ARtificial GUidance System

ARGUS

— Part B —

IST-2000-25032

May 11, 2000

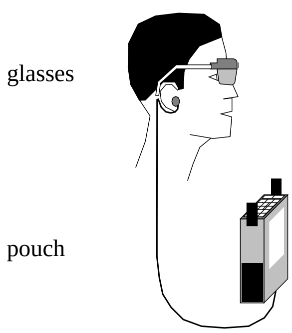
Contents

B.2 Introduction

Within the project ARGUS[1] we will develop a portable object recognition system which will meet the navigational needs of blind and visually impaired people. The system will complement existing navigation aids by providing significant spatial information to the users about their immediate environments. It will consist of two cameras on a spectacle-like device, connected to a computer. It works by interpreting the visual environment and informing the user about selected views or classes of objects by speech-output.

The basic functionality is similar to that of an electronic guide dog. A dog that can read and talk. In normal operation, the user will request (by pressing a key or combination of keys on a keypad) information about the location of a particular (type of) object from the system (e. g. a mail box or zebra crossing, taking into account the different models used in the individual member states of the EC), and the system will inform the user as soon as such an object becomes visible. Additionally, it will be possible to have ARGUS read signs or remember particular views of an object which will be recognised the next time the same object is approached from approximately the same position.

ARGUS is intended as a secondary mobility aid, i. e. an extension to the long cane (or, to a lesser degree, a guide-dog). It is assumed that the user has already completed mobility training and is reasonably proficient with his primary mobility aid to traverse a well known path with acceptable confidence. ARGUS will then add to the user's ability to react to unforeseen situations; it will greatly enhance his perception of his surrounding; and it will ultimately allow him to explore hitherto unknown areas of his surrounding with confidence, and without the need to rely on other people's help. By doing so, ARGUS will greatly increase the quality of life of blind and visually impaired people, as well as their potential employment prospects.



The figure to the left shows a possible implementation of ARGUS in its final stage. It consists of a spectacle-like device (the glasses) and a microcomputer (the pouch), connected by a cable. The glasses contain two sub-miniature CCD-cameras, an electronic plummet, and two miniature speakers; the pouch contains the electronics for the camera as well as a GPS and a mobile computer, the power supply for the system, and a keypad for interaction with the computer. A finished industrial product based on ARGUS might additionally contain ultrasonic sensors for obstacle avoidance and use speech-recognition for input — both are readily available today and could easily be integrated into an industrial product, but are not part of this proposal.

The functionality planned for ARGUS as outlined in this proposal (of which a more detailed description can be found at the beginning of Section B.6) is based on an analysis of the orientation and mobility requirements of the blind and visually impaired which was carried out as part of a similar project (see [5, 16, 25, 33], compare Part C for additional information on the consortium's background, which includes several National and European projects). However, further analysis of user needs and user requirements is an integral part of the project and will be carried out by partner 8, an umbrella organisation of organisations for the blind. ARGUS has been designed under the assumptions that the long cane will remain the primary mobility aid in the near range (as mentioned above), and that all other senses, in particular the sense of hearing, should remain fully usable. Therefore, ARGUS will be strictly demand-driven, i. e. only if and when the user is actively requiring information will ARGUS become active, and it will represent the result of its analysis as concisely as possible by audio (using speech synthesis software, which is readily available).

In the following, we will describe the individual scientific and technological objectives aimed at within this project.

[1] Named after both the four-eyed — or, in a classical case of mythological inconsistency, up to a hundred eyes — Arcadian hero from ancient Greek mythology, as well as Odysseus' faithful dog.

B.3 Objectives

B.3.1 Recognition of Special Objects

Blind persons do not have access to much of the infrastructure we take for granted — pedestrian crossings, benches, or letter-boxes, to name just a few — simply because they do not know where in their vicinity such objects are located. ARGUS will be able to recognise a number of particularly important objects — selected in collaboration with organisations for the blind — and inform the blind user about the distance and direction to these objects.

We will develop fast and reliable algorithms for the detection of typical objects, as well as a number of relevant traffic signs (e. g. bus-stop) and public information symbols according to ISO 7001 [15], also called icons throughout this proposal. The algorithms must be able to locate at least 70 % of these objects and 50 % of all relevant traffic signs within a distance of 7.5 m, and 35 % of all icons within a distance of 3 m, assuming a maximum slant angle of 15 ° (DIN 66 079 [14]), with a rate of false positives of usually no more than 1 %, and less than 0.1 % for safety-critical objects (e. g. pedestrian-crossings).

B.3.2 Recognition of Trained Landmarks

The ability to train ARGUS to recognise arbitrary views plays an important role within the overall concept of ARGUS, as only the ability to train and recognise arbitrary views allows the user to employ ARGUS for tasks which could not possibly have been foreseen by the developers of the system; a simple example is the recognition of an acquaintance's doorway. These algorithms will naturally interface with the GIS described below. While the work described above will be realized by matching relational descriptions based on prominent edge and region features, we will also be working on a structural representation of the environment which can arrange these objects in order of their relative depth and produce a rough 3D sketch.

The result will be fast and accurate algorithms which will allow the correct identification of at least 75 % of all trained landmarks if the actual view and the trained view overlap at least 60 %, the orientation difference is less than 10 ° and the lighting conditions fall into the same category (either daylight or fixed artificial light).

B.3.3 Text Recognition

One functionality high on the list of priorities developed in collaboration with organisations of the blind [5,16,25,33] is the recognition of text as it surrounds us in everyday life. This could be signs like "Bank" or "Supermarket" or the names of individual shops on one end of the spectrum, and street-names or the number and destination of an approaching bus on the other end. The main problem to solve is that of segmentation, i. e. the identification of text within a very varied and cluttered environment, and subsequent rectification. This differs markedly from the usual application of OCR, where text is always encountered head on and within a well-structured environment. However, the necessary research into low-resolution text-recognition might also be a good starting point for a future general (book) reading function beyond the scope of this proposal, but high up on the wish-list of the blind.

We will develop fast and accurate algorithms for the detection of printed text. The algorithms must be able to locate at least 50 % of all text on signs and windows etc. with a letter size of at least 15 cm, between 50 cm and 2.5 m from the ground level, and within a distance of 5 m. A correct recognition with a slant angle of up to 45 ° and a tilt angle of up to 15 ° must be obtained.

B.3.4 Obstacle Detection and Centre-Path Travel

Another functionality with high priority is centre path travel and obstacle detection. Centre path travel refers to the system's ability to guide a blind person along a path, usually delimited by a wall, fence, or hedge on one side and a kerb on the other. Obstacle detection in this context refers to objects which are not easily or reliably detected by the use of the long cane, in particular small objects (about the size of a brick or smaller) and depressions, e. g. stairs or road works[2]. Railway platforms are an example where both functionalities merge.

Specifically, it is our aim to detect and track the ground plane orientation and its distance from the camera system. The location of the kerb should be correctly detected and tracked, and the user can be informed whenever he is found walking towards the kerb. False positives should be less than 1 %, while the number of false negatives should be below 10 %. Small obstacles (e. g. boxes or bricks) with heights between 10 cm and 30 cm that lie between 5 m and 10 m ahead of the person will be detected and their position, depth and height will be estimated. Pits in the pavement will also be detected. The rate of false negatives will be less than 10 % and the rate of false positives less than 2 %. At least 70 % of all stairs will be detected.

B.3.5 Geographic Information System

We believe that several of the computer-vision tasks mentioned above could benefit from the use of additional location information as provided by a Geographical Information System (GIS) in conjunction with (differential) GPS, while on the other hand the information provided by the GPS could be improved upon using computer-vision algorithms.

We will develop a fast and efficient GIS based on input from (differential) GPS and image analysis (IA). This GIS must be able to track the trajectory with combined IA/GPS inputs to estimate the actual position with an accuracy of 20 m or much better. It must be able to provide the IA with all known information in a (semi)circle with a radius of 20 m around the actual position within 10 ms and all specific queries of a blind person in a town district of 1 km^2 within 5 s. Two specific sub-objectives are the realisation of a user interface that allows non-experts to enter or change information interactively on a PC screen, and to establish a list of important queries from the blind.

B.3.6 Hardware Development

Previous experience with similar projects [5,16,25,33] has shown that the availability of a hardware-prototype is essential for the success of a project like ARGUS, since only early access to the actual hardware used will guarantee the development of adequate solutions on the computer vision side.

We will therefore develop two different hardware-prototypes over the course of the project. The first one will consist of the cameras alone, and will interface directly to a standard laptop. This prototype will be available early on in the project (after month 6) and allow the development of algorithms adequate to the cameras' resolution, as well as the specification of a final prototype. This final prototype will be developed over the first two years of the project and will be produced in a small series of 8 units, which can then be used for evaluation purposes by the individual partners. The extreme demands in terms of processing power and portability, as well as the need to interface to various hardware requires a custom design.

B.3.7 Software Integration

The results of the individual partners concerned with the algorithmic side will be fused into a coherent software system capable to execute on the target hardware, which will be stable and easy to use by a blind person, so that field tests can be performed under as real as possible conditions.

B.3.8 Consultations with Organisations for the Blind

Partner 8 is an umbrella organisation of organisations for the blind. They will guarantee close involvement with these organisations throughout the duration of the project, and work on a continuous refinement of the catalogue of desirable features. They will also perform field tests with users from three different age-groups, where at least two blind and two visually impaired users from each group are selected (making for a minimum of 12 users per test).

[2]Obstacles from approximately waist height up are better detected by ultrasound, for which industrial products exist, compare Section B.5.1.

B.4 Contribution to Key Action Objectives

The objectives of the Key Action I — Systems and Services for the Citizen — encompass the cost effective and flexible access to general interest services for everybody, in particular with a view to people with special needs, including the elderly and people with disabilities. The implementation strategy is based on mid- to long-term research and technological development, and emphasis is placed on the creation of easy to use, dependable and cost-effective access to general interest services, meeting identifiable user needs.

In addition, the Action Line I.2.1 — Intelligent assistive systems and interfaces to compensate for functional impairments — within the RTD I.2 — Persons with Special Needs, including the Disabled and the Elderly — has the objective to enable citizens with specific impairments, especially those related to aging (and the elderly constitute approximately 70 %–90 % of the blind population), to benefit as fully as possible from intelligent assistive systems. The focus is on innovative assistive systems for supporting mobility, orientation, transportation, and vision as well as secure home and living environments, capitalising on recent advances in personal devices, and even allowing for longer-term research, based on an improved and detailed understanding of the nature of cognitive and sensory processes, which might be required into how advanced interfaces can increasingly compensate for the effects of impaired functionality on human performance. Finally, significant industrial participation is required to ensure commercial exploitation.

The above is reinforced by the Action Line I.3.1 — Systems and Services for Independent Living — which places the main emphasis on the development of new tools which will enable people with special requirements to live independently, addressing mobility and aiming to improve access to a wider range of services, as well as greater participation in social and community activities for people with reduced mobility or impaired functions. This includes extended employment and learning opportunities.

All of the above, which for the most part has been taken literally out of the European Commission's 1999 and 2000 workprogramme for IST, is almost an explicit description of the objectives aimed at within ARGUS. A rich infrastructure of everything from street-markings and traffic-signs to telecommunication and mail services, public transport and recreational infrastructure (benches as well as public toilets, to give just two examples) is already in place throughout Europe — and is virtually inaccessible to Europe's 6.5–7.4 Million blind and visually impaired[3]. It is the aim of ARGUS to develop easy to use, dependable and cost effective technology which will make this rich wealth of existing infrastructure accessible again. This will greatly enhance many blind persons' ability to go about their daily business without the help of a seeing person; it will add to the user's ability to react to unforeseen situations; it will greatly enhance his perception of his surrounding; and it will ultimately allow him to explore hitherto unknown areas of his surrounding with confidence, and without the need to rely on other people's help.

The latter is of particular importance for the elderly, which constitute the major part of the blind population. While the younger, and in particular people born blind, will often venture into unknown territory armed with nothing more than the long cane and a lifetime of experience, mobility doesn't come that easily to the elderly. These will generally lose their mobility together with their eyesight[4], making them dependent on other peoples' help in all but the most standard situations. ARGUS will allow them to go about their business with bolstered confidence, knowing that using ARGUS they cannot get lost and, what's more, will have access to much of the public infrastructure they took for granted before turning blind. This will not only increase their mobility, but also their independence, allowing them once more the active participation in social and community activities that sighted people take for granted.

[3] With the exception of a very limited area, usually along two or three well known routes, were the blind person was informed about the existence of that kind of infrastructure by a seeing person.

[4] While of course depending on the individual's personality, this is pretty much true for anybody loosing his eyesight after age 40.

B.5 Innovation

The project ARGUS will both lead to an innovative navigational aid for the blind as well as advance the state-of-the-art of Computer Vision System Design in several respects, due to its innovative use of object recognition as a navigational aid for blind people. We will first review the navigational aid technology and describe the innovative contributions of ARGUS for this application field.

B.5.1 Contributions to navigational aids

State of the Art Navigational aids for the blind and visually impaired go by many names: electronic travel aids (ETAs), electronic mobility devices, mobility aids, obstacle detectors, navigational aids, blind aids, electronic spatial sensors, spatial orientation aids, and others. We will use ETA in the following.

There is a long history of attempts and partial successes to provide ETAs for the blind, with patents issued as early as 1950. A list of currently available ETAs is kept by the Royal National Institute for the Blind [45] (RNIB) and can be ordered at: RNIB Joint Mobility Unit, 224 Great Portland Street, London W1N 6AA.

The classics among the ETAs include the Mowat sensor and the Nottingham obstacle detector (NOD) [17]. These are both hand-held ultrasonic torches, developed around 1980, that map ultrasonic reflections into vibrotactile or auditory signals, respectively. An improvement of the Nottingham obstacle detector (NOD) is known as the Sonic Pathfinder (1984) [12, 13]. This is a head-mounted pulse-echo sonar system comprising three receivers and two transmitters, controlled by a microcomputer. It does not give information about surface texture, and normally the auditory display indicates only the nearest object.

Individual research on sonar devices led to the "sonic glasses," or Sonicguide, around 1974 [17]. More recently, the angular resolution of this Binaural Sensory Aid (BSA) system was improved using a third narrow beam transmitter for creating an additional monaural signal. This new system is now known as the KASPA system (1994). KASPA represents object distance by pitch, but also represents surface texture through timbre.

More recently, interesting new results have been achieved in several EC projects funded through the TIDE program (Technology Initiative for Disabled and Elderly people).

The ASMONC system (Autonomous System for Mobility Orientation, Navigation and Communication) uses stereo vision, sonar and infra-red sensors for obstacle detection with GPS and mobile phone for navigation and communication [31, 32]. The ASMONC system performs an extensive 3D and motion analysis for ground plane tracking to enable reliable stair-case and kerb detection. Processing takes place in a backpack carried by the user.

In the project MoBIC (1994 - 1996) an aid was developed which built on the technologies of GIS (Geographical Information Systems) and GPS [35]. The MoBIC Travel Aid (MoTA) consists of the MoBIC Pre-journey System (MoPS) to assist users in planning journeys, and the MoBIC Outdoor System (MoODS) to execute these plans by providing users with orientation and navigation assistance during journeys. The MoTA is seen as being complementary to primary mobility aids such as the long cane or guide dog.

The project OPEN (Orientation by Personal Electronic Navigation) has developed an infra-red reactive audible sign based on beacons mounted in public environments. A user with a special receiver can hear the message which gives relevant directions, e. g. to ticket offices and platforms.

In further TIDE projects (OMNI, SENARIO) sensory aids have been developed for intelligent wheelchair navigation. The technology used is less relevant for ARGUS. The same is true for the PAM-AID [26] project, funded by the EC Telematics Applications Programme, which aims to develop a mobility aid for people who are both frail and visually impaired. It is motivated by the fact that the combination of visual impairment and frailty severely limits a persons independence.

A headline-grabbing experiment has been the cortical implant approach, pioneered in the 1960s [2] and early 1970s [4]. It uses an array of electrodes placed in direct contact with the visual cortex. This is also called a neuroprosthesis. The cortical implant is an invasive approach, requiring major surgery. Maximum resolution of the implants is so far [4] on the order of 10×10 pixel, a resolution of 512 pixel (e. g. 16×32 pixel) is foreseen for the near future. Individual pixels are perceived as phosphenes (sensations of light, flashes), while the input signal is a video signal after pre-processing (edge detection). It might be worth noting that even with the best acuity forecast for the near and midrange future the patient is still legally blind and dependent on some additional mobility aid, e. g. the long cane.

A different approach maps grey-scale images from a video camera into corresponding sound patterns, or soundscapes. This device, nicknamed The vOICe, has been implemented in low-cost hardware (1991). It attains real-time performance through a pipelined design, and offers a resolution of 64×64 pixel and 16 shades of grey.

Comparison The approach taken in ARGUS is similar to ASMONC regarding the use of stereo cameras and advanced Computer Vision methods. The functionality of ARGUS is much richer, however, due to its innovative recognition capabilities, as ASMONC was only attempting part of the functionality described as "Obstacle Detection and Centre-Path Travel" in Section B.3. ARGUS, on the other hand, will be able to recognise, firstly, selected generic object classes (e. g. pedestrian crossings, telephone booths) using model-based recognition technology. Secondly, ARGUS, with the help of algorithms capable of learning and adaption, offers the recognition of landmarks (user-selected views) which will be a completely new ETA facility. Thirdly, ARGUS will include sign reading functionality (traffic signs and public information symbols as well as limited text-reading capabilities) which is high up in the list of user requirements. In summary, ARGUS tackles several advanced recognition tasks which can provide the blind with highly useful information about their visual environment.

ARGUS will be similar to both ASMONC and MoBIC with regard to the use of a GPS-based GIS. The main difference to MoBIC is that the GIS only has a supporting role within ARGUS, while the complete functionality of MoBIC is based on a GIS. This makes ARGUS much more flexible with regard to missing information within the GIS, which, unfortunately, is still the case for virtually all but a very few selected places around the world (due either to the lack of a signal from the GPS, e. g. in indoor environments, or, much to often, the lack of adequate maps). Also, ARGUS' proposed ability to build its own GIS (at least to some extend) not only allows it to adapt to the users' needs in an unprecedented fashion, but will potentially also provide interesting input for other areas like robotics and unsupervised ground vehicles.

Finally, it should be noted that none of the more complex ETAs mentioned above have so far reached a product stage (only some of the ultrasonic devices are currently available as a product). The continued development of an ETA, building on the experience of prior projects funded by the EC and individual European countries, would give the EC a competitive edge in a field otherwise dominated by companies from the USA, which provide most of the assistive technology available today.

B.5.2 Contributions to Computer Vision

The Computer Vision tasks tackled by ARGUS are steps towards real-time recognition of real-life objects within a real-life environment, which is a central goal of Computer Vision. While the intended functionality of ARGUS is far from providing general object recognition capabilities, the research will undoubtedly contribute towards that goal.

One aspect of the ARGUS research is particularly likely to provide an innovative push: the investigation of the variability of visual features under real-life conditions. In order to achieve reliable recognition under a wide range of lighting conditions, varying viewpoints and accidental occlusions, landmark descriptions and models for generic object classes must incorporate knowledge about the effects of such changing conditions.

For example, generic descriptions will be considered as resulting from experience. This is particularly appropriate for landmark descriptions which are actually derived from experience. Experience-based Vision is a fairly new research direction within Computer Vision which stresses real-life and learning aspects.

Another aspect of ARGUS which is of interest beyond ETA applications concerns the real-time performance. The requirement of response times in the order of one second is representative for a large field of Computer Vision applications. Hence system design decisions taken in ARGUS and its outcome in general will be of use to future applications such as robot navigation, service robotics, unsupervised ground vehicles, sensory process control and others.

B.6 Project Workplan

B.6.1 Introduction

This section describes the structure of the workplan and the overall methodology used to achieve the objectives. In order to do so, it first gives an overview over the intended functionalities of ARGUS in Section B.6.2, and from there describes the different workpackages needed to achieve this functionality. Section B.6.6 explains the interconnections between the individual workpackages and, where needed, the management structure employed to cast the individual workpackages into a unified prototype.

B.6.2 Functionality

ARGUS in its final form will be a portable device capable of capturing images of the visual environment, analysing and interpreting these images, and presenting the resulting analysis in form of auditory output (sounds, words). It is designed as an aid to the orientation and mobility of the blind and visually impaired. Numerous additional applications can be conceived, ranging from navigation to automation.

The functionality planned for ARGUS is based on an analysis of the orientational and mobility requirements of the blind and visually impaired which was carried out during a previous similar project [6,16,25,33], see each partner's background in Section C. A repeated review and refinement of the functional requirements through partner 8, an umbrella organisation of organisations for the blind, will be an integral part of the project.

The functionality of ARGUS for perceiving and interpreting the visual environment has been designed under the assumptions that the cane will remain the most important tool in the near range, and all other senses, in particular the auditory, tactile, and, if applicable, remaining visual senses should remain fully usable.

In the following sections the functionality will be described in more detail.

Centre Path Travel and Obstacle Detection Probably the most classic applications within a mobility aid for the blind are centre path travel, i. e. the ability to travel along an imaginary line without much deviations from the intended direction, and obstacle detection. This at first seems to be the domain of the long cane, i. e. the blind person's primary mobility aid, but interviews with the blind have shown that extended functionality would be very welcome when it comes to the detection of small obstacles (approximately brick-sized) and centre path travel in difficult environment, for example on train platforms or generally in areas with sudden drop-offs (road works, etc.).

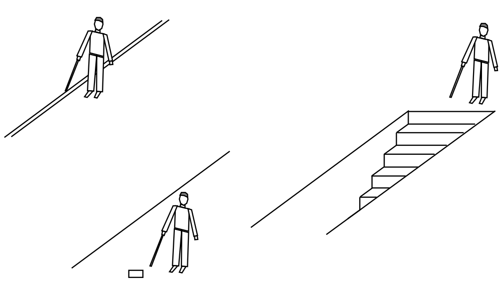


Figure 1: *Centre path travel (guiding along a line), detection of obstacles and stairs.*

ARGUS will therefore help a blind person to navigate safely. Safe navigation will be assured by detecting and tracking the relative position of the kerb with respect to the person. Small objects or depressions, which are often missed using the long cane or even special ultrasonic sensors, will be detected using computer vision techniques to estimate their location and dimensions in 3D space [3, 9, 27]. Furthermore, another useful feature of ARGUS will be the ability to detect stairs (e. g. subway stairs) and estimate their location in 3D as well as their orientation (ascending or descending).

The objectives described above will be achieved by a combination of texture segmentation and stereo-vision: the pavement and the road regions will be identified in both images based on texture segmentation. From these regions, a set of feature points will be extracted for both the region of the pavement as well as for the region of the street, and correspondences will be found. Using triangulation, the 3D coordinates of the corresponding points will be recovered and the 3D location of the ground plane will be predicted and tracked while the person is walking. Kerbs will be detected as the boundary between the two regions. Small obstacles or depressions (including sudden drop-offs) will be detected as connected regions with different texture within the pavement region. Stereo-matching [9, 41] will be used to calculate their 3D location, and the user will be informed about the obstacles. In addition, stairs will be identified by performing edge detection outside the pavement and road regions and searching for a set of concurrent lines [44]. After detecting such a set, the 3-D locations of these lines will be estimated and it can be determined whether the stairs are ascending or descending. In addition to our own work cited above, this will also build on similar work by others [7, 30, 31, 38, 39].

Recognition of Special Objects Many objects of the public infrastructure are easily identified by a combination of their 3D-form and colour-information, and ARGUS will contain models for a number of interesting objects from all participating member states. These models can be used to recognise the objects they represent independently of perspective distortions. Possibilities range from simple objects like bollards or traffic-lights to complex ones like cars. Some functionality already exists with regard to telephone booths, mail-boxes [23] and pedestrian crossings [44], which were implemented as part of a similar project [16, 25, 33]. New functionality can easily be added in form of a new model.

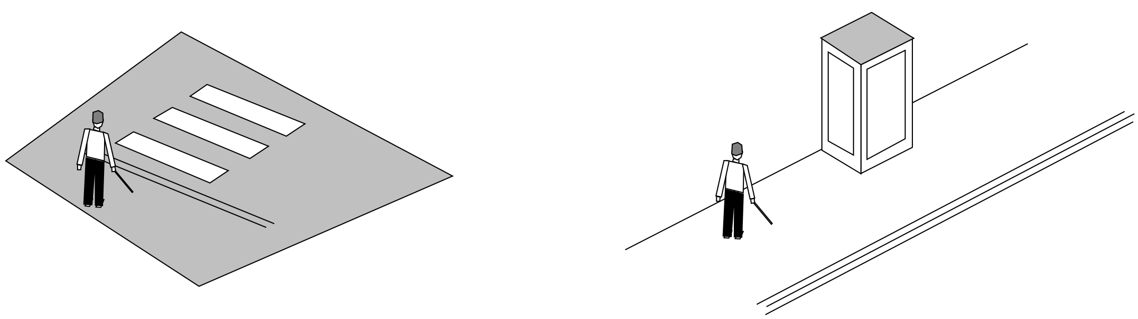


Figure 2: *Recognition of special objects (a pedestrian crossing and a phone booth).*

The approach used for the recognition of 3D-objects combines a multilevel-approach for the evaluation of the depth of selected lines (rather than computing an entire depth-map, which is time-consuming), fuzzy rules for colour segmentation, and the use of geometric hashing [21–24]. This combination results in an algorithm which is both robust and time-efficient. The algorithm needs to be complemented by a reliable verification procedure in order to weed out unwanted false-positives. Different approaches based on geometric constraints and texture analysis will be implemented in order to identify the most reliable algorithm.

Recognition of Trained Landmarks With this functionality the user can train ARGUS to recognise landmarks. Roughly speaking, landmarks are camera views which a blind user can store using ARGUS and which can be recognised later to aid his navigation, e. g. the entrance of a particular shop or a particular street corner. To store a new landmark, the user points ARGUS to obtain the desired view, possibly with assistance, and commands ARGUS to store the view as a landmark. A stored landmark may be activated by the user with the effect that ARGUS continuously checks whether the current view corresponds to that landmark. If this is the case, an appropriate sound or speech signal is generated.

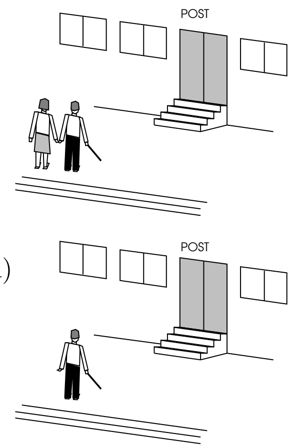


Figure 3: Recognition of trained scenes.

The recognition of landmarks will be realized by matching relational descriptions based on prominent edge and region features. Basic concepts for this have been developed during a similar previous project [33]. The recognition of arbitrary landmarks, which could be classified as experience-based vision, requires the development of feature attributes for recognition under variable lighting conditions as well as recognition with extended viewpoint variability. In addition, an effective recognition strategy including uncertainty management is needed, as well as landmark enhancement techniques based on multiple views.

While the work described above will be realized mainly by matching relational descriptions based on prominent edge and region features, we will also be working on a structural representation of the environment which can arrange objects in order of their relative depth and produce a rough 3D sketch. It is well known that from a single image it is not possible to make any reliable statements about the structure of the underlying scene. The use of stereo alleviates this problem to a certain extent, but is slow and burdened with problems when dealing with untextured surfaces. We will instead focus on the generation of a qualitative description of a scene based on a single image of the scene, but aided by additional knowledge about the general structure of the scene, geometric constraints, and possibly some selected stereo-cues, aiming for results which are comparable to the results of humans under similar circumstances (at a coarse level). In order to do so, we will combine approaches from different fields of computer vision, including projective geometry, texture analysis, graph-theory and traditional AI.

Localisation and Recognition of Icons and Text ARGUS will allow the localisation and recognition of standardised icons (public information symbols, ISO 7001 [15]), selected traffic signs, and simple texts, e. g. house-numbers or bus-lines. Many texts normally have a large size, a good contrast with the background, and a simple font. More complicated examples include shop names and street-name signs. The former can show rather exotic fonts and colours, while the latter can be relatively small, in varying locations, and often in bad repair. The recognition of all possible texts under different illumination conditions etc. is still a challenge. Nevertheless, by taking into account normal contexts like texts on rectangular backgrounds and a database of the most important texts, in general or in a specific neighbourhood, this problem can be solved. The processing steps include the detection of perspectively-deformed rectangular areas with high contrast (variance) or word-type texture, the localisation of the text outline within the rectangle, slant and tilt correction using the (deformed perspective) rectangle or text outline, thresholding

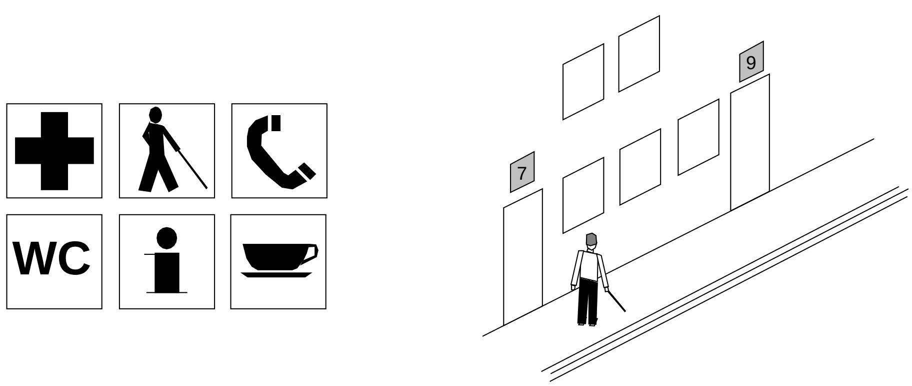


Figure 4: Recognition of icons and house-numbers.

and thinning, and character and word recognition with context help.

Similar considerations apply to the recognition of traffic-signs, which are easily recognised based on their colour and unique form, and public information symbols (icons). The latter probably form the group that is hardest to recognise, as they are usually held in black and white (i. e. no discriminating colour-information). Texture analysis will be employed to identify possible candidates.

Geographic Information System Using the current generation of microprocessors (Pentium III), real-time and unsupervised image analysis as aimed at in the the previous sections, corresponding to workpackages WP 1–WP 6, is still a formidable task. This task would be greatly simplified if a-priory information about the objects and obstacles around the observation position is provided and the image analysis can confirm this according to the viewing direction of the camera.

A GIS (Geographic Information System) together with high-precision GPS would be a relatively simple solution to the above mentioned problem. Normal GPS resolution is mainly limited by the U. S. A. Dept. of Defence through Selective Availability (SA), which causes an uncertainty of ± 100 m. Although the President has announced that SA will be turned to zero, presumably sometime between now and the year 2006, the question remains whether the uncorrelated uncertainties caused by other effects (ionospheric error of ± 5 m, HDOP of ± 5 m, ephemeris of ± 2 m, satellite clock of ± 2 m and receiver error of ± 1 m) will be tolerable. Differential GPS can in theory attain a better-than 1 m precision, although the precision measured under practical conditions is typically between 6 m and 32 m (from [20], 1-σ interval). Also, the price of differential GPS is prohibitive at this moment, even though prices are expected to go down. In other words, the experience with differential GPS in combination with a GIS within this project can be applied later when prices are no longer prohibitive. However, there are two other problems with the use of (differential) GPS and a GIS in general:

1. In streets between high buildings the satellite reception is sometimes insufficient and the precision drops. By tracking the trajectory of the blind user and by using the information from the image analysis, the GIS can still keep a relatively good position accuracy.

2. For certain regions (currently, and for the foreseeable future: most regions), a GIS might not be available, or the informations might not be up to date.

These examples show that a solely GIS-based solution would be insufficient; indeed the interaction between GIS and Image Analysis has to be bidirectional: the GIS informs the IA about expected visual input and the IA helps the GIS when the GPS temporarily drops out, or even helps to construct a rough substitute GIS if no GIS was originally available.

Note also that, while GIS/GPS will be used where available, and can greatly ease many of the tasks performed by ARGUS, we will none the less design all other functionality in such a way as to perform even without the help of a GIS/GPS, as would, e. g., be the case in many indoor environments (malls, train stations). GIS/GPS on the other hand will never be used without the additional use of computer vision (as is the case in, e. g., MoBIC), due to it's limited precision, which, for a blind person, can mean the difference between safety and a potentially life-threatening situation. Final object detection, finally, will often still remain the domain of the long cane or another primary mobility aid, as mentioned above.

The GIS in this project is an electronic (symbolic) map of a neighbourhood with streets and plots according to commercially available maps as well as maps from the land registry or military maps, extended with all important details. These include not only fixed obstacles like poles, signs and steps, but also objects like shops, bus stops, taxi stands, telephone booths, etc.

However, even without living beings, including traffic, a neighbourhood is not quite static. A street-name pole may disappear due to an accident. There may be a broken water-pipe in a sidewalk; its repair may block one side of the street or part of it and obliges everybody to pass to the other side, but where is the closest zebra? Hence, there is a necessity to regularly update the GIS map by "non-experts" including the blind because they detect new and annoying obstacles with the long cane and later use these in their navigation. It would be an interesting exercise to also use input from the computer-vision side to update the GIS. Some aspects of the work described above will build on similar work by others [35, 36].

Summary It can be seen from the above that the functionality of ARGUS, although limited to tasks well below the capabilities of human vision, requires contributions from a significant number of distinct computer vision subfields (lead contractor in parentheses):

- Segmentation of natural scenes (2)
- Texture analysis (3)
- Colour, illumination, photometry (2)
- Learning, experience-based vision (2)
- Binocular stereo (4, 1)
- Invariance theory (1)
- Geometric uncertainty management (1)
- 3D Structure recognition (1)
- Motion detection (2)
- Text localisation and OCR (3)
- Template matching, sign recognition (1)
- Search and decision strategies (2)

The large number of contributing subfields, most of them with their own methodological basis, is typical for real-life vision tasks. In order to obtain excellent and innovative solutions, it is necessary to bring together research teams whose combined expertise covers all aspects of the proposed task, and this has been done within ARGUS. The partner number behind each item (in parentheses) indicates the coverage within the proposed consortium, which for the software development alone combines the expertise of 4 university or private research-institutes from 3 different European countries, all of which have extensive experience from other European and Nationally funded research projects.

B.6.3 User Tests

User- and field-tests are an integral part of the proposed project. Only the involvement of users right from the beginning can guarantee the success of the project not only as an intellectual exercise, but also as a future, commercially viable, and competitive product; only the constant review and verification of the progressive computer vision algorithms employed within ARGUS under real-life conditions can guarantee the required robustness and flexibility. It is for this reason that one of our partners is an umbrella organisation of organisations for the blind which will provide us with detailed user requirements and accompany the development of ARGUS with questionnaires and field tests in at least 5 different European countries, namely France, Belgium, Italy, Spain, and Germany[5]. This will help considerably to increase the acceptance of a device like ARGUS within the blind community throughout Europe and should provide additional interest once a product stage will be reached (not within the scope of this project).

User tests will be made on a comparison basis where possible to determine the added advantages of ARGUS over other approaches. To facilitate this, the user groups will undertake a survey of existing navigation aids and see what state they have reached.

As ARGUS will operate in several different countries, it will be important to identify clearly the sentence structure for each "spoken phrase", which may well be different in different languages. Different message structures needs to be identified and tested. The required level of quality of the speech output under different ambient conditions is another parameter which needs to be identified.

[5]Additional project partners are located in Greece and Portugal, and we hope to enlist the cooperation of a British partner.

B.6.4 Hardware Requirements

The following objectives are of particular importance under the aspect of user-friendliness:

- low weight
- small dimensions
- pleasing look
- can be worn comfortably and securely
- robustness against mechanical strain
- robustness against varying lighting conditions
- easily usable
- reasonable battery-life.

In addition, the hardware for ARGUS needs the following I/O capabilities:

Input Obvious input channels for ARGUS include the capture of stereo-images, for which a special lightweight head-mounted stereo-camera will be built, as well as an electronic plummet (information about the cameras orientation relative to the ground-plane). The numerous functions intended for ARGUS additionally require the possibility of some interaction between ARGUS and the user. The foremost method of input will be a keypad on the pouch. We place large importance on the user-friendliness of all interactions with ARGUS (with an eye towards the usability of ARGUS by the elderly), as well as cosmetic aspects connected with the device. Alternative methods of interaction are possible for later development steps (outside the scope of this proposal), speech-recognition being one of the more obvious possibilities.

Output Output is generated by an open headphone. Using an open headphone means that any ambient noise can still be heard and reacted upon by the user. For the same reason the audible output itself will be reduced to the bare minimum, usually simply naming distance and direction to selected objects or landmarks. User tests as mentioned above will provide detailed information about the exact nature of the audible feedback for different tasks and in varying environments.

In normal operation the user will select only one particular functionality — or at most a small set of functions — at any given time. This guarantees a selective output which will not overburden the user's remaining senses.

Design

Figure 5: *A spectacle-like night-vision system for persons with night-blindness, developed by partner 5.*

As can be seen from the above list, we need to place special importance on the design of the visible part of the hardware to ensure user-acceptance. The extensive experience of partner 5 in the development and production of optoelectronic spectacle-like devices for visually impaired persons has shown that, in addition to the actual function provided, particular importance has to be placed on low weight, minimum dimensions, and an attractive yet unobtrusive design. The user does not want to be identifiable as visually impaired by the use of such a device, and this is one of the reasons why we are aiming to integrate the cameras into a pair of glasses. User groups, represented by partner 8, will play an important role in final design decisions.

Summary It is clear from the above requirements that the use of custom hardware will be essential for a project like ARGUS, and its development has therefore been included in the workpackages. All in all, 3 different companies from 2 European countries will be responsible for the development of the hardware and the integration with the software mentioned above, and it is these companies, which all have close connections with their country's organisation for the blind and the blind community, that will play an important role in the dissemination of the results and the possible further development into a commercially viable product (outside the scope of this proposal). The early availability of a first prototype (Deliverable D 2) is of particular importance to facilitate the development of functionality under realistic conditions.

All in all, the ARGUS consortium is made up of 8 partners from 5 European countries, among them 3 universities, one non-profit organisation, 3 companies and one user-organisation.

B.6.5 Exploitation and Dissemination

As ARGUS is not expected to reach a product stage during the time-span covered by this proposal, most of the exploitation and dissemination efforts will, at least initially, by geared towards dissemination within the scientific society through the usual channels, in particular through the creation and continuous update of a web-site, preferably with its own domain `www.argus.*` and ftp-site `ftp.argus.*` both for the distribution among the individual partners of the consortium (containing e. g. information regarding workshops, seminars, conferences, call for papers, but also the source of the actual reports and algorithms), as well as to a wider public; presentation of ARGUS at congresses, symposia, conferences, technical events, etc. in addition to the individual partners' contributions to scientific conferences.

However, it is felt that in order to improve the chances of a possible future product derived from ARGUS it will be important to inform the public in general and the blind community in particular about the efforts being undertaken, and this will be done mainly through contacts with the press, and in particular reports in local, national and international newspapers and on radio and television. Experience from a similar previous project has shown that the press is usually quite interested in this kind of work. In addition, we will use partner 8's established communication links within the blind community to inform all parties concerned in even more detail.

In addition, work will be devoted to the production of an adequate and realistic business and exploitation plan for future commercialisation of the project results, including the identification of potential markets and potential competitors. Although it is neither planned nor expected that ARGUS will reach product stage over the next three years, it is still possible that individual functionalities could be marketed on a stand-alone basis. In addition, the exploitation plan mentioned above will, among other things, deal with the question of who will continue developing ARGUS once funding from the EC has stopped (all 3 companies involved have indicated their interest in continued development and manufacture of a finished product), as well as questions such as who could provide updates for the GIS employed (bringing up additional questions of standardisation and interface definitions) and service or backup for the entire product (including localisation for the individual EC member states). Also included in such an exploitation plan should be aspects such as awareness packs in different languages, video demonstration packs, and training packages, all of which would need to be developed in close cooperation with user groups, represented by partner 8.

B.6.6 Structure of the project

From the above outline of the intended functionality, we can see that the entire project ARGUS separates into three logical parts, namely software-development, hardware development, and interaction with the users, comprising both the determination of user-requirements at the beginning of the project as well as user tests during the last year of the project. This leads to a structure as can be seen in Figure 6 on Page 17, where the user interaction can be said to act as a bracket around the two concurrently implemented blocks, software-integration on the one side and hardware-integration on the other. The concurrency of these two blocks should however not be viewed as a logical and organisational separation; rather the opposite is the case: both blocks are intimately interwoven and build onto each other, where hardware specifications are based on the requirements of the software side, while on the other hand the availability of particular hardware is of paramount influence in the design of the algorithms.

Following the essentially vertical structure of the two main development blocks, we have mostly designed the individual workpackages to follow the same vertical layout, that is each of the workpackages WP 1–WP 9 comprises all the work necessary to take a certain task from the specification stage to the integration stage. Only at this stage does the usual horizontal structure set in, and with the workpackages WP 10–WP 12 we have three workpackages which, in this order, serve to integrate the different software components, integrate the software with the hardware, and perform user tests and final adjustments. The workpackage WP 13 finally is concerned with dissemination and exploitation and spans the entire duration of the project. We think that this structure has several advantages over the more traditional horizontal structure, the most important ones being ease of management, avoidance of bottlenecks and subsequently slipping milestones, and increased robustness of the results achieved. This will be discussed below in more detail.

Ease of management The development of an algorithm is often not easily parallelised, but is best assigned to only one or at most two researchers. Added manpower will not necessarily result in faster conclusion of the project, but will simply increase the overhead and the amount of communication necessary between the different co-workers. This is already true for projects that are centrally managed and executed; it is all the more true for a project like ARGUS, where researchers from several European countries, with different professional and cultural backgrounds, cooperate on a single project. We think that, instead of trying to overcome these problems, the EC's limited resources are better spent on the work at hand.

This does not mean that our vertical model does not provide for the exchange of ideas and cooperation among the partners. The workpackages WP 10–WP 12 are expressively designed in such a way as to foster the exchange and integration of results. In addition, we will organise 6-monthly meetings among all participants, which are meant to do exactly that — foster the exchange of ideas. Half of these meetings will also be organised as workshops, enlisting the participation of outside researchers, which will establish the active exchange of ideas even beyond the partnership. These (restricted) meetings also serve to evaluate the work done, to synchronise the participants, to discuss problems and improvements, and to prepare the annual reports as well as the work planned for the next six months.

In addition, partners that are working on related workpackages will normally have set aside a certain amount of time for active cooperation with and participation in other workpackages, compare Table 1. Cooperation in this context will reach from the simple discussion of similar problems and a comparison of approaches used and results obtained, to joined publications and even the exchange of researchers.

Avoidance of bottlenecks In the classical horizontal structure, individual workpackages build on top of each other. Such a structure is inherently susceptible to bottlenecks, since as a rule some or all functionality in one layer will depend on some or all of the functionalities in earlier layers. As soon as one milestone slips, or one deadline is missed, the entire project is jeopardised.

In the vertical model used here, on the other hand, only very few hard dependencies exist (the main exception is the availability of certain hardware-components at certain times). So even if a milestone is missed in any of the workpackages, this will usually only pertain to this individual workpackage, jeopardising only one individual component of the overall system rather than the system itself.

		Partner							
		1	2	3	4	5	6	7	8	Sum
WP 1	(Obstacles)	1		1	30			2	1	35
WP 2	(Text)	1	1	24				2	1	29
WP 3	(Signs)	24	1	1				2	1	29
WP 4	(Objects)	24	1	1	2			2	1	31
WP 5	(Scenes)	1	18		2			2	1	24
WP 6	(Landmarks)		24					2	1	27
WP 7	(GIS / GPS)			24		1	1		2	28
WP 8	(Cameras)					33	3		2	38
WP 9	(Hardware)					3	45		2	50
WP 10	(Software)					1	12	24		37
WP 11	(User tests)								24	24
WP 12	(Integration)	12	12	18	12	7	8	6		75
WP 13	(Dissemination)	3						3	2	8
	(Management)	18	3	3	2	2	3	3	2	36
Sum		84	60	72	48	47	72	48	40	471

Table 1: *Allocation of man-months by workpackage (WP) and partner.*

Robustness of results One of the main requirements for the results of a project such as ARGUS is robustness. Many algorithms perform safety-critical tasks, and ARGUS will be a success only if the blind user can completely depend on it. And even though no product-stage will be reached, and in user-tests ARGUS will only be used in a controlled environment, where no real danger for the user exists, we do none-the-less feel that robustness is easily the single most important feature to strive for.

How a vertical structure can improve robustness will be made clear on the (admittedly rather extreme) example of colour segmentation. All workpackages WP 1–WP 6 use colour segmentation to some extent — the workpackages WP 1, WP 3, WP 4 and WP 6 all depend on colour-information, while the remaining workpackages, WP 2 and WP 5, make use of colour-information to simplify their task. However, the actual requirements differ sharply from workpackage to workpackage. Where in WP 3 colour-information is used to identify regions of a particular, well-defined colour with very little variance, it is used in WP 1 to identify regions of equal colour, but with an a-priori only approximately known value and distribution.

Rather than form a workpackage which will develop an algorithm that tries to be all things to all people, we have deliberately decided that each group selects its own colour-segmentation algorithm — although in close cooperation where appropriate. This way, each group can work with the algorithm it considers best for its particular application, while at the same time comparing its performance to the other groups' algorithms (this of course requires a common interface specification, which will be provided as part of workpackage WP 10). We believe that this process will ultimately lead to an algorithm which is superior to each individual groups approach, while a horizontal structure would stifle creativity and positive inter-project competition.

B.6.7 Graphical Representation of the Project Components

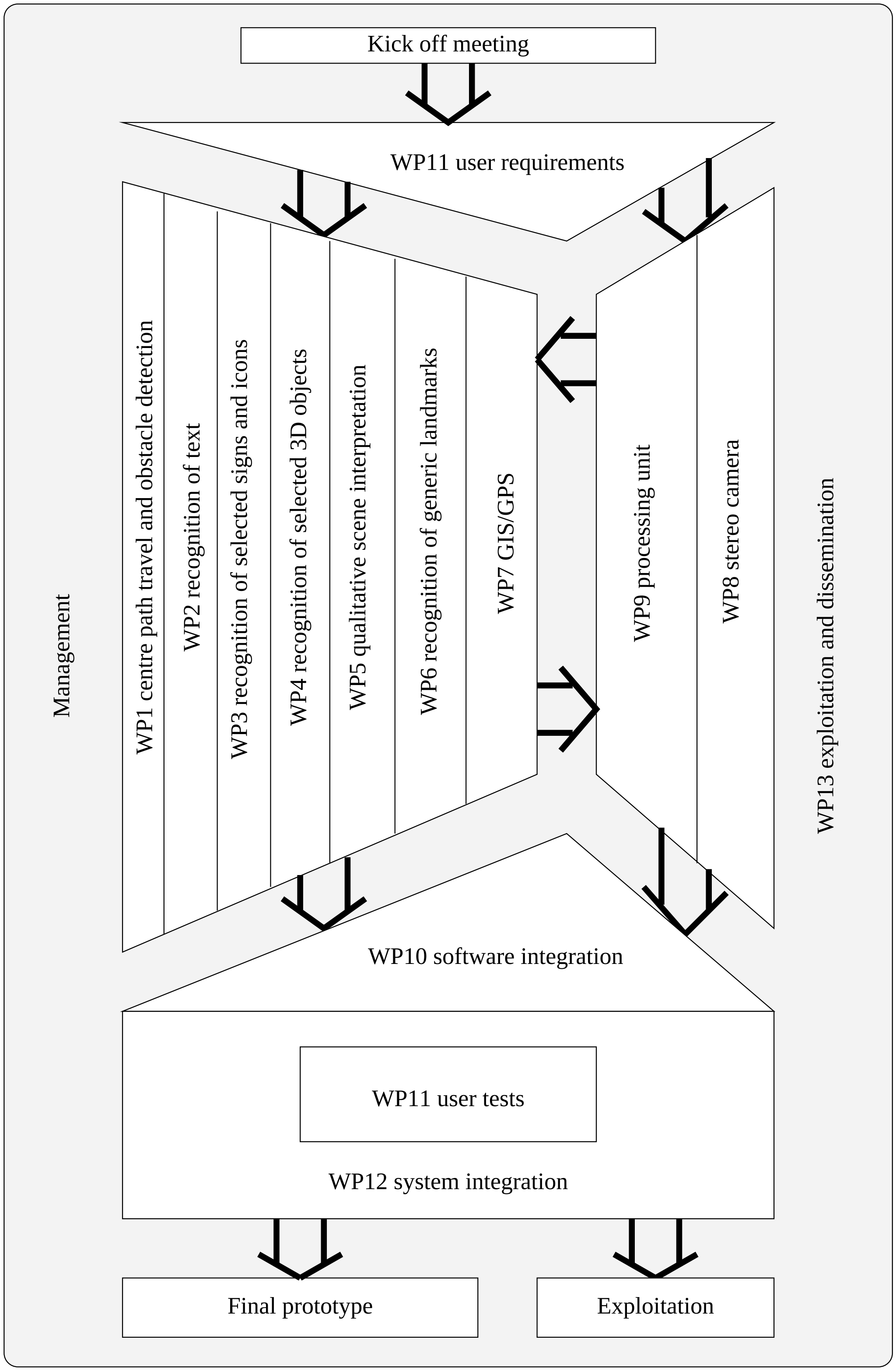


Figure 6: *Graphic representation of the project structure*

B.6.8 Project Planning and Timetable

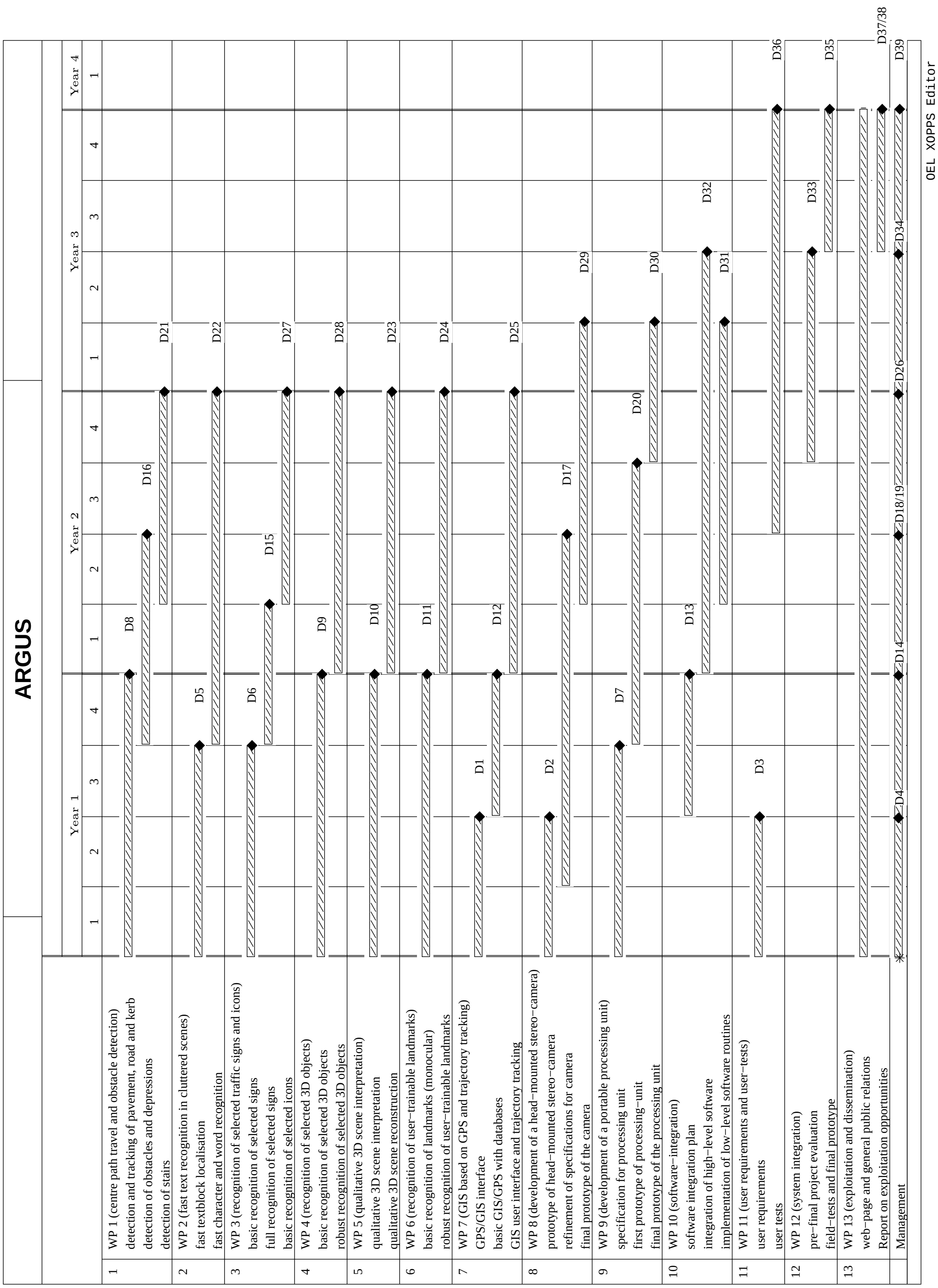


Figure 7: *Gantt chart of workpackages and tasks. This chart shows the target dates and deliverables (milestones ⧫).*

B.6.9 Detailed Project Description

B1.	Workpackage list

Work-package No	Workpackage title	Lead contr. No	Person-months	Start month	End month	Phase	Deliv-erable No
WP 1	Centre path travel and obstacle detection	4	35	0	24	–	D 8, D 16, D 21
WP 2	Fast text recognition in cluttered scenes	3	29	0	24	–	D 5, D 22
WP 3	Recognition of selected traffic signs and icons	1	29	0	27	–	D 6, D 15, D 27
WP 4	Recognition of selected 3D-objects	1	31	0	27	–	D 9, D 28
WP 5	Qualitative 3D scene interpretation	2	24	0	24	–	D 10, D 23
WP 6	Recognition of user-trainable landmarks	2	27	0	24	–	D 11, D 24
WP 7	GIS based on GPS and trajectory tracking	3	28	0	24	–	D 1, D 12, D 25
WP 8	Development of a head-mounted stereo-camera	5	38	0	27	–	D 2, D 17, D 29
WP 9	Development of a portable processing unit	6	50	0	27	–	D 7, D 20, D 30
WP 10	Software-integration	7	37	6	30	–	D 13, D 31, D 32
WP 11	User requirements and user-tests	8	24	0	36	–	D 3, D 36
WP 12	System integration	1	75	21	36	–	D 33, D 35
WP 13	Exploitation and dissemination	1	8	0	36	–	D 37, D 38, D 40
	Management	1	36	0	36	–	D 4, D 14, D 18, D 19, D 26, D 34, D 39
	Total		471				

B2. Deliverables list

Deliv. No	Deliverable title	Deliv. date	Nature	Dissem. level
D 1	GPS / GIS Interface	6	P	PU
D 2	First prototype of head-mounted stereo-camera	6	P	CO
D 3	Report on user-requirements	6	R	PU
D 4	First management report	6	R	CO
D 5	Text block detection software	9	RP	PU
D 6	Basic recognition of selected traffic-signs	9	P	PU
D 7	Specification for the processing-unit	9	R	CO
D 8	Detection and tracking of pavement, road and kerb.	12	RP	PU
D 9	Basic recognition of selected 3D-objects	12	RP	PU
D 10	Qualitative 3D scene interpretation (monocular)	12	RP	PU
D 11	Basic recognition of landmarks (monocular)	12	RP	PU
D 12	Basic GIS plus databases system	12	RP	CO
D 13	Software integration plan	12	R	CO
D 14	Second management report	12	R	CO
D 15	Recognition of the full set of selected traffic signs	15	RP	PU
D 16	Detection of obstacles and depressions	18	RP	PU
D 17	Full specification for the head-mounted stereo-camera	18	R	CO
D 18	Mid-project evaluation	18	RP	CO
D 19	Third management report	18	R	CO
D 20	First prototype of the processing unit	21	RP	CO
D 21	Detection of stairs	24	RP	PU
D 22	Word recognition software	24	RP	PU
D 23	Qualitative 3D scene reconstruction	24	RP	PU
D 24	Full recognition of landmarks (3D-based)	24	RP	PU
D 25	GIS user interface and trajectory tracking	24	RP	CO
D 26	Fourth management report	24	R	CO
D 27	Recognition of the full set of selected icons	24	RP	PU
D 28	Full recognition of selected 3D-objects	24	RP	PU
D 29	Final prototype of the camera	27	RP	CO
D 30	Final prototype of the processing unit	27	RP	CO
D 31	Implementation of low-level software routines	27	RP	CO
D 32	Software integration	30	RP	CO
D 33	Pre-final project evaluation	30	P	CO
D 34	Fifth management report	30	R	CO
D 35	Fully functional, optimised and field tested prototype	36	P	CO
D 36	Report on user-tests	36	R	PU
D 37	Dissemination report	36	R	CO
D 38	Exploitation plan	36	R	CO
D 39	Final report	36	R	CO
D 40	Web-page and general public relations	–	O	PU

B3 Workpackage description

Workpackage number: WP 1, Centre path travel and obstacle detection
Start date or starting event: 0
Participant number: 4; 1, 3, 7, 8
Person-months per participant: 35: 30 (4), 1 (1), 1 (3), 2 (7), 1 (8)

Objectives: To develop and implement a fast, stereo-based algorithm for centre path travel and the detection of small obstacles (e. g. bricks, bottles) that cannot easily be detected using the long cane. As intermediate results, the ground plane orientation with respect to the camera system will be estimated and the location of the kerb will be detected and tracked. In addition, depressions, pits and stairs (e. g. subway stairs) on the pavement will be detected.

Description of work: The pavement and road will be identified in both images based on texture segmentation. A training procedure will be used to optimise the parameters of the texture segmentation procedure using sample images from the test environment. Then, a set of feature points will be extracted from the region of the pavement and correspondences will be found. Using triangulation, the 3D coordinates of the corresponding points will be recovered and the 3D location of the ground plane will be tracked while the person is walking. Texture segmentation and stereo matching will also be used to detect small obstacles or depressions (connected regions with different texture), kerbs and stairs (set of coincident lines with given crossratio), where the stereo-matching will be used to verify the initial assumption, and to calculate their 3D location.

Deliverables: Deliverable D 8 is a set of subroutines, including a documentation report, which are used to identify the pavement and road, and detect and track the 3D location of the ground plane and the kerb. Deliverable D 16 consists of a set of subroutines, including a documentation report, which use results from D 8 to detect obstacles and depressions within the region of the pavement. Deliverable D 21 will consist of the software to detect stairs originating from the ground plane and determine whether they are ascending or descending, as well as a corresponding report. It will be based on results from D 8.

Milestones and expected result: Colour and texture identification algorithms for the detection of the pavement and road will be developed during the first 3 months, while the implementation and optimisation of the algorithm used to detect and track the ground plane from a set of feature points will be completed over the next 6 months. During the final 3 months of the first year, the kerb detection algorithm will be developed (month 12: D 8). The development of the obstacle and depression detection algorithms can start after month 9 and will require another 9 months (month 18: D 16). The development of the stair detection algorithm and the associated parameter estimation algorithms will take another 9 months within the second year (month 24: D 21).

B3	Workpackage description

Workpackage number: WP 2, Fast text recognition in cluttered scenes
Start date or starting event: 0
Participant number: 3; 1, 2, 7, 8
Person-months per participant: 29: 24 (3), 1 (1), 2 (7), 1 (2), 1 (8)

Objectives: To develop and implement a very fast algorithm for the detection of words on rectangular backgrounds with perspective projections and to correct for the deformation; to develop and implement a very fast algorithm for character and word recognition, including a small dictionary to solve ambiguities.

Description of work: By detecting vertices by means of a model of cortical end-stopped cells and edges by means of e. g. the Canny edge detector, in combination with texture measures that are optimised for characters in words (normally vertical strokes etc.), text blocks and their deformed background can be detected and the projective distortion can be calculated. Characters and words can then be detected within the rectified pixel-block by successive application of the following algorithms (to be developed): (1) correction of contrast and other illumination effects, (2) adaptive thresholding, (3) thinning or skeletonisation, (4) construction of a list of most likely characters and (5) word extraction. To solve ambiguity problems in (4) and (5) a manageable word list needs to be developed with the blind. All algorithms have to be highly optimised.

Deliverables: Deliverable D 5 is a set of subroutines, including a documentation report, which receive GIS input about expected positions and heights of texts and which produce a (set of) corrected rectangular pixel block(s) which can be used as input for the actual character and word recognition. Deliverable D 22 is a set of subroutines, including a documentation report, which, with input from D 5, produce a list of output strings.

Milestones and expected result: Starting with publicly available software, we can concentrate on complex images and complex geometries right from the beginning. However, in view of the non-triviality of the segmentation-problem a first version for the detection of text-blocks will take at least 6 person-months and its CPU-time optimisation 3 months (month 9: D 5 tested and optimised). The selection of the most stable algorithm for each link of the remaining chain — and its subsequent optimisation — will require a considerable amount of work (month 24: D 22 ready and tested with complex still images). We expect to be able to detect all texts in still images with limited slant / tilt angles, but field tests during the third year (WP 10) and using the actual system are required to assess the performance in practice.

B3	Workpackage description

Workpackage number: WP 3, Recognition of selected traffic signs and icons
Start date or starting event: 0
Participant number: 1; 2, 3, 7, 8
Person-months per participant: 29: 24 (1), 1 (2), 1 (3), 2 (7), 1 (8)

Objectives: To develop and implement a fast and robust algorithm for the detection and classification of a number of selected traffic signs and icons as specified in ISO 7001 [15]. This includes algorithms for the colour-based detection of Regions of Interest (ROI), contour detection, calculation of the projective distortion (and subsequent rectification), as well as the actual algorithm used for the recognition and verification.

Description of work: Traffic signs can be recognised based on their distinctive colour, and we will therefore employ a two-step scheme which first identifies ROIs and only then performs the actual recognition independently on each ROI. The main problem here is to devise an algorithm capable of working with the comparatively low-resolution 1-chip CCD-cameras which will be used within ARGUS, taking into account the different responses of different pixels depending e. g. on their position. Icons as specified in ISO 7001 [15] on the other side are usually simply black and white and as such much harder to segment. Here an algorithm will be used based on the rectangular outline of the icon, compare WP 2. Segmentation of Regions of Interest is in both cases followed by the actual recognition and verification based on invariant geometric features.

Deliverables: Deliverables include a basic algorithm for the recognition of selected signs (D 6), which already contains all the main parts necessary for a successful recognition procedure, but has not yet been optimised for speed, nor does it include all selected signs. Both the optimisation as well as the extension to different signs will be part of deliverable D 15. Deliverable D 27 finally is an extension of the deliverables D 5 and D 6 to the recognition of icons. Both deliverable D 15 as well as deliverable D 27 will consist of the actual program and an accompanying report.

Milestones and expected result: Much of the work on colour segmentation as well as the recognition and verification of traffic signs can build on existing work done by partner 1, leaving the adaption to the camera used within ARGUS as the main task. We estimate 3 person-months for a first algorithm and another 6 person-months for the adaption of the algorithm (month 9: D 6). Adding additional signs to this algorithm and optimisation should not take more than an additional 6 person-months (month 15: D 15). The recognition of icons is in comparison much more difficult, mostly due to the lack of colour cues, but we assume that based on the work in D 5 and D 6 it should not take more than 6 person-months for segmentation and, using the results from D 15, 3 person-months for the actual recognition (month 24: D 27).

B3 Workpackage description

Workpackage number: WP 4, Recognition of selected 3D-objects
Start date or starting event: 0
Participant number: 1; 2, 3, 4, 7, 8
Person-months per participant: 31: 24 (1), 1 (2), 1 (3), 2 (4), 2 (7), 1 (8)

Objectives: To develop and implement a fast and robust stereo-based algorithm for the detection and classification of a number of selected 3D-objects and classes of objects. This will include colour segmentation, the extension of a line-based stereo-algorithm, the design of an adequate database and robust verification.

Description of work: Many objects of the public infrastructure can be identified by a combination of their 3D-form and colour-information. The combination of a multilevel-approach for the evaluation of the depth of selected lines (rather than computing an entire depth-map, which is time-consuming), fuzzy rules for colour segmentation, and the use of geometric hashing will allow us the construction of an algorithm which is both robust and time-efficient. This algorithm needs to be complemented by a reliable verification procedure in order to weed out unwanted false-positives. Different approaches based on geometric constraints and texture analysis will be implemented in order to identify the most reliable algorithm.

Deliverables: The deliverable D 9 includes the basic algorithm for the recognition of three-dimensional objects, combining the efficient estimation of depth-information by a multilevel-approach, fuzzy rules for colour segmentation and the use of geometric hashing. Deliverable D 28 will add to this additional objects as well as a reliable verification procedure. Both are supplemented by one report each describing the work done and results obtained.

Milestones and expected result: Much of the work proposed can be based on a long history of similar work done by partner 1 [23–25]. It is expected that the basic algorithm will require approximately 12 person-months — 3 person-months each for stereo, colour-segmentation, and the geometric hashing, and another 3 person-months for integration and optimisation (month 12: D 9). The extension of the object-database and in particular the verification and final optimisation are expected to take up another 12 person-month (where the main part will be allocated for the verification), with a second milestone after month 24: D 28.

B3 Workpackage description

Workpackage number: WP 5, Qualitative 3D scene interpretation
Start date or starting event: 0
Participant number: 2; 1, 4, 7, 8
Person-months per participant: 24: 18 (2), 1 (1), 2 (4), 2 (7), 1 (8)

Objectives: To develop and implement a fast and robust algorithm for the qualitative interpretation of 3D-scenes based on geometric constraints, but also using cues from colour- and texture-segmentation as well as graph-theory and traditional AI.

Description of work: It is well known that from a single image it is not possible to make any reliable statements about the structure of the underlying scene. The use of stereo alleviates this problem to a certain extent, but is slow and burdened with problems when dealing with textureless surfaces. This workpackage will instead focus on the generation of a qualitative description of a scene based on a single image of the scene, but aided by additional knowledge about the general structure of the scene, geometric constraints, and possibly some selected stereo-cues, aiming for results which are comparable to that of humans under similar circumstances (at a coarse level). In order to do so, it will combine approaches from different fields of computer vision, including projective geometry, analysis of texture and colour, graph-theory and traditional AI.

Deliverables: Deliverable D 10 is an algorithm for the qualitative interpretation of a scene from a monocular image, specifying relations such as "in front of" as well as "horizontal" or "vertical". Based on this, deliverable D 23 will be an algorithm providing a qualitative reconstruction of usual street scenes, similar to the rough sketch a human could provide when shown the image. Both are complemented by a report describing the work done.

Milestones and expected result: The work proposed within this workpackage requires the combination of knowledge from different fields of computer vision and image understanding. Work will initially start with the projective-geometry constraints and by-and-by add other features. It is expected that a first qualitative interpretation of general street scenes will be possible after at least 9 person-months distributed over 12 months (month 12: D 10), and the optimisation and extension to an algorithm for the qualitative reconstruction after another 9 person-months, again distributed over 12 months (month 24: D 23).

B3 Workpackage description

Workpackage number: WP 6, Recognition of user-trainable landmarks
Start date or starting event: 0
Participant number: 2; 7, 8
Person-months per participant: 27: 24 (2), 2 (7), 1 (8)

Objectives: To develop the landmark recognition functionality of ARGUS. Roughly, landmarks are camera views which a blind user can store using ARGUS and which can be recognised later to aid his navigation, e. g. the entrance of a particular shop or a particular street corner. To store a new landmark, the user points ARGUS to obtain the desired view, possibly with assistance, and commands ARGUS to make it a landmark. A stored landmark may be activated by the user with the effect that ARGUS continuously checks whether the current view corresponds to that landmark. If this is the case, an appropriate sound or speech signal is generated.

Description of work: Landmark recognition will be realized by matching relational descriptions based on prominent edge and region features. Basic concepts have been developed during a similar project [33]. The full functionality will be achieved in three steps. In the first step, basic landmark recognition will be developed for varying lighting conditions and viewpoints, with landmarks generated from single monocular views. In step 2, landmark recognition will be extended to exploit stereo information and multiple views. In step 3 the methods will be refined based on feedback from user tests and integration requirements.

Deliverables: Deliverable D 11 is a prototype for basic landmark recognition under variable lighting conditions and from variable viewpoints, including an experimental demonstration of landmark recognition under changing conditions (expressed in location and orientation deviations with respect to the stored landmark and qualitative differences of lighting conditions), as well as a report. Deliverable D 24 extends this prototype through the inclusion of stereo information and multiple views. It consists of the algorithms for an experimental demonstration and a report. The fully functional prototype for landmark recognition with modifications and refinements based on user feedback and integration requirements, and a corresponding report will be part of deliverables D 35 and D 39.

Milestones and expected result: Deliverable D 11 requires new research about the attribute space of real-life object features induced by changing illumination and viewpoints. Overall work is expected to last 12 person-months (month 12: D 11).
Deliverable D 24 comprises the development of new methods for learning landmarks from multiple views. The research results of this workpackage are expected to be also relevant for other recognition tasks under real-life conditions. Overall work is expected to last 12 person-months (month 12: D 24).

B3 Workpackage description

Workpackage number: WP 7, GIS based on GPS and trajectory tracking
Start date or starting event: 0
Participant number: 3; 5, 6, 8
Person-months per participant: 28: 24 (3), 1 (5), 1 (6), 2 (8)

Objectives: To implement a fast GIS with differential-GPS and image-analysis interfaces, together with a complete neighbourhood database; to develop and include into the GIS a graphical user interface for interactive updates of the GIS; to include trajectory tracking and extrapolation into the GIS; to perform first field trials in order to test the image-analysis interface in practice.

Description of work: The design and implementation of a fast GPS-based GIS including three data bases: (a) streets and buildings, including (multiple) entrance; (b) visual landmarks (including traffic signs, texts like TAXI, BUS, BANK, big shop names, etc); (c) known obstacles. The GIS will directly interface to the image analysis components (WP 1–WP 6). In the first version, developed by creating an artificial neighbourhood, this interface will already include input from the image analysis to confirm (or reject) e. g. recognised icons or text. The GIS will be extended by addition of an easy-to-use interface to create and update a neighbourhood database, plus an extension of the actual GPS coordinates with the path history and extrapolation. This work will be supplemented by real field tests.

Deliverables: Deliverable D 1 will be a subroutine that sends a request to the GPS and that receives the longitude-latitude data and converts these to the coded format used in the GIS. Deliverable D 12 will be the first GIS implementation (program and documentation) with measurements of its efficiency and processing speed. Deliverable D 25 will be the same system, but extended with a user interface and trajectory tracking, and consisting of the complete software package, description and field-test results. The latter serve for developing a strategy for final improvements and fine-tuning during the third project year (WP 10).

Milestones and expected result: Using existing libraries, we estimate: 3 months for the design of the overall strategy and data formats, GPS and image-analysis interfaces; 3 months implementation of the basic GIS structure using simple databases (month 6: D 1, first limited running version); 6 months optimisation, using complex artificial-neighbourhood data (month 12: D 12, the first version running with GPS input and complex data according to the specifications given, working within a neighbourhood of 1 km^2). Based on existing libraries, it is expected to write the GIS user interface with 9 person-months. The integration of trajectory tracking by means of a sequential list with previously encountered landmarks etc. can be realised easily (month 24: D 25, GIS tracking).

B3 Workpackage description

Workpackage number: WP 8, Development of a head-mounted stereo-camera
Start date or starting event: 0
Participant number: 5; 6, 8
Person-months per participant: 38: 33 (5), 3 (6), 2 (8)

Objectives: To develop and produce a small number of small and lightweight head-mounted stereo-cameras, including the camera-electronic and battery-based energy-supply. A first set of cameras will be used by partners 1–4 as basis for the development of their algorithms, while a set of 8 final stereo-camera setups will be used during field-tests.

Description of work: The aim of this workpackage is the development and construction of the head-mounted stereo-camera which will be used in ARGUS. Two different sets of cameras will be needed; an initial set of no more than 4 stereo-camera setups, with an interface to a standard notebook or portable computer, will by used by the partners 1–4 for their initial work. Based on feedback from the partners a second setup will be developed, which will interface to the processing unit developed in workpackage WP 9 and will be used in the final evaluation and user-tests. Both setups will include an electronic plummet, and particular importance is placed on an unobtrusive and lightweight design. The work includes extensive market research and the development of an adequate battery-system.

Deliverables: Deliverables include up to four prototypes of the stereo-camera setup (D 2), the refinement on the specifications for the camera based on input from the partners and work performed in workpackage WP 9 (D 17), and finally deliverable D 29, a set of 8 prototypes plus documentation which will integrate directly with deliverable D 30 from workpackage WP 9.

Milestones and expected result: The design of the first prototype, deliverable D 2, will be directly based on previous work in similar projects; we estimate 4 person-months for the development of the setup, and another 2 person-months for the construction (month 6: D 2). The design will then continuously be refined based on input both from the partners 1–4 as well as from workpackage WP 9, including extensive market research. This will require approximately 15 person-months and will be done overlapping the construction of the first and second prototype (month 18: D 17). An additional 12 person-months finally are reserved for final work on the second prototype (month 27: D 29).

B3	Workpackage description

Workpackage number: WP 9, Development of a portable processing unit
Start date or starting event: 0
Participant number: 6; 5, 8
Person-months per participant: 50: 45 (6), 3 (5), 2 (8)

Objectives: To develop and produce a small number of small and lightweight portable processing units appropriate for field-tests and general use within ARGUS.
The processing unit will include interfaces to: stereo video input; stereo audio output; GPS input; keypad and serial communication port. An ergonomic and light enclosure for the processing unit, camera electronic and batteries will be designed. This case will be hand-made in plastic.

Description of work: A detailed analysis of microprocessors performance (general processors and signal processors) will be carried out to choose the most promising architecture, both in terms of actual performance as well as technological projection. Since the final result of this project will not be a production system but an advanced prototype, power autonomy will be sacrificed to a certain degree in favour of computing power. The second main aspect in the design of the overall system is that of mass storage, which needs to be light-weight, low-volume and robust. Different alternatives for mass storage will be studied, paying special attention to FLASH memory technology, which promises the highest reliability, high density and affordable prices due to its popularisation in consumer electronics (e. g. digital cameras).

Deliverables: Deliverable D 7 contains the specifications for the processing-unit. It will include both a detailed specification of the processing unit in terms of computing power, memory, I/O interfaces, user interfaces and power supply requirements, as well as a Processing Unit Architecture Design Document. The next deliverable (D 20) is a first prototype of the processing unit, including the Processing Unit Detailed Design Document, electrical schematics, and a first Printed Circuit Board (PCB) assembled and debugged. Deliverable D 30 is the final prototype of the processing unit, including a Low Level Programmers Guide, a Processing Unit Testing Guide, the final PCB, and 8 processing units fully operative, enclosed in hand-made plastic cases.

Milestones and expected result: The first 9 months of the project will produce the specifications of the processing unit in close cooperation with the partners (month 9: D 7). During the next 12 month 2 engineers and 1 technician will spend 27 person-months on the detailed design (4 months), the schematic capture (3 months), PCB design (3 months) and assembly and test of the first prototype (2 months), milestone reached at month 21: D 20. Over the next 6 months, 2 engineers and 1 technician will spend a total of 9 person-months on debugging (1.5 months), validation (1.5 months) and the production of 8 prototypes for user tests (3 months). The case of the processing unit will be subcontracted during this period (month 27: D 30).

B3	Workpackage description

Workpackage number: WP 10, Software-integration
Start date or starting event: 6
Participant number: 7; 6, 5
Person-months per participant: 37: 24 (7), 12 (6), 1 (5)

Objectives: The integration of the algorithms developed within the workpackages WP 1–WP 7 and their implementation on the portable processing unit developed within WP 9.

Description of work: The integration of software from different partners requires the establishment of an integration plan (D 13) at an early stage. This integration plan will define how the various internal software components (algorithms) of the global ARGUS system are going to be integrated and implemented on the portable processing unit in order to produce the first prototype version. In particular, methods and basic software tools needed for the integration and implementation on the portable processing unit will be defined. The final integration will then commence as soon as the first results from WP 1–WP 7 are available, and will closely follow the integration plan described above.

Deliverables: The deliverables defined for this workpackage consist of the Software Integration Plan (deliverable D 13), the implementation of the low-level functionality and user-manual in deliverable D 31, and the actual integration of the software plus report in deliverable D 32.

Milestones and expected result: The development of the Software Integration Plan will start after month 6. It is expected to take 6 person-months (month 12: D 13). This is followed by a continuous software integration process, starting in month 13 and gradually increasing in intensity. Software integration will require 12 person-months for hardware-specific low-level routines (month 27: D 31) and 18 person-months for the integration of the high-level functionality (month 30: D 32).

B3	Workpackage description

Workpackage number: WP 11, User requirements and user-tests
Start date or starting event: 21
Participant number: 8; 5
Person-months per participant: 24

Objectives: To produce a detailed report about user requirements in connection with ARGUS, and to perform extensive user-tests once the final prototype is available. This includes the development of an adequate questionnaire and the execution of the interviews and user-tests in at least the following countries: France, Belgium, Italy, Spain, and Germany. Based on these tests, final modifications to the hard and software will be implemented in order to integrate the results of the work packages WP 1–WP 10 into a homogeneous and fully functioning system, compare WP 12.

Description of work: In order to facilitate easy tests, user groups will be set up in the above countries. Questionnaires will be designed on the blind users' requirements on mobility in French, and will be translated into Spanish, Italian, German and English. Based on these questionnaires, interviews will be held and the results compiled into a report.
In addition, user-tests will be performed in the above countries. Each user group in each country will be composed of at least 12 users, 6 legally blind and 6 visually impaired; within each group 2 users will be under the age of 18, two around 35, and two over 60. A special test-courses will be created or suitable locations be selected to perform tests with regard to size, illumination, distance and other scene attributes.

Deliverables: Deliverable D 3 is a report on user-requirements collected from blind users across at least five different European countries: France, Belgium, Italy, Spain, and Germany. Deliverable D 36 is a report on user tests performed in the same countries.

Milestones and expected result: The report on user-requirements will be generated during the first 6 months of the project (month 6: D 3). User tests will be performed during the second half of the project, mainly during the last 9 months, and will require 18 person-months from partner 8 (month 36: D 36).

B3 Workpackage description

Workpackage number: WP 12, System integration
Start date or starting event: 21
Participant number: 1; 2, 3, 4, 5, 6, 7
Person-months per participant: 75: 12 (1), 12 (2), 18 (3), 12 (4), 7 (5), 8 (6), 6 (7)

Objectives: To implement final modifications to the hard and software in order to integrate the results of the workpackages WP 1–WP 10 into a homogeneous and fully functional system. The work will in particular be based on the results of extensive field-tests performed within workpackage WP 11.

Description of work: This workpackage is related to the workpackage WP 11, which will help to test and evaluate ARGUS in a real environment. On the basis of the results of the field-tests undertaken, final modifications to the hardware and software components will be implemented, in particular refinements and adaptations to real life requirements and as a result of the feedback received from the users, integrating the results of the workpackages WP 1–WP 10 into a homogeneous and fully functional system.

Deliverables: Deliverable D 33 is a report as the result of the pre-final project evaluation, describing the system at that time (month 30) and highlighting recommended improvements. Deliverable D 35 is a fully functional, optimised and field tested prototype of ARGUS.

Milestones and expected result: Workpackage WP 12 is by far the biggest and most important workpackage, as the close cooperation of all partners will be required during these last 15 months of the project, with two important milestones at month 30: D 33, the pre-final project evaluation, and at month 36: D 35, the fully functional, optimised and field tested prototype of ARGUS.

B3 Workpackage description

Workpackage number: WP 13, Exploitation and dissemination
Start date or starting event: 30
Participant number: 1; 7, 8
Person-months per participant: 8: 3 (1), 3 (7), 2 (8)

Objectives: To promote the project results and exploit synergies with other groups working in similar areas, to inform the public and in particular the blind community about the current status of the project and the intended results, to analyse future exploitation opportunities in particular with a view towards a commercially viable product, and to contact possible future commercial partners.

Description of work: In order to reach the objectives outlined above, a host of different activities is required, including the following: creation and continuous update of a web-site, preferably with its own domain `www.argus.*` and ftp-site both for the distribution among the individual partners of the consortium (containing e. g. information regarding workshops, seminars, conferences, call for papers, but also the source of the actual reports and algorithms), as well as to a wider public; presentation of ARGUS at congresses, symposia, conferences, technical events, etc. in addition to the individual partners' contributions to scientific conferences; production of an adequate and realistic business and exploitation plan for future commercialisation of the project results, including the identification of potential markets and potential competitors.

Deliverables: While most of the above description contains soft targets in connection with personal contacts and public relations not easily forged into deliverables, it is none the less possible to identify the following three deliverables: creation of a Dissemination Report (D 37), creation of an Exploitation Plan (D 38), and creation and continuous maintenance of a web-site including ftp and a mailing-list (D 40).

Milestones and expected result: Deliverable D 40, the web-page including ftp-access and a mailing list, as well as general work on public relations, will be started immediately following the kick-off meeting and will be kept up to date continuously until the end of the project. This will also include contacts to the press, the presentation of ARGUS on congresses etc.. As a continuous effort it is not linked to any date; it is to some extend part of the management of the project.
The creation of the Dissemination Report (D 37) and Exploitation Plan (D 38) are assumed to take 1.5 person-months each at the end of the project (month 36: D 37 and D 38).