A Data Fusion Architecture Proposal for Visually Impaired People

Natal Henrique Cordeiro[*‡], Antonio Miguel Batista Dourado[†‡], Gustavo da Silva Quirino[‡], Emerson Carlos Pedrino[‡]
[*]Federal Institute of São Paulo - Campus Votuporanga, SP, Brazil
[†]Federal Institute of São Paulo - Campus Boituva, SP, Brazil
[‡]Computer Sciences Department - Federal University of São Carlos (UFSCAR) - São Carlos, SP, Brazil
Email: {natalhenrique, dourado}@ifsp.edu.br, {gustavo.quirino, emerson}@dc.ufscar.br

***Abstract*—Based on the difficulties that a Visually Impaired Person (VIP) has in understanding certain contexts, this article proposes a new architecture that aims to contextualize elements of that environment and, thus, help the visually disabled person to move around safely. The main objective of this project is to extract environmental features and to subsequently perform the processes of perception, comprehension and projection of the Situation Awareness (SAW) model. The proposed architecture is composed of Computer Vision techniques (CV) and Data Fusion (DF). The CV techniques were used in order to obtain the necessary characteristics for the DF to perform the search for consistent relationships for pattern recognition and allow for the projection of possible collisions of the VIP with environmental elements. For the collision projection, the specific characteristics of each context were extracted and trained, thus providing the VIP with inferences made in real time.**

***Keywords*-visually impaired person; situation awareness;**

I. Introduction

Autonomous navigation has been explored in several studies in order to produce systems in the automotive industry and in robotics. These systems are designed to provide locomotion based on an understanding of features extracted from the environment. Several projects have adopted this type of technology to supplement the functions of a person with visual impairment [1], [2], [3], [4], [5]. VIPs require technologies that go beyond just indicating the desired destination; these technologies must recognize patterns, contextualize aspects of the environment and indicate what action should be taken to ensure the safety of the VIP.

Due to the difficulties that the VIPs have with orientation in specific environments, this project presents a system composed of computer vision techniques and analysis of images that provide data for the recognition of objects present in the scene also informing their locations, their distances, their movements and directions. This information is supplemented with data from physical sensors and integrated through a fusion system. This system was developed based on the Salerno Model for high-level fusion.

According to Liggins *et al* [6], this model incorporates concepts from the Joint Directors of Laboratories (JDL) model and the SAW model proposed by Endsley. Endsley *et al* [7] suggests that SAW is formed only when the first three levels of a data fusion process are performed, in which the first level is the perception, the second is understanding and the third projection. The design in this study uses a level of perception where information is captured from the environment in which a human being moves around, such as information about objects, people on stairs, upholstery, etc. The level of understanding concerns the relationship between the aforementioned elements and the meaning of their actions. The level of projection allows the prediction of a consequence in the near future based on the relationships and actions of environmental objects; such as a warning that the person may collide with an object. Research related to the development of VIP navigation support systems is made up of three fields of activity: internal navigation, external navigation and obstacle detection [8].

The main scientific contribution of this project is to develop a new architecture containing a data fusion module that applies the Salerno Model to provide refined information to the level of projection. This module has been developed for specific contexts present in indoor environments that offer some kind of danger to VIPs. Importantly, the use of these contexts (see Section III-B2) has not yet been explored at the level of projection. Therefore, it was not possible to carry out a comparative study with other Intelligent Systems techniques, statistics and probability, which could indicate those classification techniques that are the most suitable for certain scenarios. Sensors and computer vision techniques are not able to produce decision making comparable to that of a human being. Based on this concept, a data fusion system was implemented that relates all the data provided by the sensors and the techniques developed, contributing to decisions that are both more accurate and more akin to the way that people without visual impairment think.

Contributions: Several technological solutions have been presented in order to assist the VIP to find out their position, which elements are in their way and where is the safest place to move around. In most of these methods, different types of sensors were used, such as those that detect distance, presence, motion and colors [8], [9], [10], [11]. However, most of them hardly address data fusion systems to integrate all this information and make decisions based on the SAW level of projection. Typically these sensors are used as data sources, however, the systems do not have progressions to more refined fusion phases in order to correct errors, remove redundancies and generate decisions that a human being can trust. This project fits into contexts where data fusion is applied as a fundamental technique in decision making and aims to reach higher data-fusion levels in order to forecast possible collisions in the near future.

In this project, the data fusion is carried out after the integration of all the features extracted from the vision module with a view to building a portable product that uses little energy and allows threw system to be uploaded onto credit card-sized single-board computers. Another important feature of this design is the low cost compared to that of using physical sensors and, therefore, a product of smaller size. With the use of CV techniques, image analysis and DF in the architecture, it was possible to reduce the number of physical sensors, making the project cheaper.

II. Related work

The SAW model has been applied in a variety of projects. However it has not really been explored in support systems for navigation and sensory analysis by VIPs. The level of its use decreases substantially when the data fusion reaches the levels of understanding and projection. These levels aim to use the data in search of consistent relationships and consequently detect system patterns. In this project, the data fusion reaches the level of projection and the search for patterns is done by rules of associations using Intelligent Systems (IS) techniques, statistics and probability.

Ando *et al* [11] presents an interesting approach to the use of data fusion in a system that supports VIP navigation. Its goal is to provide continuous communication between the VIP and a network of sensors distributed in the environment to provide locomotion without collisions. That approach is presented as a good strategy to be implemented in public enclosed spaces such as schools, libraries and museums. The technology carried by the VIP, the sensors and the processing center are connected by wireless networks. The communication with the user is carried out by means of sounds. Ando's project does not use a set of CV techniques to provide characteristics, and so it needs to use a set of physical sensors around the environment.

Joseph *et al* [9] developed a system that adopts the Kinect motion sensor, to obtain a disparity map and addresses the use of social sensors to provide data from sites and social networks that support VIP to find out if there is any danger to mobility in a given region. With these data sources, the SAW can be generated, which allows the production of more reliable information. However, Joseph's system depends upon data feeds provided by others. Thus, Joseph *et al* [9] and Ando *et al* [11] have a high dependence on external sensors in their systems for performing data fusion. Tamjidi *et al* [10] states in his work that a way to reduce errors in navigation is with the addition of sensors and using data fusion but says that this method can generate a heavier, more complex system. In this project, the fusion is performed only on data from the Kinect device and using computer vision techniques. The Kinect was installed on the VIP's waist. This architecture does not depend on any external sensor and has a satisfactory performance in producing environmental information and drawing conclusions concerning the safety of the VIP.

In the context of support systems for VIPs, there is nothing in the literature that describes an architecture containing computer vision techniques and embedded machine learning to establish data fusion at a level that predicts collisions or accidents between a VIP and environmental elements. An important feature of this project is to provide the VIP with the possibility of performing inferences (making decisions) in two contexts; the possibility of passing between obstacles and collisions with moving objects. It is worth noting that comparisons with the learning techniques used in this project can serve as guidance tools to choose the most appropriate technique to be adopted in a given context.

Angin *et al* [8] proposes data extraction in a similar way to the structure used in this project. It links various data sources to provide decision-making information for the VIP. In that study, GPS is used to estimate position and a stereo camera to detect the distance of objects. The system also performs pattern recognition operations to provide more information for its data fusion system. However, it is not clear how the sensor data have been fused to look for patterns to provide impact projection information for the VIP. Nor is it discussed whether techniques were used to understand systemic patterns of context and provide more reliable information. So it appears that Angin's project only presents the implementation of data fusion at the level of perception.

There are many studies that use data fusion to support decision making and use, as a source of data, physical sensors, social sensors, as well as the classification of entities (objects and people) that can be used with CV and pattern recognition techniques. However, it is clear that SAW has not been investigated very much in the context of VIPs which aims to make decisions for safe mobility.

III. Sensory Analysis System For Visually Impaired Person (SAS-VIP)

This project, entitled Sensory Analysis System For Visually Impaired Person (SAS -VIP), aims to provide a system comprising of CV and DF techniques. The CV techniques perform extraction of a set of environmental characteristics. The DF techniques produce consistent associations using an intelligent system (IS) based on the SAW model. SAW should have the perception of the elements in an environment within a given time and space, the comprehension of their meaning and the projection of their actions in the near future [6].

The IS has learning characteristics based on a particular context and provides the execution of a set of inferences that aims to provide the VIP with a projection of collisions in certain contexts. For the development of an IS with the use of the SAW model, it is important to analyze the requirements covering all specific goals, composed of object attributes able to provide the perception, comprehension and projection.

According to Endsley *et al* [7], the Goal Directed Task Analysis (GDTA) performs the analysis of cognitive tasks in order to list all the entities and their functions, thus ensuring a reliable decision. This technique allows to activate the user's knowledge for certain tasks and to efficiently produce an analysis with the definition of goals. In this project, GDTA was developed based on interviews with VIPs at the Institute of the Blind. The main difficulties mentioned were recorded by means of questions: What elements are present in this environment? In which direction can I move? Is there enough space to travel in that direction? What is the position of the obstacles that do not produce sounds? Which objects are moving around? What are their direction, speed and size? What is the classification of these objects (static and dynamic)? Could I collide with an object (static or dynamic)?

After recording these questions, the system objectives were set up: to identify common objects that allow us to understand the context as people, doors and seats; to fix alternative directions for the VIP to take; to inform which are the regions with the potential for a collision; to alert which objects are moving, considering their distance, direction, speed and area; to predict collision courses in different directions based on the movements of objects and their classification. With the recorded interviews, the GDTA was developed as the first phase of development of SAS-VIP. The objectives and the entities that were fundamental to the development of the data fusion system with the use of Salerno model are shown in Figure 1.

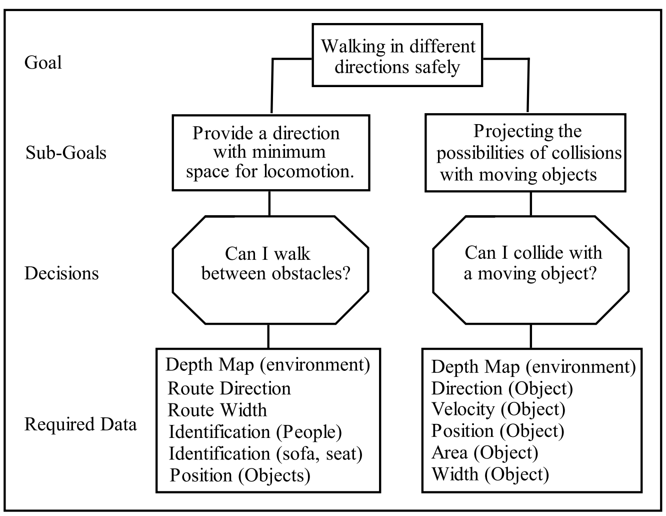


Fig. 1. An analysis of user's goals using GDTA.

After some experiments with the Kinect device, it was seen that the Kinect, combined with computer vision techniques, could provide a variety of environmental characteristics and, so, give the information required by the GDTA. Thus, a new architecture was constructed consisting of processes which produce a vector with sufficient characteristics for the IS to build solid relationships.

For the construction of this new architecture, a classification was used: modules, systems, processes and input and output devices. The modules consist of systems connected by processes which include the same purpose. This architecture (Figure 2) consists of two modules: the Vision Module (VM)(III-A); and the Fusion module (FM) (III-B).

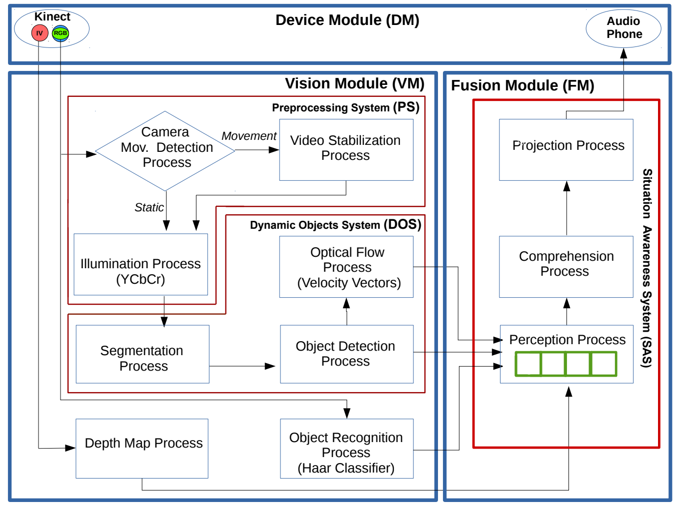


Fig. 2. Architecture of SAS-VIP

The input device used in this project is the Kinect. It is made up of a projector and infrared light, an infrared camera and an RGB camera. For the data from these sensors to generate different interpretations, they are subjected to filtering, segmentation and classification processes presented in the section III-A.

The output device is a headset that provides information and continual alerts on environmental characteristics, especially objects that are in the foreground. Using the phone, the VIP can be informed of some probabilities related to their safety, predicted by the FM.

A. *Vision Module*

The VM is composed of two systems and two complementary processes: Pre-processing System (PS); Dynamic Objects System (DOS); Depth Maps Process (DMP) and Object Recognition Process (ORP). Each VM process aims to provide the necessary requirements for feature extraction.

The Vision Module aims to extract characteristics of static objects in the environment. These characteristics are the position, distance and classification. For objects that are moving, the VM makes the extraction of the aforementioned features with the addition of the center of mass, area and velocity vectors.

1) Preprocessing System: The PS consists of three processes: Camera Motion Detection Process (CMDP); Video Stabilization Process (VSP); Illumination Process (IP). The PS is the first system to receive a given RGB sensor and check if the camera has some kind of movement. If the VIP is moving around or turning the CMDP automatically detects camera movement and sends the pictures to be stabilized by the VSP. The Pre-processing Process produces the last image in accordance with the PS data flow.

Camera Movement Detection Process: The CMDP was developed due to the presence of sharp movements in the camera. These movements impair the detection of moving objects and make it difficult to estimate speed and direction. When it detects camera movement, images are sent to the VSP for the stabilization to be performed. A phase correlation process was used to detect camera movement [12].

Video Stabilization Process: The analysis of moving objects in images is a challenging field, especially when the image acquisition sensor undergoes translational and rotational movements. This scenario becomes even more complex when the movements are produced in an uncontrolled way, such as when a camera is installed in something that is subject to changes in position. When this occurs, the video motion needs to be understood and then compensated for with respect to image objects. If the camera suffers no kind of movement, the techniques for analyzing moving objects can be more simplified. This process only stabilizes sudden movements of translation and rotation. Thus, for the stabilization to occur, the geometric transformation coefficients are estimated from a frame obtained at a given instant and a frame captured in a second moment. Importantly, the estimated processing coefficients should be automatic and fast so as not to harm system performance and the VIP's knowledge of the situation. The corner detection in previous image and its displacements in the current image allowed it to achieve the coefficients to be applied in Affine transform. Harris Corner method proposed by Harris *et al* [13] was used for corner detection.

Illumination Process: The Illumination Process (IP) consists of two phases that manipulate light. The first stage performs the radiometric correction after discovering the intrinsic parameters through the camera calibration [14]. In this same stage, the radial distortion caused by camera geometry is removed. In the second phase, the lighting control is carried out with the transformation between color spaces. It was found that the RGB color space makes it difficult to separate details where the hue, saturation and intensity are together. Zhang [15] states that each RGB color space channel is highly correlated and dependent on the others. Although the HSI color space introduced many improvements allowing control of lighting without the dependencies found in RGB, the YCbCr color space gives better results when applied to the background subtraction algorithm in the Segmentation Process (SP). The algorithm 1 shows the order of all the methods implemented in the SP to analyze the camera motion, perform stabilization and control lighting to send the images to the Dynamic Object System.

Algorithm 1: PREPROCESSING SYSTEM (PS)

Input: Video in real time.
Output: Stabilized Frame and controlled illumination.

1 **while** $(CaptureVideo)$ **do**
2 $\quad (\langle f_1, f_2, \ldots, f_n \rangle) \leftarrow$ convertFrame($video$)
3 $\quad prevF \leftarrow f_1$
4 $\quad currF \leftarrow f_2$
5 $\quad hann \leftarrow$ createHanningWin($sizeCurrF$)
6 $\quad shift \leftarrow$ phaseCorrelate($prevF, currF, hann$)
7 $\quad dfPrevF \leftarrow$ DFT($prevF$)
8 $\quad dfCurrF \leftarrow$ DFT($currF$)
9 $\quad R \leftarrow (dfPrevF * dfCurrF)$ / $|dfPrevF * dfCurrF|$
10 $\quad r \leftarrow$ DFT^{-1}(R)
11 $\quad (shift.x, shift.y) \leftarrow$ weightedCentroid(argmax(r))
12 $\quad radius \leftarrow$ sqrt($shift.x^2 + shift.y^2$)
13 $\quad$ **if** $(radius > 2)$ **then**
14 $\qquad$ cornerDetect($currF, crCurrF$[ ])
15 $\qquad$ calcOF($prevF, currF, crPrevF$[ ], $crCurrF$[ ])
16 $\qquad T \leftarrow$ estimRigidTransf($crCurrF$[ ], $crPrevF$[ ])
17 $\qquad$ AffineT($prevF, currFTransf, T, currF.size()$)
18 $\quad$ Ilumination($k1, k2, \alpha, g, b, y, cb, cr$)
19 **return** *stabilizedFrame*

TABLE I
DESCRIPTION THE PROCESSING OF THE PS ALGORITHM 1.

Line	Processing
Line 1, 2	Get video from camera and convert video to image.
Line 3, 4	Get Previous Frame (PF) and Current Frame (CF).
Line 5	Applies a Hanning window to remove edge effects
Line 6	Detect Translational Shifts with phase correlation
Line 7, 8	Computes the forward Discrete Fourier Transform (DFT) of each source array
Line 9	Computes the cross-power spectrum of each frequency domain array
Line 10	The cross-correlation is converted back (inverse DFT)
Line 11	This function computes the peak location
Line 12	Parameter Values that define camera motion
Line 13	Defines the movement of the camera based on radius value
Line 14	Detect corners in preview image
Line 15	Get all velocity vectors using optical flow theory
Line 16	Computes affine transformation between two 2D point sets
Line 17	Applied coefficients of affine transformation in PF
Line 18	Radiometric and geometric correction and control YCbCr.
Line 19	Returns adequate image for DOS.

2) Dynamic Objects System: VIP mobility is difficult in scenarios containing obstacles and people moving, especially when devices are not available that mention the presence of elements that could generate some collision. The Dynamic Object System extracts features related to moving objects in the VIP field of view. The objective of this system is to count the objects that are in motion, analyze their directions, calculate the speed and estimate their area. Acquiring these characteristics and relating them to the environment is important for the data fusion system to predict a safe path for the VIP.

The number of data sources may, at the same time, reflect the quality of the response or projection of the Fusion System. The importance of this technology, which provides analysis of moving objects to assist the disabled person in decision-making, is clear. This system provides the location of obstacles and the direction of their movement with the goal of predicting a possible collision with the VIP.

Segmentation Process: In order to provide more refined data for the analysis of moving objects, it was decided to use, in the first stage of this system, a segmentation algorithm that isolates moving objects from other objects that complete the scene (the background). The segmentation method used for this purpose is Background Subtraction (BS). This technique is often used to identify objects that are moving in a given scene. The Codebooks algorithm [16], [17] was chosen to perform segmentation due to its high performance characteristics and its application in various environments. This method allows all the processing to focus on the moving objects which consequently have higher chances of collision with the VIP. The SP can be used as a filter which eliminates irrelevant elements in the image and so provides only the data needed for the Optical Flow Process (OFP)(III-A2) discussed further below. The segmentation algorithm is also used as a preprocessing for the implementation of the edge detection algorithm used to compute the area and the center of mass of a moving object.

Object Detection Process: After applying the method of segmentation and isolation of objects that are in the foreground, a technique to detect contours [16] and the center of mass of the object to estimate its area in pixels is used. For the generation of the boundary, a threshold filter was used, followed by the Canny edge detection filter to eliminate remaining noise and, consequently, produce a better quality result for the area calculation. When the detector provides a region bounded by a curve or a closed contour, its area can be calculated.

Optical Flow Process: The OFP provides data which allows understanding of the speed and direction of a set of pixels based on their intensity patterns. To be able to understand these phenomena, a sequence of images with variations in time is used. For each pixel in this sequence, a velocity vector is produced; however, small image regions are generally used to ensure an identifiable point. According to Braski [16], when strong derivatives are found in two orthogonal directions, it is assumed that this region is unique. In that way, you can get the position (x,y) of a region defined as a corner and provide sufficient coefficients for calculating the optical flow.

For obtaining the velocity vector of the dynamic object, a RGB value was added at position(x, y) in the center of mass of the object and then was detected the new position of such a RGB value in the subsequent frame. After checking some computer models that estimate the optical flow, here a different technique is used. In this work, the Lucas-Kanade (LK) estimation method was used to obtain the optical flow. LK is a technique known for its robustness against noise and for its high performance in tracking objects. It is used in comparing stereo images and can also be used for pattern recognition. Algorithm 2 presents the sequence of techniques implemented to extract the characteristics that estimate the position, direction and speed of moving objects.

Depth Map Process: The Depth Maps Process (DMP) belonging to the VM, is responsible for generating maps which determine the depth or distance of the elements that compose the environment. The map is produced by means of an infrared emitting sensor and an infrared receiving sensor both integrated with the Kinect.

Object Recognition Process: The Object Recognition Process (ORP) aims to detect patterns that resemble the characteristics of people, obstacles and stairs, important to the data fusion module. Here the Haar Classifier method proposed by Viola *et al* [18] was used to perform a supervised training. The high performance in the detection of an object present in an image is the main feature of this method.

Algorithm 2: DYNAMIC OBJECTS SYSTEM (DOS)

Input: Video with stabilized Frame and controlled illumination.
Output: Segmented Dynamic Object.

1 **while** *(capturePSVideo)* **do**
2 $cpYCbCr \leftarrow$ copyImage($YCbCr$)
3 $model \leftarrow$ createBGCodeBookModel()
4 bgCodeBookUpdate($model, cpYCbCr$)
5 bgCodeBookDiff($model, cpYCbCr, imgCBook$)
6 segmentForeground($imgCBook$)
7 cannyFilter($imgCodeBook, cannyOutput$)
8 findContours($cannyOutput, contours$)
9 contourArea($contours$[x])
10 $(\langle CM_1, CM_2, \ldots, CM_n \rangle) \leftarrow$ CM($contours.size()$)
11 getPosition(CM_1 , $PVVPrev$[ ])
12 calcOF($CM_1, CM_2, PVVPrev$[ ]$, PVVCurr$[ ])
13 **return** *segmentedObject*

TABLE II
DESCRIPTION THE PROCESSING OF THE DOS ALGORITHM 2.

Line	Processing
Line 1, 2	Get Video from PS and Convert Video YCbCr to Image
Line 3	Initializes background model
Line 4	Building background model.
Line 5	Find foreground by codebook method
Line 6	Visualizing bounding boxes and centers
Line 7	Detect edges using Canny Filter
Line 8, 9	Find contours and Calculate the area
Line 10	Get the Center of Mass (CM)
Line 11	Get position in Previous Velocity Vector (PVV).
Line 12	Get all positions PVV using Optical Flow Theory.
Line 13	Returns Actual Position of the Segmented Object.

B. Fusion Module

The Fusion Module (FM) seeks consistent relationships between the characteristics defined by GDTA and aims to generate contextualized information to provide safe, reliable mobility for the VIP. The FM is composed of three processes: the Perception Process (PEP), the Comprehension Process (COP) and the Projection Process (PROP). Each process belonging to the FM is intended to provide the information needed to make the decisions (see Figure 1).

1) Perception Process: This process analyzes the characteristics of relevant objects in the scene. The contextualization of these objects provides the requirements for the highest levels of abstraction of the SAW. According to Salerno's Model for Higher-Level Fusion [6], this process involves analyzing: the existence and number (How many?); the identity (What / Who?); and the kinematics (Where / When?).

In the SAS-VIP, PEP aims to organize and combine the data provided by the VM, in addition to producing information for the Comprehension and Projection processes.

In this project, the PEP combines the information from the VM to estimate the direction and average acceleration of moving objects, calculating appropriate distances for movement and produce walking routes. Based on the velocity vectors from the VM, the PEP calculates the speed of a moving object and gives the direction of its path. The speed of the moving object was calculated based on the distance in pixels that the object moved. For this calculation, the value of the Euclidean distance (pixel distance) was divided by time.

For the direction, a function to generate the $arctangent$ between the start and end points of the path taken by the moving object was used. To provide alternative routes to the VIP, a method that runs the depth map in search of regions with greater distances (depth) was developed. This search is based on the gray scale of the pixels. Detecting the direction of that region, this method checks that the width of the obstacles and the suggested route allow for the VIP to pass safely. An object in a 2D image may vary in size when a VIP changes its own position. With the intrinsic parameters from camera calibration, thresholding of disparity map and contour detection provided by the VM, it is possible to calculate the width (in centimeters) of objects and passage.

2) Comprehension Process: The Comprehension Process (COP) performs the second phase of the FM. In the COP, a new data fusion is performed, however using information that has already been refined by the above process. Analysis of objects is carried out, based on their behavior, actions, intentions, relevance and capability [6]. Through this analysis, the possible impact on the goals presented in the GDTA can be checked. Here, the effects to be studied relate to providing a route that allows the VIP to move without crashing into obstacles and to detecting possible collisions with moving objects. Based on these two goals, two respective contexts were analyzed by the Comprehension process.

To understand the first context, it has been shown that prior knowledge such as how to move safely is relevant in generating decisions closer to reality. To check the possibility of clearance between obstacles, the width and position of the passage had to be considered and, also, the position of seats and people. It was decided to use Bayesian networks theory. This theory is suitable for applications where there is no need to represent ignorance, that is, when it is easy to define a representation from prior probabilities.

This technique allows for the production of a smart decision based on the data and information available regarding the environment. Thus, it allows for informing the VIP about the possibility of collisions.

Based on new relationships detected in the VIP context, Bayesian networks can generate changes in their relationships (initial probabilities), which make them adaptable and effective in certain contexts. The second affected context for the COP needed to use variables with more complex forecasting possibilities. So, in this context, a Support Vector Machine (SVM) was used for the classification.

To check the possibilities of a collision with moving objects, the distance, the horizontal position, the direction and the speed of moving objects needed to be included. The use of SVM is focused on contexts in which the data used for the pattern recognition can be separated into two classes. The SVM classifier aims to find the best hyperplane to separate classes with the presence of or the absence of a collision.

3) Projection Process: The Projection Process (PROP) is the third phase of the FM. This process needs to merge the information produced by the COP in order to design situations in the near future. The PROP projects the actions of objects in relation to the VIP to define the possible collisions for a given route. Inferences with elements and current situations are made that project potential collisions.

According to Liggins *et al* [6], beliefs networks allow relationships to be formed in a given context. From these relationships, it is possible to classify objects in such a way that the use of inference can project situations in the near future. Inferences are provided by a set of information generated by the vision and fusion modules in real time.

On detection of any relevant object present in the COP contexts, the PROP sends the identification and position of this object to the headset. After this process, the VIP may require an inference to the contexts that this object belongs to. Inferences available in the PROP are based on the two sub-goals of the GDTA.

IV. Experiments

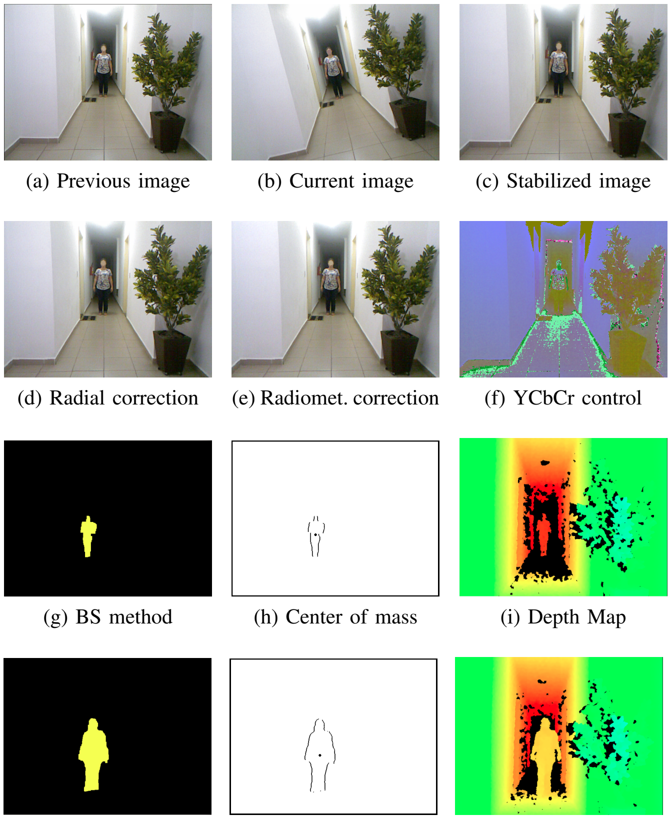
Experiments were carried out with all the processes of SAS-VIP architecture to validate the information that the Fusion module needs to provide inferences for a given context. Figures 3 (a and b) show a sequence of images acquired by the Kinect device.

Video motion (rotation to the left) between images (a) and (b) of Figure 3 can be seen. Image (c) was generated after motion was detected in the video and subsequent geometric transformation performed to stabilize the movement.

Before sending the images to the segmentation process, radial correction (Figure 3 (d)), radiometric correction (Figure 3 (e)) and lighting control (Figure 3 (f)) in the YCbCr color space were done. This control allows for noise reduction and an increase in the quality of the segmented object. The image provided to the DOS can be seen in Figure 3 (f).

Segmentation by BS method provides the position of all moving objects in the scene. However, it is a process that requires completely stable images so that camera shake does not impair the detection of moving objects. In Figure 3 (g) the segmentation of a moving person can be seen.

Figure 3 (h) shows the center of mass and the contour of this object and in Figure 3 (i) distances of all elements present in the SAS vision are given. Figures 3 (j, k and l) show images where the same techniques as those used in (g), (h) and (i) have been applied to images captured a short time later. With all the identified features sent to the FM, the PEP calculated the direction and speed of the moving objects. This calculation was based on the velocity vectors generated from the center of mass shown in Figures 3 (h) and (k).



(a) Previous image (b) Current image (c) Stabilized image

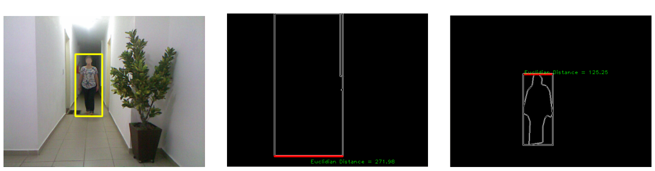
(d) Radial correction (e) Radiomet. correction (f) YCbCr control

(g) BS method (h) Center of mass (i) Depth Map

(j) BS method (k) Center os mass (l) Depth Map

Fig. 3. Preprocessing System and Dynamic Objects System

Figure 4 shows the result of the algorithm that calculates the clearance width of the passage and the detected obstacles (a) shows the result of the recognition of people. Figures 4 (b) and (c) validate pedestrian passage in certain areas.



(a) Person Detect (b) Estimated width (Passage) (c) Estimated width (Person)

Fig. 4. Object Recognition and PEP Process

To make a comparison between classifiers, the information produced by the SAS-VIP to feed the database of the two contexts elaborated in the COP (Section III-B2) were used. Table III shows the results of the training processes of two classifiers applied in each context. This table consists of the following features: the context in which the VIPs want to support decision making; the classifiers used; the Correctly Classified Instances (CCI) for each classifier; the average precision (avg Pc) of each training; the True Positive (TP) rate. For this comparison, the following classifiers were used: Bayesian Network, Support Vector Machine (SVM) and Multilayer Perceptron. You can see that the Multilayer Perceptron classifier had a higher percentage of the CCI in relation to the other classifiers in both contexts. However the Bayesian Network TP Rate shows that its results are more reliable due to better detection of true positives.

TABLE III
CONTEXTS OF PROJECTION

Context	Classifiers	CCI	AVg Pc.	TP Rate
Passage Context	Bayesian N.	81.2%	82.5%	79.1%
Passage Context	M. Perceptron	83.4%	83.5%	74.8%
Dyn. Obj. Collision	SVM	77.27%	81.2%	77.3%
Dyn. Obj. Collision	M. Perceptron	97.72%	97.8%	97.7%

With the trained dataset, the inferences were made to validate the SAS-VIP projection process. The results of the inferences are discussed in Section V . Bayesian networks are interesting because different probabilities for different variations in the context can be used, being one way to present more refined information than providing only a value with a logical data type.

V. RESULTS AND DISCUSSION

The SAW model has been little explored in contexts that support VIPs, especially in predicting collisions. Thus it was not possible to make comparisons of the Fusion Module and the inferences it made with other studies. The following tables present the results of some inferences produced for the context of accidents and collision with moving objects defined in the COP (Section III-B2). The first column (Passage Width) of Table IV, indicates whether the detected passage is smaller or larger than 80 centimeters. The second column shows whether the passage is in the center of the VIP's field of view. The third column indicates whether there is any person in the passage available. The fourth column indicates the presence of some sort of seat or obstacle in the direction of the passage. The last column shows the percentage chance of safe passage in the region found by the Perception Process (Section III-B1).

TABLE IV
CHANCE OF PASSAGE

Pass. Width	Pass. Pos.	Person Pos.	Obstacle Pos.	Chance
Passage >80	No Center	No Center	Center	29.9%
Passage >80	Center	Center	Center	60.0%
Passage >80	No Center	Center	Center	44.9%
Passage >80	Center	Center	No Center	95.0%
Passage <80	Center	No Center	Center	0.00%

The results in Table IV were obtained through a program designed specifically for performing Bayesian networks inferences, a subclass of a model of probabilistic networks. The results of the inferences were satisfactory and similar to the beliefs of a person without visual impairment. The possibilities generated by Bayesian networks are formed using the following factors: the possibility of crossing exists only when its width is greater than 80 cm; the passage must be in the center of the VIP's path; the presence of people in the crossing increases the chances of collision-free mobility past the obstacles; The presence of obstacles in the passage decreases the chances of safe passage, and trying a new route is suggested. It is worth noting that the presence of people is represented by sound signals to the VIP.

So, the VIP just waits for the person to move for the passage to be clear. The possibility of passing 29.9% presented in the first row of Table IV is justified because there is a passage with width greater than 80 cm. However, it is a small percentage due to the following: the passage is not in the central axis of the VIP; there is not a person moving in the passage; the existence of an obstacle, such as a seat, in the area of the passage. In the fourth row in this table, the possibility of 95% is a substantial increase due to: the passage being on the central axis of the VIP; there are people moving in that region; there are no obstacles in the passage. Table V is composed of four dynamic characteristics of moving objects (distance, speed, direction of movement and position) and another field indicating the occurrence of a collision.

TABLE V
CONTEXT - COLLISION WITH DYNAMIC OBJECTS

Depth	Speed	Direction	Position	Collision
$200cm$	$0.7m/s$	93°	88°	Yes
$81cm$	$2.8m/s$	150°	29°	No
$120cm$	$0.9m/s$	88°	125°	Yes
$152cm$	$2.4m/s$	25°	152°	No
$119cm$	$1m/s$	86°	122°	Yes

The values that are presented in this table were used to perform inferences that verify the possibility of collision between the VIP and a moving object. To create a training dataset, different positions, distances, speeds and directions of moving people were collected, including data on possible collisions with the VIP in a hallway (indoor environment). Thus, 60 situations were simulated between safe paths and collision paths for a VIP. The training phase was performed in supervised mode. The results of inferences presented in the table below were produced by the SVM, a technique that comes from statistical learning theory.

VI. CONCLUSION

Many of the projects developed in order to support the mobility of VIPs show the existence of obstacles, position, distance, and alternative routes for movement. However, few have applied the prediction of collisions based on a better understanding of the context. This project provides an architecture that provides a set of extractors of basic characteristics to produce the perception and understanding of the environment.

This architecture is also available to use with classifiers specific to each type of context in order to support the VIP in their decision making. The results of the Vision Module techniques are of good quality and have been validated by the Fusion Module, which, without the available features, would not be able to make the inferences discussed in the Results (Section V). Thus, this project makes an important contribution to the development of systems aimed at predicting collisions in different contexts. Among these contributions are the architecture, the method, the definition of classifiers for certain given contexts and especially the way inferences are made that predict collisions in the near future.

VII. ACKNOWLEDGMENT

We are grateful to the Brazilian funding agency FAPESP – Project No. 2015/23297-4. Our thanks go as well to UFSCar–DC, and IFSP, SP, Brazil.

REFERENCES

[1] N. Bourbakis, S. K. Makrogiannis, and D. Dakopoulos, "A system-prototype representing 3d space via alternative-sensing for visually impaired navigation," *IEEE Sensors Journal*, vol. 13, no. 7, pp. 2535–2547, July 2013.

[2] A. Aladren, G. Lopez-Nicolas, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using rgb-d sensor with range expansion," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–11, 2014.

[3] M. Yusro, K. M. Hou, E. Pissaloux, H. L. Shi, K. Ramli, and D. Sudiana, "Sees: Concept and design of a smart environment explorer stick," in *Human System Interaction (HSI), 2013 The 6th International Conference on*, June 2013, pp. 70–77.

[4] A. Alkhanifer and S. Ludi, "Towards a situation awareness design to improve visually impaired orientation in unfamiliar buildings: Requirements elicitation study," in *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*, Aug 2014, pp. 23–32.

[5] X. Z. B. L. X. L. Jizhong Xiao, S. L. Joseph and J. Zhang, "An assistive navigation framework for the visually impaired," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 635–640, Oct 2015.

[6] M. Liggins, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition*, ser. Electrical Engineering & Applied Signal Processing Series. CRC Press, 2008.

[7] M. Endsley, B. Bolte, and D. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. Taylor & Francis, 2003.

[8] P. Angin, Bharat, and K. Bhargava, "Real-time mobile-cloud computing for context-aware blind navigation," July 2011.

[9] S. L. Joseph, J. Xiao, B. Chawda, K. Narang, and P. Janarthanam, "A blind user-centered navigation system with crowdsourced situation awareness," in *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2014 IEEE 4th Annual International Conference on*, June 2014, pp. 186–191.

[10] A. Tamjidi, C. Ye, and S. Hong, "6-dof pose estimation of a portable navigation aid for the visually impaired," in *Robotic and Sensors Environments (ROSE), 2013 IEEE International Symposium on*, Oct 2013, pp. 178–183.

[11] B. Ando, S. Baglio, S. L. Malfa, and V. Marletta, "A sensing architecture for mutual user-environment awareness case of study: A mobility aid for the visually impaired," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 634–640, March 2011.

[12] A. N. J. Raj, S. S. Majeeth, and R. C. Staunton, "A comparison of fft and dct based phase correlation function for focused and defocused images," in *Machine Vision and Image Processing (MVIP), 2012 International Conference on*, Dec 2012, pp. 173–176.

[13] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[14] E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*. Prentice Hall, 1998.

[15] Z. Qiu-yu, L. Jun-chi, Z. Mo-yi, D. Hong-xiang, and L. Lu, "Hand gesture segmentation method based on ycbcr color space and k-means clustering," *International Journal of Signal Processing - Image Processing and Pattern Recognition*, vol. 8, no. 5, pp. 105–116, Jun. 2015.

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

[17] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, Jun. 2005.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, 2001, pp. I–511–I–518 vol.1.