EMS-Vision: Gaze Control in Autonomous Vehicles

M. Pellkofer and E. D. Dickmanns

Institut für Systemdynamik und Flugmechanik
Universität der Bundeswehr München (UBM)
85577 Neubiberg, Germany

Martin.Pellkofer@UniBw-Muenchen.de

Abstract

The paper describes an approach to an optimal gaze control system for autonomous vehicles. This gaze control system should not only determine the viewing direction ad hoc for the present moment, but also plan and optimize the viewing behavior in advance for a certain period of time. For planning the viewing behavior the situation must be predicted. The expression ‚situation' includes not only the physical situation, but also the so-called perceptive situation and subjective situation. The perceptive situation takes the present perception status, the present knowledge and the perception capabilities of the autonomous vehicle into consideration. The subjective situation contains locomotion tasks, status of vehicle sensors and actuators and subjective meanings of external objects resulting from the own point of view. For simulating the situation, the states and variances of all objects must be predicted. In the system for EMS-Vision, parts of the gaze control unit are already implemented and tested in both UBM test vehicles VaMoRs and VaMP. The embedding of the gaze control unit in the EMS-Vision architecture is discussed.

Keywords – Gaze Control, Attention Control, Active Vision, Autonomous Vehicles

1 Introduction

Human beings are able to change their viewing direction in a very quick and complex manner. Thereby, periods of smooth pursuit are interrupted by quick changes of viewing direction, so-called saccades. Humans carry out between 3 – 5 saccades in one second. This viewing behavior represents a highly dynamic perception-action-cycle. Saccades are triggered by optical stimuli or by intention. The field of research 'active vision' tries to transfer such complex viewing behaviors to technical systems [1,2]. An essential feature of active vision is that the sensors are used and adapted in dependence of internal knowledge, optimizing the information input by using its actuators. For active vision a knowledge representation is needed. In addition to this, an active vision system must know its limits of ability and must appraise the reliability of its perception and knowledge. The ability of self-monitoring is to be implemented already in the data acquisition [6]. Active vision enables the validation of object hypotheses and assumptions and provides 'learning by doing' by monitoring the consequences of the own actions and by using learning algorithms.

The present article deals with an optimization algorithm for a gaze control unit in an autonomous vehicle. This algorithm simulates the effects of alternative sequences of saccades and smooth pursuits to find the optimum. The optimum refers to a certain cost function, which assesses e. g. the precision of available information, symbolic object meanings and attributes. For applications in autonomous vehicles the optimization algorithm must work in real-time.

Following this introduction, those factors are discussed, which influence the viewing behavior. After that, the MARVEYE camera configuration and the optimization problem in the gaze control unit is described. The fourth paragraph portrays how the gaze control unit is embedded in the EMS-Vision system. The last paragraph shows some performance characteristics of the pan-tilt camera head (TACC) of VaMoRs.

2 Factors Influencing Viewing Behavior

The gaze behavior must be organized in such manner, that the relevant elements of the environment can be perceived. The emphasis lies on 'relevant'. It is impossible to perceive and represent everything in the environment. The viewing behavior depends on the situation in which the system is. The expression 'situation' includes many factors, which can be roughly grouped into three categories: physical situation, subjective situation and perceptive situation. The following three subsections explain the classification.

2.1 Physical Situation

The own vehicle moves with a certain velocity in the world. Therefore, all objects in the environment have a position and velocity relative to the own vehicle. These state components of other objects influence the viewing direction drastically. For example, it depends on the relative velocity of another vehicle whether it is a danger for the own vehicle or not. Such circumstances must be considered by the gaze control unit. The physical situation is also influenced by the kinds of objects present. The object class describes the shape of the object and other object properties that may have an influence on viewing direction. For example, it makes a difference whether the object in front of the own car is a shimmying paper bag or a slinging car. Another aspect of physical situation is a rough classification of the environment, called domain, which is known in advance due to maps. For example, the domain could be a highway or a country road and leads to some boundary conditions for locomotion and viewing behavior.

For optimizing gaze behavior, the physical situation must be predicted for a certain period of time in advance. For this purpose, the state vector of each physical object of relevance must be predicted. The dynamical models used by the image processing modules can predict the state vectors within short time periods (a few multiples of 40 ms). For predicting within longer time periods background knowledge has to be used. For this purpose, object observation extending over several seconds must be carried out in order to classify the behaviors of external objects. With this classification, assumptions can be made about the future behavior of objects and their future state vectors.

2.2 Subjective Situation

Biology shows that there is no 'optimal eye': animals in different habitats and with different behavioral patterns may be equipped with very different eyes. Human beings with their elaborate sense of vision are also no 'general observers'. Ballard [2] writes that the attempt to perceive all situation aspects exceeds all available computing power systematically and Aloimonos [1] states that perception is always linked to a task or an intention.

Therefore, the following question must be asked: What tasks the viewing system has to perform and what informations are profitable for this tasks? Often less but accurate information is sufficient for a special task and many viewing systems for special tasks have already been developed.

In autonomous vehicles, the current and planned locomotion maneuvers influence the viewing behavior. Also, detection tasks for unexpected objects must be taken into consideration. Other aspects of subjective situations are the properties and abilities of the own vehicle. Taking such aspects into account, the gaze behavior can be altered, if system components fail or are not available, in order to get by with reduced functionality.

2.3 Perceptive Situation

Besides the physical and subjective situation, the viewing behavior must also correspond to the perceptive situation. The perceptive situation describes the different ways of perceiving objects in the environment, e. g. sensors and image processing modules needed, boundary conditions for perception and so on.

Up to four cameras with different focal lengths may be mounted on the pan-tilt camera head (TACC) in the EMS-Vision system (see also paragraph 3). This multi-focal configuration in combination with changing viewing directions leads to changing visibility of objects in the different camera images [3].

Objects can be perceived in different images by different perception abilities. These abilities supply information of different quality. The image processing modules in the EMS-Vision system supply the following quantities describing perception quality according to the object hypothesized [7]:

- State vectors including positions, velocities and shape parameters.
- Variance vectors taken from the covariance matrices P of the Extended Kalman-filters.
- The ratio between the number of measured edges to the number of expected edges, which describes the certainty of identification of the object class.
- The magnitudes of the residues which cannot be explained by the shape model.

The variance describes the uncertainty of the present state variable. Analog to the state vector, the corresponding variances can be collected to a variance vector. Physical and subjective situation aspects determine to what precision a special value must be known. For example, the meaning of another vehicle with respect to the own locomotion determines their maximally allowable variances in position and velocity. The difference between present variance and maximally allowable variance determines attention demand and viewing direction.

Also, the perceptive situation must be predicted for a certain period of time in advance. Starting from a certain gaze behavior it must be examined in what sensor at what time the objects are visible. If no sensor perceives an object, the quality of its object information will decrease with increasing prediction time. This information decay leads to increasing variances and must be modeled in so-called 'knowledge decay functions'. If an object is perceived by a sensor, its variances are influenced in depend-

ence of the sensor resolution and the processing effort: higher resolution and higher processing efforts generally lead to smaller variances. The dependencies between the different perception abilities and the variances achievable must be modeled in so-called 'knowledge gain functions'. A system can only optimize its viewing behavior, if it knows how to improve the quality of its knowledge.

3 MARVEYE

Before asking what viewing directions are reasonable, the camera configuration must be determined. The MARVEYE (Multi-focal active/reactive Vehicle Eye) camera configuration is an arrangement of four cameras with three different focal lengths (figure 1). Two cameras are equipped with wide-angle lenses set up as a horizontal stereo pair with skewed optical axes. The third and fourth camera are mounted with a mild and a strong tele-lens. The respective focal lengths are: either 6 mm, 24 mm and 75 mm resulting in viewing angles of 58^0 (CCD-chip size ½''), 11.4^0 (CCD-chip size ⅓'') and 5^0 (CCD-chip size ½''). Objectives with slightly different focal lengths are also used [5]. The angle between the main optical axis and the wide-angle cameras is set in a manner that the wide-angle and the mild tele camera overlap in the field of view of the mild tele camera.

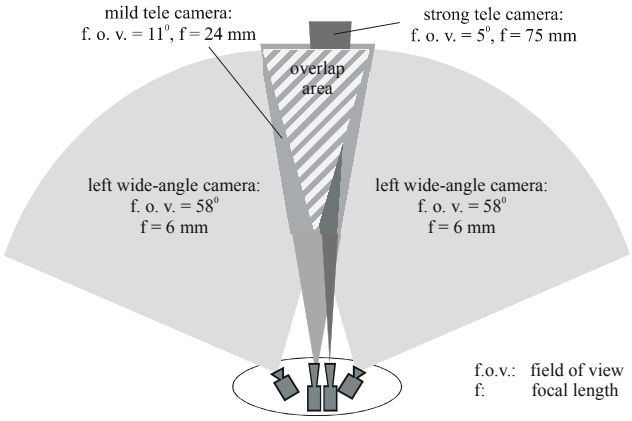

Figure 1: MARVEYE camera configuration

The simultaneous field of view (~ 105^0) of the MARVEYE configuration supports the following capabilities [4,7]:

- detecting new objects, e. g. overtaking vehicles or landmarks,
- state estimation of known objects,
- tight maneuvering on low order roads,
- estimating the distance to leading vehicle in the overlap region using stereo processing.

The MARVEYE camera configuration is mounted on the TACC, so that the gaze control unit can orient the tele cameras to an arbitrary part of the front hemisphere. The image resolution of the tele cameras is 3 resp. 10 times higher than the resolution of the wide-angle cameras. This high resolution and the different look-ahead distances are useful for:

- identifying new objects, especially at high speed,
- estimating object states and parameters with high precision,
- road curvature estimation and
- landmark navigation.

4 Optimization of Viewing Behavior

The second paragraph has shown that very many factors influence the viewing behavior. In comparison to this, the gaze control unit, as it is defined in the EMS-Vision project, has only two actuators: the motors of the pan and tilt axes of the TACC [5]. Using these two motors, the gaze control unit can change the orientation of the cameras mounted on the TACC in a large range. But how must the TACC be moved in order to optimally meet the situation aspects discussed in the second paragraph?

The TACC in the EMS-Vision system has an angular range of approx. 100^0 in pan and of approx. 40^0 in tilt. But not all angles in this range represent useful, alternative viewing directions. The existence of objects, object hypotheses and detection ranges determines alternative viewing directions of interest. For example, figure 2 shows the alternative viewing directions for four objects. If objects can be imaged simultaneously, a viewing direction covering both can be taken into consideration.

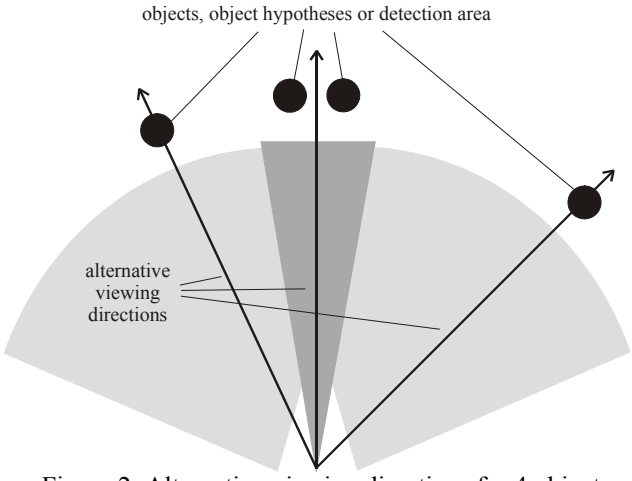

Figure 2: Alternative viewing directions for 4 objects

The gaze control unit plans the viewing behavior for some seconds in advance, so that another degree of free-

dom in determining gaze behavior is the timing of the single viewing directions. For planning the gaze behavior, object positions and velocities must be predicted. Moving objects induce changing alternative viewing directions. To get by with changing alternative viewing directions the gaze control unit handles objects in four different manners:

- the object is imaged by a tele camera,
- the object is fixated in a tele-lens image for compensating relative object motion (smooth pursuit),
- the object is imaged by a wide-angle camera,
- the object is not imaged.

The handling of different objects by the gaze control unit and their timing represent different gaze behaviors. The optimization of gaze behavior is done with a cost function. The cost function should assess all situation aspects described in the second paragraph in a suitable manner. Generally, missing information of precision necessary raises costs. For example, if an actual variance approaches its maximally allowable value, the cost function delivers a high cost value. Perceiving an object reduces its variances and raises the difference between actual and maximally allowable variances. For large differences the cost function supplies small costs. Predicting the physical, subjective and perceptive situation and using the cost function, the total costs of alternative viewing behaviors are calculated. The viewing behavior with the smallest total costs is optimal with respect to the cost function and represents the optimal viewing strategy.

5 Gaze Control in EMS-Vision

Maurer [9] suggests an architecture containing three behavior decision modules for the different aspects of behavior (see figure 3): Central Decision (CD), Behavior Decision for Gaze & Attention (BDGA) and Behavior Decision for Locomotion (BDL). CD, BDGA and BDL have to work on a uniform model for behavior and a common scene representation. The model for behavior, which consists of behaviors of different levels of abstraction and the scene representation are under development.

The module BDGA contains the planning part of the gaze control unit. Figure 4 shows the functional parts of the behavior module BDGA: Situation Assessment for Gaze & Attention (SAGA), Visibility Analysis for Gaze & Attention (VAGA) and Optimization of Viewing Behavior (OVB).

The paragraphs 5.1 – 5.3 discuss SAGA, VAGA and OVB. Paragraph 5.4 describes the server process Gaze Control (GC), which realizes the executive part of the gaze control system.

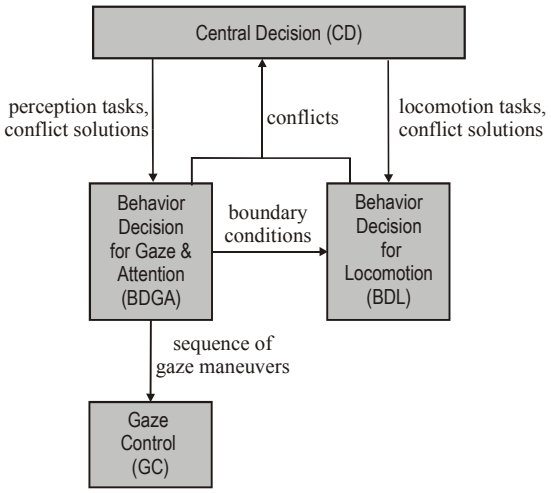

Figure 3: Modules for behavior decision and gaze control

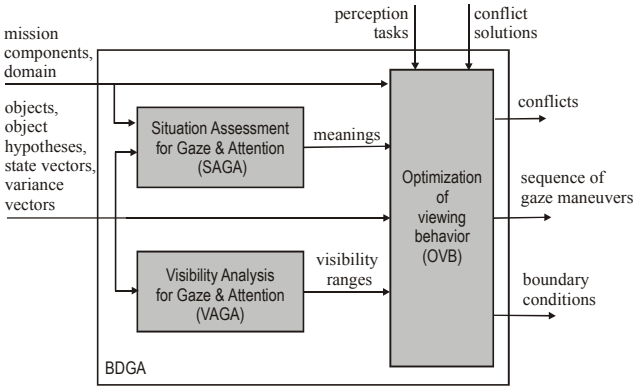

Figure 4: Behavior module BDGA

5.1 Situation Assessment for Gaze & Attention

Besides their physical values for position, velocity and shape, objects can also have symbolic attributes and meanings. These are not used for a complete scene representation, but for a specific one with respect to special tasks.

As discussed above, for predicting state variables over several seconds, background knowledge is needed, which must be stored in the scene representation. An example: The observation of a car in front of the own car indicates an abnormal driving behavior. A car with an abnormal driving behavior is far more dangerous than a car with a normal driving behavior. Due to this, a symbolic object attribute '*abnormally driven vehicle*' makes sense and can be stored in the scene representation. Such symbolic attributes and meanings can be taken into account by the optimization of viewing behavior in OVW. In the same way, other symbolic attributes may be useful, for example the attributes '*leisurely driven vehicle*' or '*aggressively*

driven vehicle', which correspond to the frequency of lane crossings and the distances being kept to other vehicles.

Also, the mission context may lead to classification features for symbolic situation assessment. Objects of the same class can have different meanings due to the current locomotion task [4,11]. For example, two road crossings can simultaneously appear in the scene representation, but only one can have a special meaning with respect to a planned turning off.

As figure 4 shows, the following information is used by SAGA: the present and planned mission components, the domain (classification of the environment) and states and variances of all objects. Within the module SAGA, decision trees use the incoming information to assign symbolic meanings and attributes to physical or mental objects of the scene representation. Thereby, the object class restrains the possible attributes and meanings. The module OVB receives and uses symbolic attributes and meanings for optimization.

5.2 Visibility Analysis for Gaze & Attention

The module Visibility Analysis for Gaze & Attention (VAGA) uses simple shape models for examining the visibility of objects in different cameras. The visibility of objects depends on the viewing direction of the TACC and the distance between camera and object. But also occlusions between objects and the aspect conditions are taken into account. VAGA sends sensor-specific visibility intervals (windows) for every physical object to the optimization algorithm in OVB. If the viewing direction of the TACC lies within the visibility interval, the object is visible in the respective camera. Outside the visibility interval, the object is not visible in the respective camera. Here, the expression 'visible' involves also the object size in the image. For image processing, the object size must not be too large or too small.

Besides VAGA, also the perception experts can send visibility intervals for their objects to OVB. This procedure can be reasonable for objects with complex shape models or for large objects, e. g. roads [7].

5.3 Optimization of Viewing Behavior

For optimizing viewing behavior the following informations are considered: the present and planned mission elements, the domain, all objects with their states, variances, symbolic meanings and attributes and visibility intervals. The module CD sends perception tasks to OVB. The perception tasks specify the objects to be perceived and object hypotheses to be tested. The corresponding maximally allowable variances are also given.

If the optimization algorithm finds an optimal viewing behavior in form of a sequence of gaze maneuvers, this sequence is sent to the executive part of the gaze control unit, called server process Gaze Control (GC) (see figure 3). If the optimization algorithm finds no viewing behavior, which suffices all perception tasks, a conflict message is sent to CD. A perception task fails, if a variance exceeds its maximally allowable value.

A situation may require to specify boundary conditions on locomotion. For example, it can be favorable to reduce the speed of the own vehicle in order to get more time for perception or to change the lane in order to get better aspect conditions. Such boundary conditions are sent directly to the module BDL.

5.4 Gaze Maneuvers

The GC process communicates with the TACC embedded controller system and connects it with the PC-net. GC offers and performs gaze maneuvers, monitors the performance of active maneuvers and writes the TACC state and status in a buffered scene node representing the TACC within the Dynamic Object Database (DOB) [5,10]. Through this exchange, every process in the system can read the angles, velocities and status of the TACC for any point in time within the buffered time period. Among others, the gaze maneuvers of the GC process include the following functionality :

- With a **saccade** an arbitrary camera can be oriented to a physical object or a point in object coordinates. The start and the end of the saccade is signaled to the system.
- With a **smooth pursuit** a moving object can be kept in a camera image. If the discrepancy exceeds a certain threshold value, an intermediate saccade is started to center the object in the image.
- With **scan paths** a certain part of the environment can be scanned with a high resolution sensor for detecting new objects. For this purpose, search paths are planned and performed by the TACC.

The GC process sends the control parameters being optimal for the current gaze maneuver to the TACC and commands set values (angle, angular velocity or both). In such manner, GC executes the gaze maneuvers determined by BDGA.

6 TACC Performance

The TACC is able to perform saccades in a very quick manner. Figure 5 shows a data plot of the TACC in the experimental vehicle VAMORS. There were three cameras mounted on the TACC. The TACC executes a sequence of saccades with different amplitudes. In figure 5, it can be seen that the setting time of a saccade depends on its amplitude. The setting times are between 160 – 390 ms.

There is no great difference between the setting times of pan and tilt axes. The higher moment of inertia of the pan axis is balanced by a stronger pan motor. TACC reaches angular velocities of up to 280^0 per second.

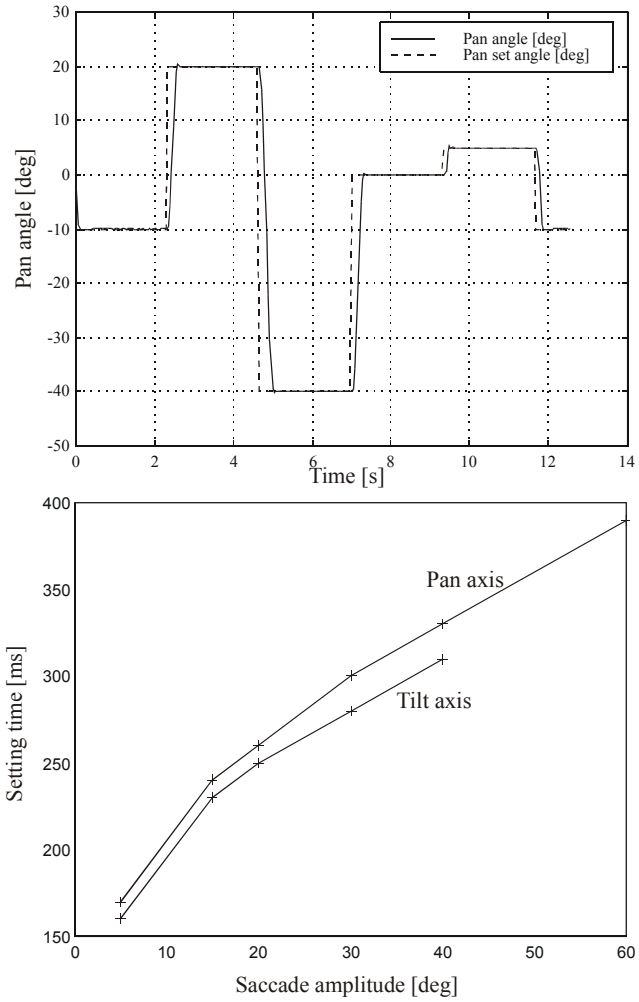

Figure 5 : Sequence of saccades (TACC of VAMORS)

7 Conclusions and Outlook

The paper has described a gaze control unit for an autonomous vehicle. First, situation aspects influencing the viewing behavior have been discussed. After that, the optimization algorithm in the gaze control unit has been portrayed. The algorithm proposed simulates the effects of alternative sequences of saccades and smooth pursuits to find the optimum. The cost function assesses the situation aspects which influence gaze behavior, e. g. the precision of available information, symbolic object meanings and attributes. It has been shown how the gaze control unit is embedded in the EMS-Vision system. Situation aspects influencing the viewing behavior are reflected in the EMS-Vision architecture. Some parts of the gaze control unit have already been implemented and tested in the UBM test vehicles VAMP and VAMORS. The last paragraph has shown some performance characteristics of the TACC in VAMORS.

Presently, an autonomous turning-off maneuver with the test vehicle VaMoRs is under development. First results are discussed in [7,11]. Tests with elaborate gaze strategies will follow. The uniform model for behavior and a common symbolic scene representation for the behavior modules in the EMS-Vision system are under development.

References

[1] Y. Aloimonos, Introduction, “Active Vision Revisited”, *Active Perception*, Y. Aloimonos (Ed.), pages 1-18, Erlbaum, Hillsdale, 1993

[2] D. H. Ballard, Ch. M. Brown, “Principles of Animate Vision”, *Active Perception*, Y. Aloimonos (Ed.), pages 245-282, Erlbaum, Hillsdale, 1993

[3] E. D. Dickmanns, “An Expectation-based, Multi-focal Saccadic (EMS) Vision System for Vehicle Guidance”, *Intern. Symp. on Robotics and Research*, Salt Lake City, Utah, October 1999

[4] R. Gregor, E. D. Dickmanns, “EMS-Vision: Mission Performance on Road Networks”, in [8]

[5] R. Gregor, M. Lützeler, M. Pellkofer, K. H. Siedersberger, E. D. Dickmanns, “EMS-Vision: A Visual Perception System for Autonomous Vehicles”, in [8]

[6] F. H. Hamker, H.-M. Groß, “Intentionale Aufmerksamkeit: Ein alternatives Konzept für technische visuo-motorische Systeme”, *Aktives Sehen in technischen und biologischen Systemen*, B. Mertsching (Ed.), pages 101-108, Infix, Hamburg, 1996

[7] M. Lützeler, E. D. Dickmanns, “EMS-Vision: Intersection Recognition on Unmarked Road Networks”, in [8]

[8] I. Masaki (ed.), *Proc. Int. Symp on Intelligent Vehicles*, Dearborn, USA, IEEE Industrial Electronics Society, Oct. 2000

[9] M. Maurer, “Knowledge Representation for Flexible Automation of Land Vehicles”, in [8]

[10] A. Rieder, “Fahrzeuge sehen”, Dissertation, Universität der Bundeswehr München, LRT, to appear 2000

[11] K. H. Siedersberger, E. D. Dickmanns, “EMS-Vision: Enhanced Abilities for Locomotion”, in [8]