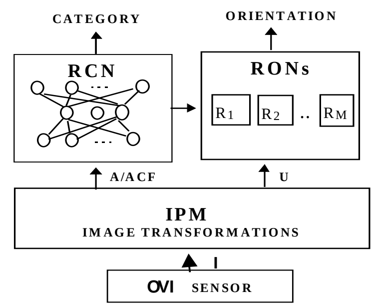


Fig 4. The Architecture of RUNN

The RCN and RONs are set up using *Nworks* tool[7], which provides various of ANN architectures and the flexibility of parameter controls. The IPM is composed of a pipeline image processing machine named PIPE [8], hosted by a PC 486 with the Nworks environment. The polar transformation can be carried out in the PIPE at the video field rate (60 fields per second). The projection transformation is also implemented in the PIPE at the frame rate. The 1D Fourier transform is carried out by the PC486. Fig. 4 shows the system architecture. The RUNN works in the following steps:

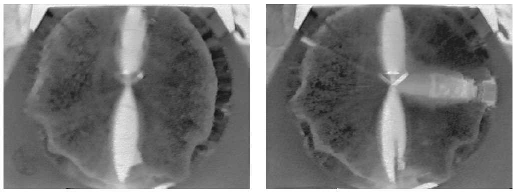
(1). The omni-view image is captured and transformed by the IPM. The orientation projection **U** and rotation-invariant Fourier amplitudes $\mathbf{A} = \{a_k, k = 0,\ldots, N-1\}$ are obtained. For comparison, the auto-correlation function (ACF) sequence is also calculated using the inverse Fourier transform of the energy spectrum.

(2). The rotation-invariant data (**A** or ACF) is used to decide the road category by the RCN.

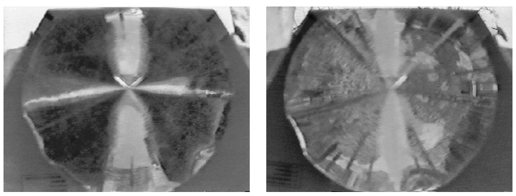
(3). The road category estimation is used to activate the corresponding RON, and the rotation-dependent data (original orientation projection **U**) is fed into that selected RON to estimate the road orientation.

The separation of the road classification and orientation estimations has several advantages. First, as the rotation invariant data are used as the input of the RCN instead of the original image, the distinctiveness of the input units is increased, and therefore the complexity of the network is reduced. If the Fourier amplitude A is used as the input of the RCN the number of the input units can be reduced

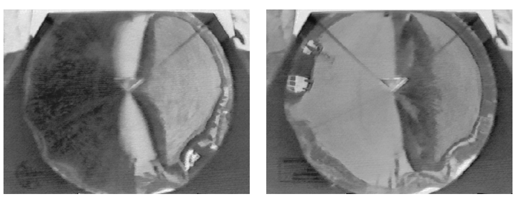
$(P \leq \frac{N}{2})$. Second, since a separate RON is used to estimate the road orientation for each road category, and the classification result is used to select the right road orientation estimation, the efficiency of the networks will be improved.



(a) Paved straight road ("||") . (b) T junction ("T")



(c) Intersection ("+"). (d) Earthy road("D").



(e) Curved road ("C") (f) Square ("S")

Fig 5. Sample images of the six road categories.

5. Results

5.1. Collecting the data

The data for the RUNN are collected while the robot is moving on the road. At the beginning of data collecting for each road category, the robot heads for the front road (i.e., the road orientation angle is 0), and the desired outputs of the RCN, representing the road categories, are input (assigned) by human supervisor. For the current experiments, The input data of the RCN are N=32 sub-sampled elements of the orientation projection. The road images are classified as 6 categories (M=6). Fig. 5 shows one typical sample image for each of the six categories. Fig. 6 shows the corresponding projections, Fourier amplitudes and phases, and the auto-correlation functions of these images. The sample number of real images captured by the OVI sensor is 2342 for the training and testing of the RCN. For RONs, each sample has as many as L rotated versions.

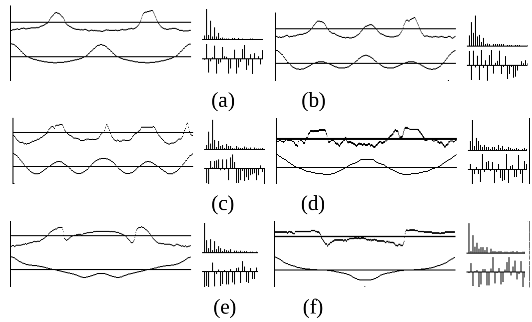


Fig. 6. The normalized projection (N=360, upper left), the first 30 Fourier amplitudes (upper right) the first 30 Fourier phases (lower right) and the ACF (lower left) of the corresponding image of each category in Fig. 5.

5.2. Training and testing

The back-propagation learning strategy is used to train the network. In our implementation, the normalized cumulative delta learning rule [7] is used for the RUNN. First the rotation-invariant Fourier amplitudes (FAs) are used to train the RCN. In this case the RCN has 16 inputs ($a_1,...,a_{16}$) and 6 outputs. The training and testing process indicates that 4 is the minimum number of the hidden units for proper classification, and 12 is the best number. The recognition rates for the training and testing sets are 98.5% and 94.4% respectively when the number of hidden units is 12. The rates are 96.7% and 92.5% respectively when the number of hidden units is 6. Comparing with the learning process of the networks using rotation-independent orientation projection as inputs in [6], the network RCN converges much slower and the recognition rate is slightly lower, and more hidden units are needed for the best classification. The reason may be that the Fourier amplitudes lose the phase information of the orientation projection and the Fourier transform compacts the energy in the first several terms so that weights between the input and the hidden units are not balanced.

So we also try to use the auto-correlation function (ACF), which is calculated as the inverse Fourier transform of the energy spectrum of the orientation projection and is still rotation- invariant, as the inputs of the RCN. Although this transform does not add or lose any information, experiments indicate that this kind of data representation may be more distinctive for the BP network (refer to Fig. 6). The testing results using ACF show improvements in convergence speed (half the iteration number as using FA) and recognition rate while the number of hidden units is fewer. The recognition rates for the training and testing sets are 97.0% and 92.7% respectively when the number of hidden units is 6. Analysis of the weight patterns shows that the connections between the input and the hidden layers extract more distinctive features from these kinds input patterns, which are more suitable for road classifications. However the improvements of the RCN with input ACF require more computation power for inverse Fourier transform and processing more input units.

Table 1. Performance of the two-layered RONs

class	‖	T	+	D	C	S
training set	100.	99.9	100.	83.9	96.2	89.4
testing set	99.0	100.	100.	83.1	95.7	86.7

After the road category is determined, the corresponding RON for this category is activated. We use the original orientation projection U as the inputs of the RCN. Experiments indicate that the RONs with no hidden units (i.e., two layered BP networks) perform best for all the road categories. Table 1 shows the correct orientation estimation rate, measured by percentage of errorless estimation for each road category. Analysis of the RONs weights reveals that the operations of the two layered orientation networks are quite similar to correlation functions. The estimation accuracy decreases when the input data become noisy and scattered for a certain category (e.g., “D”), which can be further divided into several categories. This is the reason why we classify the roads before orientation estimations.

5.3. Practical considerations

In the current experiments of the RUNN training and testing, the orientation projection is re-sampled from 360 to 32 since the software Nworks running on a PC486 is much slow with large number of PEs. Correspondingly the number of the outputs of the RCN, the estimated orientations, is sampled to 32, which means the angle resolution is about 11 degrees for the 360 degree view. This problem can be simply solved by using 360 inputs of original orientation projection to the RONs. Since the energy of Fourier spectrum compacts to the first few items, so the number of the inputs of the RCN, P, is much smaller than N. Experiments show that P=30 is enough to represent each category (refer to Fig. 6), so the scale of the RCN does not increase. Even if the ACF is used, the number of the input data is N/2 since ACF is symmetric about N/2. However if we expect 1-degree angle resolution for the orientation estimation, the number of both input and output units, Q and L, should be 360. It means there are N^2 (about 130K if N=360) connections and weights for each of the RONs. In the initial stage of the operation or the recovery stage when the robot has missed the road, The system should search for all the 360 directions. Fortunately in the normal conditions of continuous road following, the system only needs to search for a narrow range of orientations, for example, from -16 degrees to 16 degrees with 1-degree interval. Hence most of the outputs and the corresponding connections could be disabled.

6. Conclusion and discussions

In this paper we present the current results of training and testing the backpropagation network for the outdoor road scene understanding using omni-view images. Both the road orientations used for robot heading and the road categories used for robot localization are determined through the integration of invariant image analysis and adaptive neural networks. Experimental results with real scene images are promising. In order to actually apply the neural networks to the real world autonomous robot navigation in the outdoor natural scene environment, further works are needed.

Keeping the requirement of extracting rotation invariant image features in mind, and at the same time reducing the dimension of the original 2D omni-view images, we use the principal component analysis (the KL transform) along the radius axis r of the omni-view polar image I(θ,r) to obtain eigenfeatures for a given θ

$$\mathbf{U}(\theta) = (u_0(\theta), u_1(\theta), \ldots, u_S(\theta)) \qquad (13)$$

where S is expected to be much smaller than the original dimension of r. Preliminary experiments show that the first 3 components of KL coefficients of the 1D radius image can properly represent the original 2D omni-view image with r from 0 to 128. The sampled eigenfeature sequence along the orientation direction can be expressed as

$$\{\ \mathbf{U}(n) = (u_0(n), u_1(n), \ldots, u_S(n)),\ \ n = 0, \ldots, N-1\ \} \ (14)$$

Sequence $\{u_0(n)\}$, the first component of $\mathbf{U}(n)$, is approximately the orientation projection defined in the paper. The rotation-invariant vector sequence can be obtained by applied the 1D Fourier transform to each component sequence.

Spatio-temporal pattern recognition (SPR) network was proposed by S. Grossberg to explain certain cognitive process for recognizing sequence of events. The primary application of SPR networks appears to be in the area of recognizing repetitive audio signals. It is straightforward to apply the SPR network for recognizing image sequence of outdoor road scene. Since the same road category will last for a period of time, SPR network should not be sensitive to the occasional image events and could give a more robust recognition.

References

[1] R. T. Elkins and E. L. Hall, Three dimensional line following using omnidirectional vision, Proc. SPIE vol. 2354 , 1994: pp 130-144

[2] Y. Yagi, S. Kawato, S. Tsuji, Real-time

omnidirectional image sensor (COPIS) for vision-guided navigation, IEEE Trans Robotics and Automation, vol 1, no 1, Feb 1994 :pp 11-27

[3] J. Hong, X. Tan, B. Pinette, R. Weiss, E. M. Riseman, Image-based homing, Proc Int Conf Robotics and Automation, April 1991, pp 620-625.

[4] F.Stein and G. Medioni, Map-based localization using the panoramic horizon, Proc. ICRA, 1992

[5] D.A. Pomerleau, Neural network based autonomous navigation, In Vision and Navigation: the CMU Navlab, Kluwer Academic Publishers, Boston, 1990.

[6] Z.G. Zhu, G.Y. Xu, Neural networks for omni-view road image understanding, J. of Comput. Sci. and Technol., vol 11, no 4, July 1996.

[7] Neuralware Inc., Neural Computing / Reference Guide/ Using Nworks of Professional II/Plus, 1991.

[8] Aspex Inc., The PIPE reference manual , 1989.

[9] Z.G. Zhu, H.J. Xi, G.Y. Xu, Combining rotation-invariance images and neural networks for road scene understanding, Proc IEEE ICNN, 1996: pp 1732-1737.

[10] T. Jochem, Using virtual active tools to improve autonomous driving tasks, Technical Report, Nov, 1993.

[11] T. Jochem, D.A. Pomerleau, C. Thorpe, Vision guided lane transition, IEEE Sym. on Intelligent Vehicles, Detroit, USA, Sep 25-26,1995