Title

Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation

Journal

IEEE Transactions on Vehicular Technology, 53(6)

ISSN

0018-9545

Authors

Trivedi, Mohan Manubhai
Cheng, S Y
Childers, EMC
et al.

Publication Date

2004-11-01

DOI

10.1109/TVT.2004.835526

Peer reviewed

Occupant Posture Analysis With Stereo and Thermal Infrared Video: Algorithms and Experimental Evaluation

Mohan Manubhai Trivedi, *Senior Member, IEEE*, Shinko Yuanhsien Cheng, *Student Member, IEEE*, Edwin Malcolm Clayton Childers, and Stephen Justin Krotosky, *Student Member, IEEE*

***Abstract*—Dynamic analysis of vehicle occupant posture is a key requirement in designing "smart airbag" systems. Vision-based technology could enable the use of precise information about the occupant's size, posture, and, in particular, position in making airbag-deployment decisions. Novel sensory systems and algorithms need to be developed for capture, analysis, and classification of dynamic video-based information for a new generation of safe airbags. This paper presents a systematic investigation in which stereo and thermal long-wavelength infrared video-based real-time vision systems are developed and systematically evaluated. It also includes the design of several test beds, including instrumented vehicles for systematic experimental studies for the evaluation of independent and comparative evaluation in automobiles. Results of extensive experimental trials suggest basic feasibility of stereo and thermal long-wavelength infrared video-based occupant position and posture-analysis system.**

***Index Terms*—Head detection, infrared imaging, machine vision, real-time vision, stereo vision, tracking.**

I. Introduction

ACCORDING TO the National Highway Traffic Safety Administration (NHTSA) [1], in the past ten years, airbags were deployed more than 3.3 million times in the U.S.. Airbags are credited with saving more than 6000 lives and preventing a much greater number of serious injuries. These numbers clearly highlight the life-saving attributes of airbag technology. Alas, there are other rather disheartening numbers that are presented in the same report. It states that since 1990, over 200 fatalities have been recorded as a direct result of an airbag deployment. The majority of these deaths have been children. The number of severe injuries is much higher. Obviously, these deaths and injuries must be prevented by exploring and adopting the most appropriate technologies, policies, and practices. A new law, which went into effect in the U.S. in the beginning of 2004, requires that airbag systems are able to distinguish persons that are too small or persons in unsafe positions from persons of a safe size and in safe positions for airbag deployment [1]. One of the main difficulties encountered by the decision logic systems used in airbag deployment deals with the critical assumption about the occupant size and position in the car at the time of a crash. Most airbag systems consider a single standard for the occupant's size and the nature of the crash. Vision-based technology enables the use of precise information about the occupant's size, position, and posture to aid the single standard airbag system in deciding whether the occupant is of the right type for deployment.

Our overall research objective is to describe the design, implementation, and evaluation of vision-based occupant posture-analysis systems to control the deployment of an airbag in a safe manner. These efforts resulted in the development of the following:

1) framework for sensing the most relevant visual information;
2) set of robust and efficient algorithms for extracting features that characterize the size, position, and posture of the occupant;
3) pattern-recognition module to classify the visual cues into categories, which can trigger the safe deployment logic of the airbag system.

In this paper, we describe our experiments on two systems that estimate occupant body position and pose inside a vehicle using long-wavelength infrared (LWIR) imagery and stereo depth data. We will show that these systems are capable of reliably and of accurately extracting and tracking the position of the head in real time.

A novel test bed was constructed to perform the comparison between these vision-based "smart airbag" systems. The test bed is built around a Volkswagen Passat outfitted with the necessary equipment to perform real-time side-by-side tests. This framework not only subjects the systems to realistic conditions, but also allows us to quickly accommodate new systems for comparison under identical conditions.

II. Research Objectives and Approach

The objective of the proposed research is the development of a highly reliable and real-time vision system for sensing passenger occupancy and body posture in vehicles, ensuring safe airbag deployment and helping to prevent injuries. The design of the "smart airbag" system can be divided into three parts: 1) real-time scene sensing; 2) feature selection; and 3) body size,

Manuscript received July 22, 2003; revised March 29, 2004 and May 13, 2004. This work was supported in part by the University of California Discovery Grant with the Volkswagen Research Laboratory under the Digital Media Innovation Program.

The authors are with the Computer Vision and Robotics Research Laboratory, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: mtrivedi@ucsd.edu; sycheng@ucsd.edu; echilde1@san.rr.com; skrotosky@ucsd.edu).

Digital Object Identifier 10.1109/TVT.2004.835526

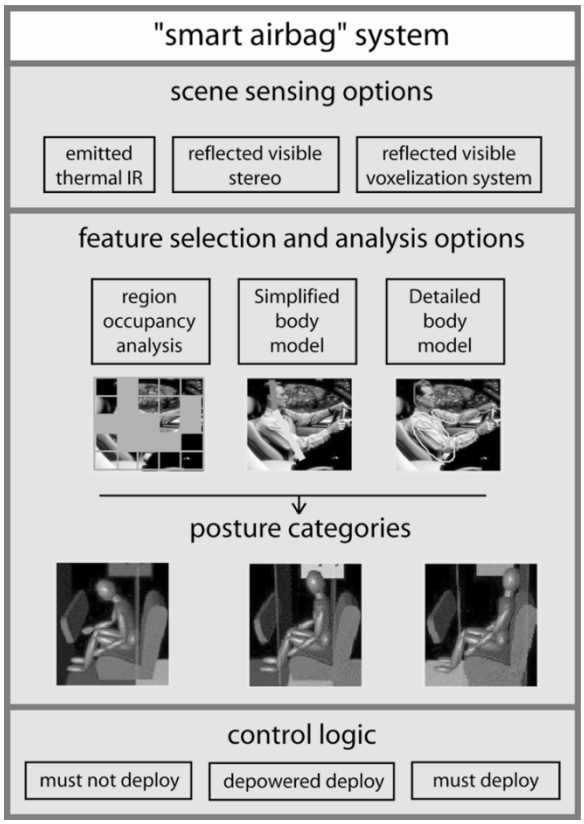

Fig. 1. Occupant position and posture-based safe airbag deployment design approach.

posture, and movement analysis, followed by decision logic for various levels of airbag deployment (see Fig. 1). We propose using video cameras for their unobtrusiveness and potential for other purposes beyond "smart airbags."

For scene sensing, we consider emitted LWIR imaging and stereo depth imaging. For feature selection and analysis, we consider both simple region occupancy features to detailed human body model pose estimation. For both scene sensing and feature selection and analysis, the choice of method will depend on its robustness and cost. However, using stereo or multicamera systems with high-level human body modeling would provide information that is useful not only for optimal airbag deployment, but for other applications with minimal extra effort. High-quality input data and detailed analysis of body pose can also be used to enhance safety by analyzing driver alertness and could also be used to build intelligent interfaces to different in-car devices, such as the mobile phone or radio [2].

To determine whether a person is in the right position for airbag deployment, the area between the back of the seat and the dashboard can be divided into sections. A diagram of the in-position (IP), out-of-position (OOP), and critically out-of-position (COOP) areas in the passenger seat is shown in Fig. 2. By analyzing these regions, we can categorically examine the human body under various positions that an occupant can take in the passenger seat, including sitting in a normal position, leaning forward, reaching down, seated with the seat advanced, reclined, slouched, knees on the dashboard or the edge of the seat, etc.

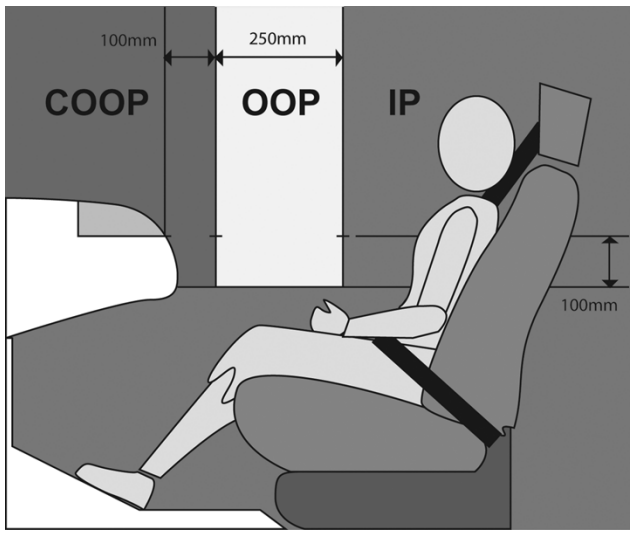

Fig. 2. Schematic showing the IP, OOP, and COOP regions of the passenger seat.

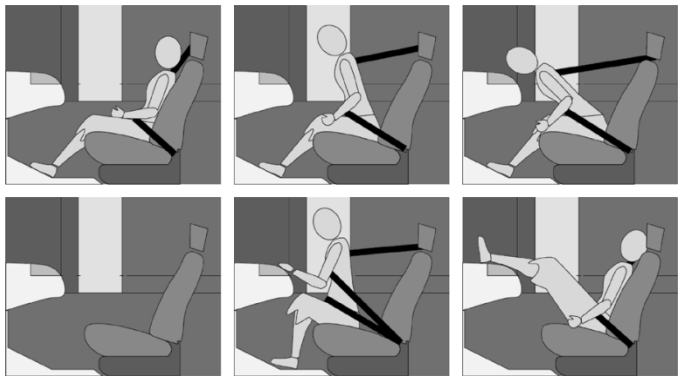

Fig. 3. Selected occupant positions considered in the development of "smart airbag" systems. In row-major order, the correct classifications for each situation is IP, OOP, COOP, not found, OOP, COOP.

The set of possible positions can be extended to include these positions and others that infants, three-year-old children, six-year-old children, fifth-percentile female adults, and 50th-percentile male adults can take on as occupant. These occupant types are used to define compliance with the new airbag law in the Occupant Crash Protection Standard (FMVSS 208) by the NHTSA [1].

Associated with each position is a desired airbag response, which is to deploy normally, deploy with limited power, or to suppress deployment. Associated with each desired operation is a cost of making the wrong decision. This cost is weighted relative to the other positions that the systems are to recognize. The average cost of making an error is system's primary performance measure. Fig. 3 shows a few positions that are considered in our investigations.

III. Related Studies

The focus of this research is to investigate vision-based systems that estimate occupant size, position, and pose. Vision is attractive because of its passive nature and its potential to provide a multitude of cues for safe airbag deployment, as well as for other uses from a single sensor.

Alternative sensor modalities for the purpose of safe airbag deployment include a system of measuring the weight distribution placed in the seat. Technologies for measuring the presence and position of an occupant include ultrasound, capacitance, and near-infrared spotlight grid. For detecting a rear-facing infant seat (RFIS), which is a "must not deploy" occupant, resonating proximity sensors, much like radio-frequency identification (RFID) cards for unlocking doors in offices, have been employed to detect presence of RFISs.

There exists a number of experiments using active illumination to capture the scene features [1]–[7]. They range from unobtrusive near-infrared light-emitting diodes (LEDs) to projecting light patterns and emitting multiple flashes to light the scene. The benefits of active illumination is a system that is less sensitive to different types of scene-illumination changes, but for some systems that gain comes at the cost of being obtrusive to the environment. In contrast, our focus is purely on unobtrusive active illumination, such as near-infrared illuminators or passive scene sensing, and we show that obtrusive lighting schemes are not necessary for robust estimation of occupant posture information.

Reference [8] presents an approach in which four image features and a series of learning algorithms are used to classify conditions that are safe or unsafe for airbag deployment. These four features were two-dimensional (2-D) and range (2-1/2-D) features, edges, motion, shape, and range. However, they do not consider emitted energy from the passenger nor volumetric representations of the passenger acquired from multiple points of view in the context of side-by-side system comparisons. Furthermore, these systems rely on the fact that a complete and varied reference image set of all passenger seat occupant types and positions, including the many types of infant seats, are known *a priori* for the training of the system to recognize occupant posture.

Faber [9] previously documented the approach that restricts the problem of describing the occupant space as occupied or unoccupied by using sparse range data of the passenger seat back. Stereo-based range data was used to detect the presence of a backrest to determine occupancy. Although this system was able to acquire occupancy information from the seat well, this system lacked detailed occupant size, position, and pose information.

In [10], Krumm and Kirk augmented the class space to include the presence of rear-facing infant seat, which together with the unoccupied classification requires the airbag to be turned off. Krumm took both image intensity (2-D) and stereovision-based (2-1/2-D) range data and found the principle components for each class, with which nearest neighbor classification was performed. Both systems were found to be quite successful in classifying the passenger seat into these three classes. However, this approach required an even larger training data set to include the types of car interiors along with person type and position to achieve this accuracy. And, like [9], the information that this system provides lacks detailed occupant position information.

In describing occupant position, the head appears to be the easier human part to detect that simultaneously provides rich implicative positional information on the rest of the body. Remaining in the 2-D category, Reyna *et al.* [11] uses a modified support vector machines (SVM) method to estimate the location of the head, much in the same way that face recognition is performed using SVM. They found an accuracy rate of 72% correct detection and 33% false alarm. Drawbacks in such a system assume that a representative set of head images are *a priori* known and hardware requirements for real-time execution of this algorithm are considerable. In contrast, we propose a method that gathers thermally emitted, stereo depth, and volumetric information that requires no training and, with the exception of the third, operates in excess of 15 frames per second (f/s) on an off-the-shelf Windows PC.

In the 2-1/2-D and three-dimensional (3-D) sensing category, the research community has made some effort. We can subdivide their approaches into two main classes: one based on region occupancy detection and the other based on object tracking. The first is an approach in which the region in which the feature resides, regardless of whether it is a head, arm, limb, or noise, determines the outcome of the classification. The other is an object-tracking-based approach in which a particular body part is tracked, providing unambiguous body-pose information such as limb orientation, angle with respect to other body parts, sizes, and lengths. The lack of a fit to a body model usually implies the existence of an object other than a human body, requiring other models to be used in detection of child seats, inanimate objects, etc.

Lequellec *et al.* [5] approached the problem of modeling occupants by way of projecting a pattern of dots onto the occupant and detecting their position in 3-D space using epipolar geometry, a stereovision-based technique (2-1/2-D). Devy *et al.* [12] get rid of the active illumination requirement of [5], relying solely on stereovision and features in the scene to provide a dense stereo reconstruction of the occupant, 3000–5000 as opposed to 400 3-D points in [5]. Both systems by Devy *et al.* and Lequellec *et al.* and our stereo-based approach rely on the surface of the occupant. Our effort differs in that we fit the acquired depth data to a model that we then track from frame to frame.

Farmer and Jain [13] presented work on an occupant classification system that addresses the *static suppression* requirement of the NHTSA standard by a system that is able to discern between four categories of occupants with high detection rates. The *static suppression* requirement specifies occupant types that, when present in the passenger seat, the airbag system must automatically suppress deployment [1]. The requirements specify automatic suppression of the airbag for occupant types such as a three- or six-year-old child, rear-facing infant seat, forward-facing child seats, and booster seats, while activating the airbag for a fifth-percentile adult female or larger. While our approach is not yet able to detect the presence of a rear-facing infant seat or a forward-facing child seat or booster seat, our approach was designed to detect the presence and exact location of a head of any type of passenger, including that of a child or a fifth-percentile adult female. Our approach addresses alternative requirement of the NHTSA standard for *dynamic suppression*, where the problem is instead to detect whether the same occupants are in or out of position.

Fig. 4. Laboratory for Intelligent, Safe Automobiles-M (LISA-M) test bed for data collection and algorithm development.

IV. Design of a Multimodal Video-Capturing System, Test Beds, and Instrumented Vehicles

To provide an adaptable experimental test bed for evaluating the performance of various sensing modalities and their combination, two test environments, based upon a Daimler-Chrysler S-class [Laboratory for Intelligent, Safe Automobiles-M (LISA-M)] test frame and a Volkswagen Passat (LISA-P) vehicle, were outfitted with a computer and a multitude of cameras and acquisition systems. Of principal importance in the hardware specification and software architecture was the ability to capture and process data from all the sensor subsystems simultaneously and to provide facilities for algorithm development and offline testing. Fig. 4 shows the laboratory-based test bed. Various sensory and computing modules used in this laboratory test frame LISA-M and the instrumented automobile LISA-P include the three types of camera modules [thermal long-wavelength infrared, trinocular stereo, and color National Television Standards Committee (NTSC) cameras], synchronized video-stream-capturing hardware, and high-volume storage. LISA-P is outfitted with a power inverter to supply 120 volts alternating current (vac) power. Details of these test beds are presented later.

The LISA-P is equipped with a 120-V alternating current power-distribution system. This comprises a marine-type 1.0-kW true sine wave inverter, an auxiliary battery, and appropriate isolation circuitry. The availability of 120 vac avoids the alternative of adapting equipment to the vehicle's 12-volt direct current (vdc) supplies and yields an easier transition from laboratory to vehicle.

The computing platform consists of a commercial Xeon PC with a high-throughput disk subsystem for streaming video data. This subsystem consists of four 15-K r/min Ultra320 SCSI disk drives in a RAID0 array that achieves in excess of 220-MB/s sustained throughput. This data rate allows for the capture of several simultaneous high-resolution video streams. The computing platform allows for a good deal of processing to be done in real time as well, but normally in the course of algorithm development speed is achieved after functionality and data collection is expensive. Hence, the overriding requirement is to capture the data for efficient offline development.

The video-capture hardware currently consists of a dual CameraLink PCI-X capture board, an analog color RS-170 PCI capture board, and an IEEE 1394 PCI host controller. Additionally, a quad combiner is available to combine quarter resolution images to be acquired simultaneously using the analog capture board. The variety of acquisition devices allows for experimentation with a wide range of imaging sensors. Three video-sensing systems are being used in the experiments described in this report.

The first sensor system is a trinocular stereo system from Pt-Grey Research, which provides 2-1/2-D stereo disparity maps. This consists of three black and white 640×480 element 1/3" charge-coupled device (CCD) sensors mounted with 3.8-mm focal length lenses in a common enclosure as two stereo pairs in an isosceles right triangle. Integrated digitization, buffering circuitry, and an IEEE 1394 interface allows transfer of video from all three cameras to the host. This system is supplied with host software that performs the correspondence calculation and provides a disparity image. The nearest object that can be seen by all cameras is roughly 30 cm from the face of the enclosure, which poses some restrictions on camera placement in the automobile.

The second sensor uses a miniature 2-D thermal long-wavelength infrared sensor, Raytheon model 2000 AS. This device provides response in the LWIR spectrum (7–14 μm) and an analog video output. The camera uses a 160×120 element amorphous silicon focal plane array and lens that produces a $35^\circ \times 50^\circ$ field of view. It requires no cooling; the absence of active cooling provisions allows the sensor head/lens assembly to be quite small (~ 125 cm^3) so that it can be unobtrusively mounted on the dashboard. The camera has no absolute calibration and is subject to considerable drift in both gain and offset with temperature. It does have a mechanism for correcting per pixel gain variation, which employs a shutter that momentarily closes in front of the focal plane array every few minutes [13]. Miniaturization and cost reduction is moving at a rapid pace in LWIR cameras, with roughly a four-fold decrease in both size and price in the last two years. In selecting this device, we intend to test something representative of what may be reasonably added to a passenger car a few years from now.

The third sensing system provides 3-D imagery through shape-from-silhouette (SFS) voxel reconstruction. The hardware is comprised of four color 1/3″ CCD cameras, each producing NTSC output with 2.3-mm focal length lenses and a quad video-signal combiner to merge the four video signals into one, such that each input video frame occupies a quadrant of the output video frame.

The placement of the cameras is shown in Figs. 5 and 6. The output of all seven cameras captured synchronously and the stored images of one experiment with a male subject are shown in Fig. 7.

The software architecture was designed to allow for efficient development of multiple video-processing algorithms and their incorporation into an integrated framework. To this end, the acquisition, storage, user interface, and display functionality are modular and separate from the image-processing functionality. Two applications have been developed, one for processing and capturing live data and one for processing the captured data for offline algorithm development in the laboratory. Both applications use the same frame processor C++ object interface encapsulating video-processing algorithms. The block diagram for this basis frame processor is shown in Fig. 8. This standard interface ensures that algorithms developed separately in the laboratory can be painlessly integrated on the test bed.

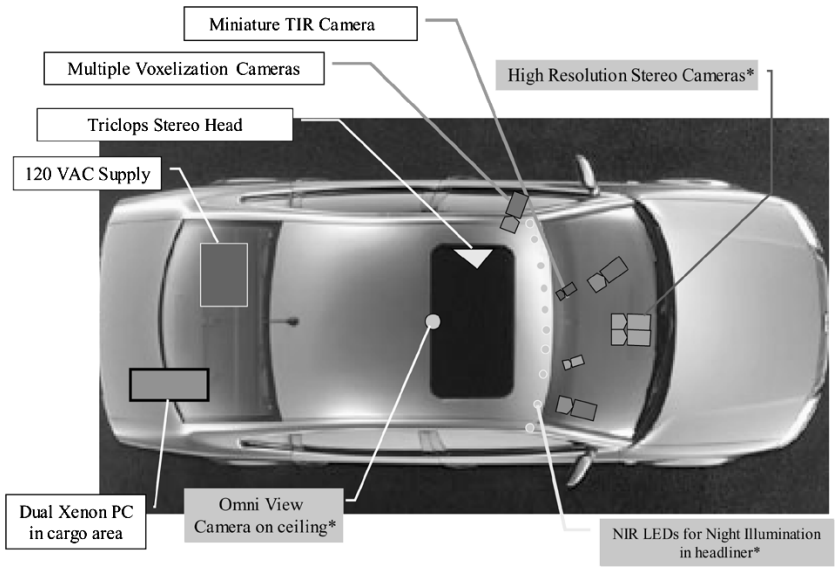

Fig. 5. Multimodal video cameras, synchronized capture, computing, storage, and power modules in the LISA-P instrumented vehicle test bed.

Fig. 6. Top left: LISA-P test bed with the power supply, PC-based capturing system. Right: Three NTSC cameras (front driver and passenger windows and front dash), thermal camera (front dash), and a stereo camera (roof rack) in view. Bottom: Test subject in LISA-P during an experiment.

Fig. 7. Example acquired image data from LISA-P. Top left, top middle, bottom left: Multiple perspective NTSC images. Bottom middle: LWIR image. Top right: One of three grayscale images for depth map calculation. Bottom right: Resulting depth image from stereo correspondence.

The live-data application is configurable at compile time to capture one or multiple streams using a digitizer-independent image-acquisition framework derived from the Microsoft Vision SDK. This approach minimizes the effort required to run the live data system on different hardware, so laboratory machines fitted with various makes and models of digitizer may be utilized for algorithm development. The live data application acquires multiple video streams in a synchronous manner, so that, for each time step, one frame from each stream is available. Therefore, the results of the various sensor modalities may be compared on a frame-by-frame basis. The processed and/or raw data is combined in a single AVI file to minimize disk seeks and written to the RAID0 array.

The captured-data application allows captured raw video streams to be played back and routed to the same frame-processor object used in the live system. Similar display functionality is provided and the identical processor specific GUI may be summoned, simulating the test-bed environment precisely

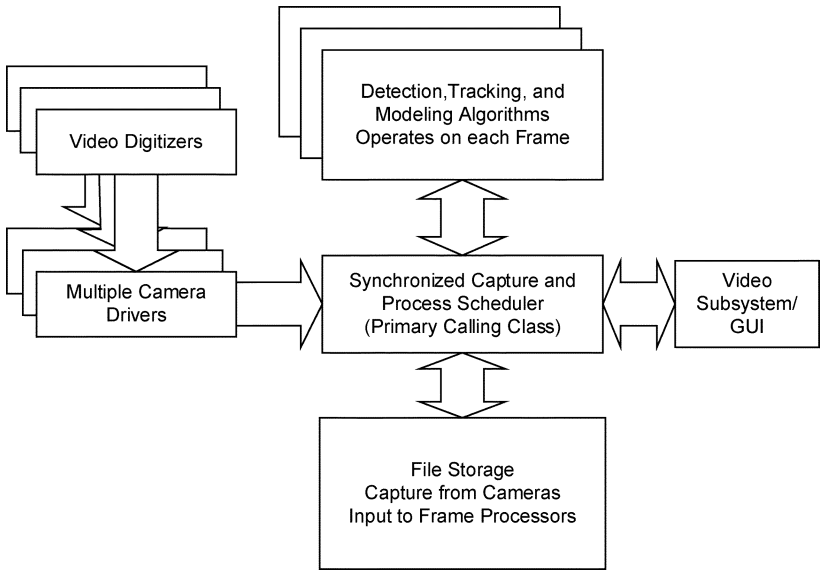

Fig. 8. Software architecture for live side-by-side processing.

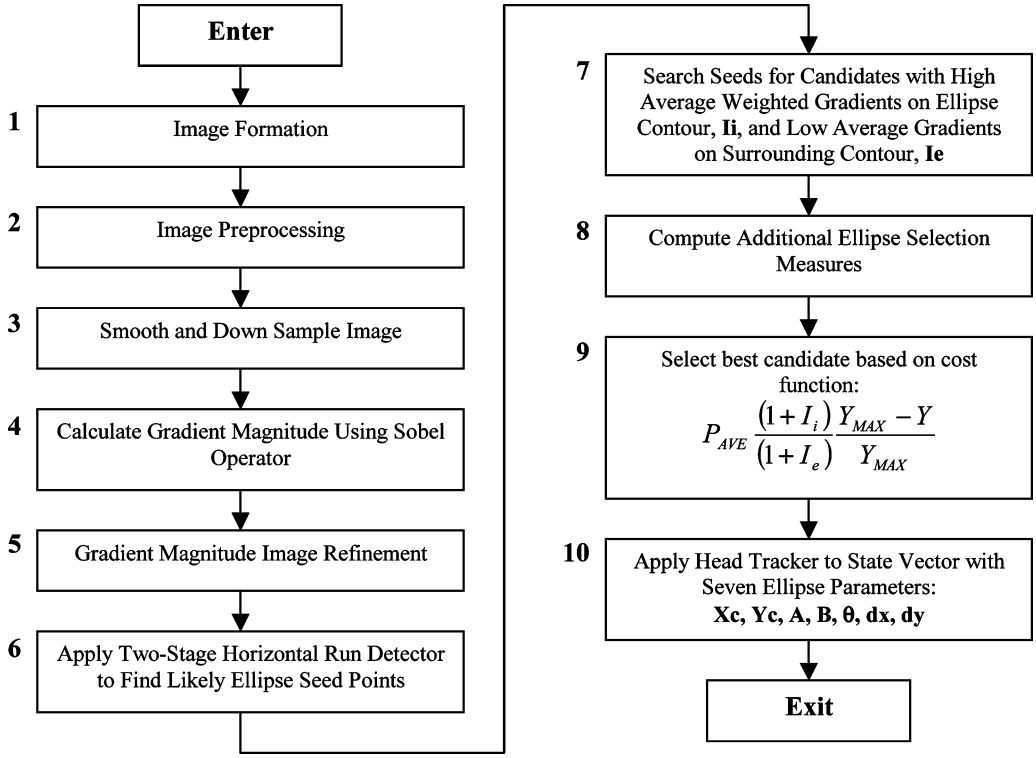

Fig. 9. Flowchart for the generic edge-based face-detection algorithm. Details of 1, 2 5, and 8 are specific to the capture modality. See the appropriate sections for descriptions of these steps for stereo and LWIR capture methods.

from the viewpoint of the frame-processor object. In this way, each frame-processing object may be developed separately offline and later incorporated into the live data application with a minimum of integration issues.

V. Head Detection and Tracking

Both the Stereo and LWIR-based head-detection algorithms are derived from Eleftheriadis and Jacquin [14], who propose an edge-based head-detection algorithm that provides head-pose and size estimates that are easily implemented in real time (in excess of 30 f/s on commercial PCs). The basic edge-based algorithm is discussed here and the specific stereo and LWIR modifications are outlined in their respective sections.

The edge-based algorithm flowchart is shown in Fig. 9 and each step is described as follows.

Step 1) For each camera, the image on which the edge-based algorithm operates is derived in a specific way. Details on the image formation are discussed in the Stereo and LWIR sections.

Step 2) For each stereo and LWIR image, specific preprocessing operations are performed before head detection is attempted. Again, these details are discussed in the Stereo and LWIR sections.

Step 3) Resultant input images are then smoothed with a Gaussian kernel and down-sampled by a factor of 4, similar to the algorithm described in [14].

Step 4) Gradient magnitude image is calculated using 1×3 Sobel operators in vertical and horizontal orientations. This is stored as a floating-point image.

Step 5) Gradient magnitude image can then be refined to eliminate gradients associated with areas that are unlikely to contain the head location. This is designed to speed up processing by reducing the seed points and also helps to reduce false detections. Specific refinements are discussed in the Stereo and LWIR sections.

Step 6) A two-stage horizontal run detector, as described in [14], is applied to the gradient magnitude image. This process consists of first applying a threshold to the gradient magnitude image to obtain a binary edge image. In our implementation, this threshold was set to 0.5 standard deviations of the gradient magnitude above the mean gradient magnitude, which gave a binary edge image of fairly constant density. This image is then examined in two stages: a coarse scan and a fine scan, as in [14]. In the coarse scan, the image is divided into 5×5 pixel blocks; the blocks that have at least one edge pixel are marked. In the fine scan, each run of consecutively marked horizontal blocks is examined. For each such run, the first scan line with an edge pixel is found. On this scan line, within this run of blocks, all pixels between the first and last edge pixels possibly correspond to the top of a head and are marked as possible seed points for ellipse contour template matching.

Step 7) A set of pairs of precalculated templates for ellipse contours and ellipse boundaries are overlaid on the gradient magnitude image from Step 5). These template pairs consist of one template that represents the contour of the ellipse and one that represents the boundary just outside the ellipse contour and are described in [14]. However, two slightly different figures of merit are used to filter ellipse candidates. In [14], the ellipse templates were applied against the binary edge image and the figures of merit were calculated by the number of edge pixels underlying the templates normalized for the size of the template, with the pixels at the top of the ellipse contour given 50% more weight. In this algorithm, the templates are applied to the gradient magnitude image itself and the average gradient magnitude underlying the template is used as a figure of merit (also with 50% extra weight given to those pixels at the top of the ellipse.) This approach yielded better fitting than justified its small impact on computation speed. Only ellipses that simultaneously exceeded a threshold on average gradient underlying the contour and were below a threshold for average gradient underlying a boundary were given further consideration.

Step 8) At this point in the processing, we have a list of ellipse candidates. Whereas in [14] this step was followed by the selection of the "best" ellipse, we found that for both stereo and LWIR images, this approach would sometimes find elliptical objects that were not body parts or elliptical contours that occurred due to the chance arrangement of several nonelliptical objects. To eliminate most of these candidates, additional selection measures for each seed point are calculated. These selection measures are specific to stereo and LWIR methods and are discussed in their respective sections.

Step 9) The best candidate ellipse is selected based on a cost function specific to stereo and LWIR. The cost function for each method shares the same basic form. We select the "best" candidate by maximizing the expression

$$P_{\mathrm{AVE}} \frac{(1+I_i)}{(1+I_e)} \frac{Y_{\mathrm{MAX}} - Y}{Y_{\mathrm{MAX}}}. \tag{1}$$

P_{AVE} is a selection measure based on the average of intensities inside the ellipse candidate. For the stereo and LWIR methods, P_{AVE} is based on different values and is discussed further in their respective sections. I_i is the mean gradient value on the contour and should be a large value. I_e is the mean gradient value on the outer contour and should have a small value. Y and Y_{MAX} are the ellipse height and maximum height in the image, respectively. These values are used to give preference to candidates higher in the image.

Step 10) To reduce the effects of measurement noise, seven ellipse parameters are tracked using a Kalman filter. The tracked parameters are the ellipse center coordinate, ellipse axes size, and inclination angle, as well as the position change from the previous detected location.

VI. Stereo-Based Head Detection and Tracking

A. Stereo-Based Head-Detection and Tracking Algorithm

The stereo-based head-detection and tracking algorithm is a modification of the edge-based head-detection algorithm described in Section V and displayed in Fig. 9. Modifications to the algorithm specific to the stereo algorithm are discussed later.

In Step 1), a background model is generated based on N frames of disparity data captured in an empty cabin. Using the disparity image in the background model allows for a larger degree of lighting changes than using the reflectance images. Once the background model is obtained, the current foreground data is computed. The SVS API can give high-valued noisy response in regions where the stereo is incomputable, so disparity values that are too high are removed from the current image. Disparity

Fig. 10. Example reflectance and disparity images for a subject in the LISA-P test bed.

TABLE I
STEREO HEAD-DETECTION AND TRACKING RESULTS

Occupant	Correct Detected Location	% Correct Detected	Correct Tracked Location	% Correct Tracked
Male 1, 5' 8", Average Build	2830	95.3%	2832	95.4%
Male 2, 5'9", Average Build	5374	94.8%	5378	94.9%
Female 1, 5'0", Petite Build	5287	97.9%	5310	98.3%
Female 2, 5'8", Average Build	3878	94.1%	3876	94.0%
Female 3, 5'11", Average Build	2945	99.8%	2946	99.8%
All Occupants	20314	96.3%	90342	96.4%

values that fall outside the car, specifically those in the car's window region, are also removed from the current disparity image. This refinement helps to remove extraneous and invalid foreground data. After this threshold is applied, a foreground map is generated through background subtraction. An example of the raw reflectance image and its corresponding depth map are shown in Fig. 10.

In Step 2), to eliminate more of the extraneous stereo data, a median filter operation is performed, followed by a morphological opening. Then, connected component analysis removes all areas smaller than the minimum head size. This final binary foreground map is combined by a logical AND with the current disparity image. The result is the current foreground disparity image. The input images are the foreground disparity image and the corresponding reflectance image.

In Step 5), the gradient magnitude image is computed using the reflectance image and only those values that fall within the current disparity-based foreground map are kept.

In Step 7), the additional selection measure computed is the average normalized disparity value within the bounding box enclosing the candidate ellipse. The inclusion of this measure is meant to ensure that the ellipse location corresponds to a region of valid and consistent depth. The average normalized disparity value is denoted as P_{AVE} in (1).

B. Performance Evaluation

The performance of the stereo algorithm was evaluated on a data set consisting of 21 379 frames of video acquired in a moving automobile with five different passenger subjects. Each subject was asked to assume several poses, listed in Table I.

Fig. 11. Examples of successful stereo-based head detection and tracking of an occupant in four different situations.

TABLE II
LWIR HEAD-DETECTION AND TRACKING RESULTS

Occupant	Correct Detected Location	% Correct Detected	Correct Tracked Location	% Correct Tracked
Baby	415	92.2%	401	89.3%
Male 1, 6'2", Large Build	555	92.5%	547	91.2%
Male 2., 5'8" Average Build	2688	90.5%	2678	90.2%
Female 1, 5'2", Average Build	425	94.4%	428	95.1%
Female 2, 5'1", Petite Build	492	83.8%	483	82.3%
Female 3, 5'8", Average Build	3761	91.2%	3746	90.9%
Female 4, 5'11", Average Build	2656	90.0%	2646	89.6%
All Occupants	10992	90.7%	10929	90.1%

Head-detection results were considered to be correct if the algorithm placed an ellipse center point somewhere on the subject's head. For a successful detection, the estimated ellipse size needed to be comparable to occupant's head size. Examples of successful detection are shown in Fig. 11 and the detected location is denoted by the white circle and corresponding center cross.

Similarly, head tracking results were considered to be correct if the algorithm placed the tracked ellipse center point somewhere on the subject's head, denoted in the examples by the dark gray cross. Results of stereo head detection and tracking are listed in Table II.

With an overall detection rate of 96.4%, this head-detection algorithm is very successful. There are however, certain instances in which the algorithm could be improved. The detection rates drop when the occupant leans forward or to her left. These drops can be explained by the nature of the stereo data and the camera setup.

Fig. 12. Examples of unsuccessful head detection due to invalid disparity data because of camera limitations.

Fig. 13. Examples of unsuccessful head detection due to competing elliptical objects in the scene.

When using the SVS API, the left-most 64 columns are invalidated in the disparity image, as it is the value of the disparity search range used in this test. This corresponds to the area in the left image that is presumed to not overlap with the right image. If the head falls in or near this region, denoted by the white vertical line in Figs. 11–13, the invalid stereo data region can distort the head contour, causing detection problems.

Similarly, if the occupant leans too close to the camera, the head can fall out of the range of valid disparities. The SVS software will return invalid data in that area where the head exceeds the minimum distance from the camera where stereo data is computable. Naturally, this invalid data can cause problems detecting the head. These out-of-position errors are indicated in Fig. 12.

These errors are due entirely to camera selection and placement. This can be resolved easily by selecting a camera with a field of view and baseline that contains the entire range of potential occupant positions. Under the current setup, the Digiclops' enclosure size and baseline limits the camera positioning options and the position selected for these tests is the most optimal. These errors directly contribute to lower results for leaning forward and left.

The largest falloffs in detection rates occur in the hand-motion and object tests. Specifically, when the occupants put on or remove the hat or put their hands on their face, the detection rates drop significantly. This occurs because other objects in the cabin, namely the hat and hands, look similar to the head in the disparity image and may maximize the detector cost function better than the head location for that frame. Examples of these errors are in Fig. 13.

These competing object errors are critical since the head is not only being detected incorrectly, but there also may be potentially dangerous foreign objects in the scene that could cause further harm in the event of an improper airbag deployment.

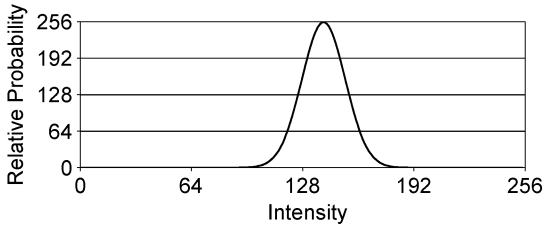

Fig. 14. Gaussian intensity to probability of skin-temperature mapping. The appropriate mean of the intensity map is chosen on a per-session basis.

C. Enhancements

Further processing is necessary to remedy the beforementioned errors. It is not enough to only search for the head in the current proposed manner. To further validate the choice of head position, subfacial cues such as eyes and noses could be searched for inside of the ellipse area. This would help to eliminate areas in which the stereo data presents a valid option for a head location, since the reflectance data would invalidate it.

It may also prove to be useful to classify all the stereo-disparity data as in a safe or unsafe position relative to the real-world automobile cockpit. This would allow for validation of the occupant's safety by a method in combination with the head location and would help to make correct decisions in situations where foreign objects are in the scene. This is critical because when foreign objects are in the scene, decision errors could occur both when the head location is detected correctly (by using the correct head location to decide that it is safe when the foreign object could cause injury) and incorrectly (by detecting the head location in the incorrect position, thereby giving an inaccurate decision). In [15], methods are proposed to extend the stereo-detection algorithm to include other body parts and foreign-object detection.

Despite the potential for these detection errors, the overall error rate is very low. The errors also seem to occur in short isolated bursts. Most of the time, errors occur in a single frame or two and are corrected in the next frame. The time for each error is a fraction of a second. Considering that the Kalman tracked head location is often correct when these detection errors occur, the time when both the detector and tracker are wrong is even smaller.

VII. Thermal Infrared-Based Head Detection and Tracking

A. Face Detection and Tracking Based on Face and Background Contrast in LWIR Signatures

This algorithm is based on the edge-based head-detection algorithm described in Section V and displayed in Fig. 9. Modifications to the algorithm specific to the LWIR algorithm are discussed below.

In Step 1), the eight-bit per pixel intensity image is remapped to approximate the probability of membership in the human skin class from LWIR intensity based on [16], using a simple Gaussian probability density function (pdf) shown in Fig. 14, with the mean and variance manually, empirically set. This is implemented using a precalculated lookup table and yields an eight-bit output image with pixels valued 255 most likely to

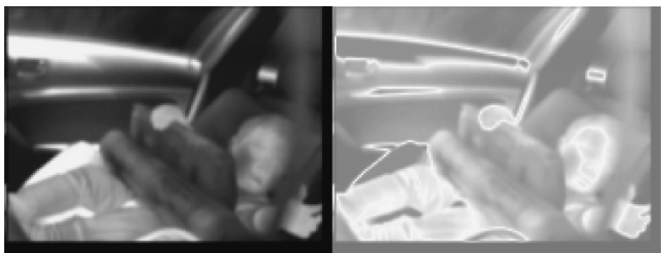

Fig. 15. LWIR images before and after intensity to probability of skin-temperature mapping. Note that areas containing human skin are now bright relative to other objects in the scene.

be skin and pixels valued 0 least likely. The probability lookup table $\mathrm{P}(I)$ is populated with values given by

$$\mathrm{P}(I) = 255\frac{(I-\mu)^2}{2\sigma^2} \tag{2}$$

where intensity $I = 0, \ldots, 255$, mean μ, and variance σ^2 are manually set on a per-run basis.

In Step 2), the probability image obtained from Step 1), shown in Fig. 15, is subjected to an iteration of grayscale erosion where each pixel is replaced by the minimum of its eight neighborhood. This eliminates the consideration of one-pixel-wide edges in the probability image caused by the boundaries of hot objects passing rapidly through the human skin-emission region.

In Step 5), no further gradient refinement is performed. In Step 7), to refine the list of these candidates, the average skin probability of each ellipsis's bounding box is calculated by summing the pixel values in the image from Step 2). This quantity is referred to as P_{AVE}. Only those candidates exceeding a threshold for this average probability are considered further.

B. Thermal Infrared Performance Evaluation

The performance of the LWIR algorithm was evaluated on a data set consisting of 12 132 frames of video acquired in a moving automobile with seven different passenger subjects. Each subject was asked to assume poses, as listed in Table I.

Results were deemed correct if the algorithm placed a head center point somewhere on the subject's head or neck if a head was present in the image or if the algorithm failed to find a head if none was present. It is not always possible to accurately determine the jawline in the LWIR imagery. For the purpose of head tracking in the context of intelligent airbag deployment, the neck/head position is a sufficiently accurate indicator. Failures by the algorithm were further classified into categories of the head not being found although present, the head being found in the wrong location, or a head being found when no head was present. A summary of LWIR head-detection and tracking results is listed in Table III.

The LWIR algorithm's performance is quite robust, with an average of 90% accuracy. It too suffered the occasional failure by misidentifying a hand as a head, but only when the hand's aspect was quite elliptical (e.g., palm out, fingers not spread). It rarely found some other body part, usually only for a frame or two in sequence. This usually occurs when another part of the person's body emitted the same LWIR characteristics as the head. Such examples occur when the subject's pants are highly emissive and allow for much of the body heat to escape.

TABLE III
OCCUPANT TEST SCRIPT

Test	Occupant Task
Position Test	Enter car and sit normally
	Lean halfway forward
	Lean completely forward
	Return to normal position
	Lean back completely
	Return to normal position
	Lean to right against window
	Lean to left towards driver
	Return to Normal Position
Hand Motion & Object Test	Move Hands about cabin
	Open the glove box
	Put hands on face, stretch
	Adjust car radio
	Place hat in lap
	Move about cabin while wearing hat
	Remove hat
	Place feet on dashboard
Free Motion Test	Subject is free to move about cabin as they wish

The algorithm would sometimes fail to find the head when a head was present. This happens most often during rapid head movements. The nature of the thermal camera's focal plane array is such that the pixels' time constants are on the order of the frame rate. This causes an afterimage and, therefore, motion blur. Since the algorithm requires a sharp edge (high gradient magnitude) in its very first, coarse scanning, the appropriate seed points for the subsequent ellipse fitting are never found. On several occasions, with a strong wind blowing in the open window at roughly 35 mi/h, the passenger's skin temperature was lowered to the point where it was momentarily cooled and classified as background.

The fact that the failures of the algorithm were only occasional and usually consisted of a sequence of few wildly incorrect frames (for instance, finding a hand) among a series of a hundred or more correct frames caused the tracker to degrade the performance of the system as a whole. Although it did improve the stability of the head location during those sequences where the head was found correctly, a single result far from the true head location would require a few frames for the Tracker to settle on the new position, whereas the raw reading was correct in the very next frame. The Kalman filter used in the head tracker is designed to model a system with Gaussian process and measurement noise, whereas this noise is of an impulse flavor. It is probably appropriate to apply some nonlinear temporal filter (i.e., a median filter) to the sequence of head-position readings before they are incorporated into the tracker estimate.

C. Enhancements

The performance of the LWIR algorithm is quite good, but it does occasionally suffer from a preference to body parts other than the head. In these cases, an ellipse candidate was usually found for the head, but scored lower than the hand when the final selection of the best ellipse was made. One such approach is to search the candidate ellipses for one or more subfeatures that one would expect to find on a face, but not on a hand. Eyes,

Fig. 16. Inferring head depth from ellipse size estimation in LWIR images.

ears, and mouths show up quite distinctively in LWIR imagery, although they appear to be completely different than in visible spectrum images.

The algorithm may also benefit from calculating P_{AVE} as the estimated probability underlying the ellipse, as opposed to that underlying the ellipse's bounding box, at the cost of slightly more computational complexity.

The LWIR camera also exhibits some undesired characteristics. Currently, the LWIR camera sensors have a nonstationary skin temperature-to-intensity mapping. Over time, the intensity for skin changes in the LWIR image. This is a problem for the LWIR algorithm, which relies on a skin intensity pdf that becomes invalid if the intensity mapping changes. An adaptive recalibration of the intensity mapping would rectify this problem.

The size of the ellipse may be used to get a rough depth from size estimate, although currently the head is sometimes found from the top of the head to the chin and sometimes found from the top of the head to the neck, yielding two distributions of head size. If we can robustly measure one or the other, then the head-image size data might be combined with switches in the seatback to give a reference head image size for the passenger sitting fully back. It would then be possible to give an estimate of the passenger distance from the camera (and, hence, the dashboard) by comparing the head image size with that in the reference position. An illustration of inferring depth information from the ellipse size in LWIR images is shown in Fig. 16.

VIII. Stereo and LWIR Head-Detection and Tracking Algorithms: Comparative Analysis

Experiments were performed using the LISA-P test bed. A series of experiments were conducted in which data was captured for both the stereo and LWIR methods simultaneously. The desire is to have a direct comparison of the two-head tracking methods on a frame-by-frame basis.

The tests were conducted on three nonconsecutive mornings from 8:00 to 11:00 AM using the LISA-P test bed. The stereo camera was placed on the driver's side roof rack, looking down on the passenger's seating area. Left and right images we captured at 320×240 resolution at 15 f/s. Stereo data was computed using SRI's Small Vision System API. For these tests, the LWIR camera was next to the stereo camera in the same orientation, so that approximate depth from the dashboard could be inferred. Infrared data was captured simultaneously at a resolution of 640×480, also at 15 f/s. Each captured stereo image has a corresponding and comparable LWIR image. Test subjects were asked to enter the vehicle and perform a series of movements and tasks while the car was driven at road speeds.

The occupant script listed in Table III was divided into a position test, which tests the algorithm's ability to detect the head at various positions in the cabin and to track the head's movement; a hand-motion and object test, which was designed to evaluate the algorithm's robustness to competing objects and hand motion in the scene; and a free motion test, which was designed to catch other potential situations for detection error, as the subject was free to move as they wish during this test. Example images of each of the occupant script poses are shown in Fig. 17. In all, stereo and LWIR data was collected with three subjects for a total of 10 045 frames of collected data.

Ideally, a method of comparison would consist of answering the following question: how well does a system make decisions on the most critical occupant sizes, positions, and poses, at the same time addressing the less critical airbag deployment conditions? In the standard issued by the NHTSA [1], it states that airbag systems must be able to disable itself when occupants of five standardized sizes are in various positions. This is the so-called *dynamic suppression test*. One way is to associate with each occupant size and position the cost for making the erroneous decision and having the airbag deploy or not deploy by mistake. For every system, there is an associated miss-detection rate. False alarms are not considered, since there is a comprehensive set of occupant sizes and positions. An acceptable cost function would be the sum of false-alarm and miss-detection rates weighted by the cost of making an erroneous classification for each occupant size and position. This "likelihood of safe deployment" is the measure of goodness for the system under test.

In the stereo and LWIR tests, the distance from the detected head to the dash is the only variable to be considered when classifying a person to be in or out of position. Since all subjects were normal-sized adults, the other variable, occupant size, is constant. Correct head detection, therefore, is equivalent to a valid distance estimate between the dashboard and the head, within some distance variance. Table IV lists the detection rates for the stereo and LWIR head-detection and tracking algorithms.

Fig. 17. Example images of occupant script poses. From top left: sitting normally, leaning halfway, leaning completely forward, leaning back, leaning right, leaning left, moving hands about cabin, opening glove box, hands on face, stretching, adjusting radio, hat in lap, putting on hat, moving while wearing hat, removing hat, feet on dashboard.

TABLE IV
STEREO AND LWIR HEAD-DETECTION AND TRACKING COMPARISON

Occupant Task	Male 1, 5'8"		Female 1, 5'8"		Female 2, 5'11"		All Occupants	
	Stereo	LWIR	Stereo	LWIR	Stereo	LWIR	Stereo	LWIR
Sit Normal	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Lean Halfway	100.0%	73.0%	100.0%	92.9%	X	X	100.0%	82.8%
Lean Forward	76.4%	0.9%	X	X	X	X	76.4%	0.9%
Return to Normal 1	100.0%	95.9%	98.0%	98.0%	100.0%	100.0%	99.6%	97.4%
Lean Back	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Return to Normal 2	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Lean Right	100.0%	52.1%	100.0%	100.0%	97.8%	96.7%	99.1%	92.1%
Lean Left	100.0%	98.9%	X	X	97.7%	100.0%	98.4%	99.7%
Return to Normal 3	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Position Test Totals	97.3%	80.3%	99.8%	98.7%	98.7%	99.1%	98.4%	91.7%
(Number of Frames)	(940)	(776)	(537)	(531)	(676)	(679)	(2153)	(1986)
Move Hands about cabin	78.1%	100.0%	100.0%	97.4%	97.8%	99.1%	91.6%	99.2%
Open the glove box	100.0%	100.0%	100.0%	95.5%	74.3%	97.6%	91.2%	97.8%
Put hands on face & stretch	81.7%	100.0%	100.0%	85.2%	87.8%	89.4%	90.0%	91.3%
Adjust car radio	100.0%	100.0%	100.0%	100.0%	99.4%	100.0%	99.8%	100.0%
Place hat in lap	100.0%	100.0%	100.0%	97.5%	100.0%	97.7%	100.0%	97.9%
Put hat on head	90.0%	84.3%	90.5%	35.7%	100.0%	93.3%	95.2%	85.2%
Move with hat	98.8%	87.9%	100.0%	68.3%	92.6%	62.8%	96.5%	71.0%
Remove Hat	100.0%	100.0%	100.0%	62.1%	100.0%	100.0%	100.0%	94.9%
Feet on Dashboard	100.0%	94.5%	100.0%	76.4%	93.9%	100.0%	98.3%	87.3%
Hand Motion & Object Test Totals	92.6%	97.4%	99.8%	85.7%	92.0%	90.5%	94.8%	90.9%
(Number of Frames)	(1399)	(1471)	(1939)	(1665)	(2258)	(2221)	(5596)	(5357)
Free Motion Test	100.0%	87.4%	99.8%	95.5%	95.8%	86.1%	97.9%	88.9%
(Number of Frames)	(493)	(431)	(470)	(450)	(942)	(846)	(1905)	(1727)
All Test Totals	95.4%	90.2%	99.8%	89.6%	94.0%	90.9%	96.2%	90.3%
(Number of Frames)	(2832)	(2678)	(2946)	(2646)	(3876)	(3746)	(9654)	(9070)

X denotes that the subject moved out of the camera frame for this test, and the results were invalid.

Fig. 18. Four perspectives of the occupant inside LISA-M and the silhouettes generated from background subtraction with shadow removal.

IX. Discussion

A. Stereo and LWIR Head-Detection and Tracking-Evaluation Summary

This test setup is particularly unique in that it allows for a direct comparison of the stereo and LWIR head-detection methods on a frame-by-frame basis. The test bed allows for the simultaneous capture of multiple inputs from differing modalities all synchronized to the same clock. The test bed is also set up to allow for easy modularization, so that both new cameras and new algorithms can be introduced into the system and tested with minimal further development. As both camera and algorithm advancement are anticipated, the test bed gives an ideal circumstance to extend this research.

Both algorithms achieve a high average accuracy in detecting and tracking the head. At success rates of 96.4% and 90.1%, respectively, for various occupant types, it can be concluded that both the stereo and LWIR systems can be used to robustly detect the location of the head with high reliability. The resulting head-location information can help to decide the manner in which an airbag should be deployed.

Although the algorithms achieve a high success rate, both algorithms suffer from certain limitations that should be resolved through further research.

Stereo methods outperformed LWIR methods when the subject moved with a hat. Because the head-detection algorithm is searching for elliptical objects, the detection rate decreases in the LWIR case, because the hat changes the emissive properties of the subject's head, making it less elliptical. Conversely, in the stereo case, we compute edges based on the reflectance data, in which the occupant head looks similarly elliptical with or without a hat.

Stereo methods also outperformed LWIR methods when the subject leans completely forward. This, however, is not a function of the occupant's position, but rather due to the subject's head being turned from the camera so that only the back of the head was visible. Naturally, the occupant's hair does not have the same temperature as the face and resulted in the low detection rates. This also was the case for the low results for the lean right test for Male 1. This indicates that modeling only skin temperature may not be enough and that other occupant properties, such as hair and clothing, should be taken into account.

LWIR methods outperform stereo methods when dealing with competing elliptical objects in the scene, especially hands. This is because the skin temperature of the hands is usually different enough from the face as not to confuse the LWIR detector. However, if the hands are of a similar size and depth as the head, the stereo detector can give erroneous results. Potential solutions include using subfacial cues to verify ellipse candidates, as well as introducing more sophisticated body modeling to account for hands and arms in the scene.

Despite the success of these initial tests, further testing is imperative. This test of the stereo and LWIR systems included only three subjects at a particular time of day in fair weather. Clearly, many different subjects need to be tested on the system. It is still untested how well the algorithms will perform when subjects have features such as facial hair, large hats, are eating or drinking, are very large or very small, or are sitting in unconventional positions. The algorithms are also untested in driving conditions other than a sunny day. Although an exhaustive test of the permutations of subject type and driving condition is impractical, a much larger and extensive test of these variations is necessary to deem the algorithms reliable enough for commercial use.

B. Extension Using SFS Voxel Data

An extension to the problem of designing a reliable occupant posture estimation system using vision-based techniques is our investigation of using SFS voxel data with the multicamera setup to extract occupant posture information. SFS is a technique that reconstructs the visual hull of the occupant from the occupant's silhouette images. Visual hull is the closest reproduction of an object's actual volume using an arbitrary number of silhouette images of the object [17]. Small [18] envisioned a real-time algorithm for the computation of the visual hull using volume elements or voxels.

Four images are collected from various perspectives around the occupant inside the LISA-M test bed. These images are shown in Fig. 18. Given camera-calibration parameters and the silhouette images, the resulting voxel reconstruction of an occupant is shown in Fig. 19. The space of the voxels are demarcated by the IP, OOP, and COOP and voxels are colored green, yellow, and red to illustrate which voxels are within these regions.

It is not difficult, however, to envision a decision scheme based on occupancy information of the voxels alone to decide a person to be IP. It furthers the case when head and torso positions are known. It has been shown that the head and torso can be found from the voxel data of occupants of various sizes with consistent regularity [19]. The key components for that voxel-based occupant posture estimation system are camera placement, camera calibration, silhouette generation, voxel reconstruction, and body modeling from voxel data.

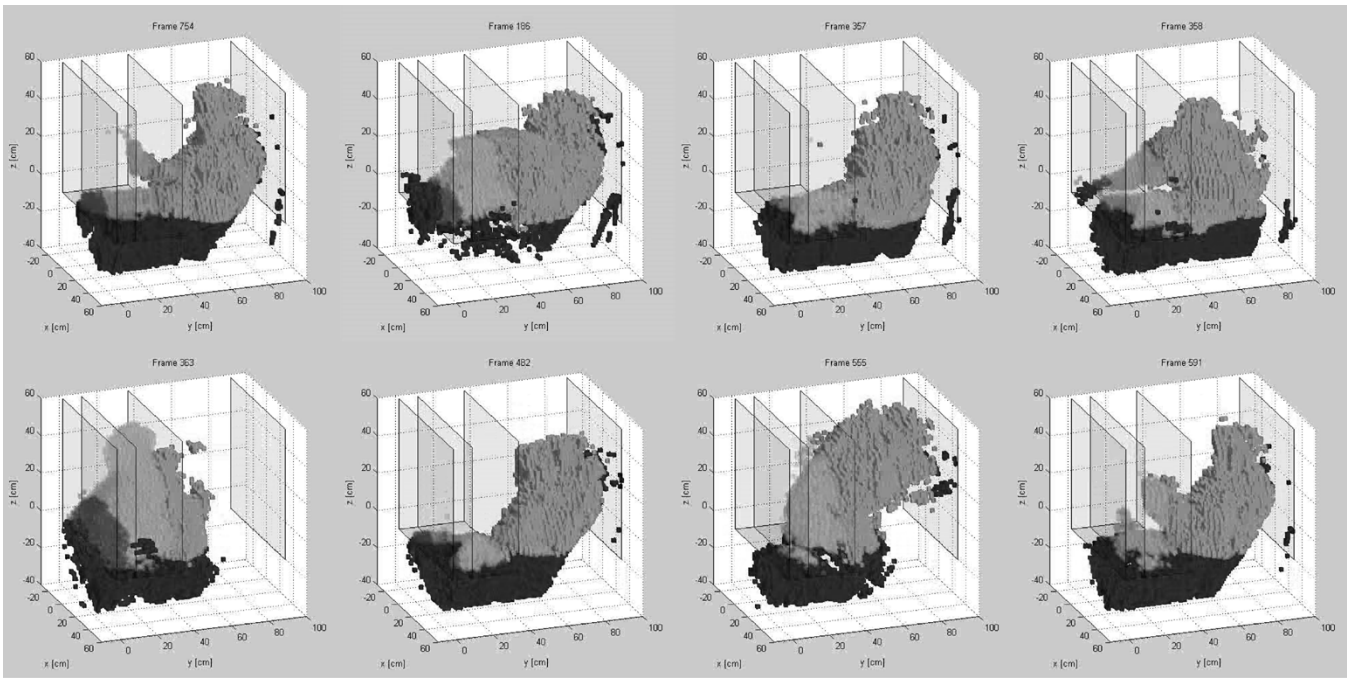

Fig. 19. Illustration of the boundaries based on IP, OOP, and COOP regions.

Currently, to generate silhouette images, a statistical background subtraction technique is employed. This technique with no further modifications is an inadequate image-segmentation method for inside the car. Alternatives are currently being investigated.

However, the uses of a full or even partial body can be undeniably far reaching. The potential applications goes beyond occupant posture estimation for the purpose of "smart airbag" deployment and enters driver fatigue analysis, driver attentiveness, and human–machine interfaces inside the car.

X. Conclusion

A new generation of airbags will incorporate information about the position and posture of the occupant in making deploy normally, deploy with limited power, or to suppress deployment decisions. Development of vision-based systems for human posture analysis in an accurate and robust manner within the challenging constraints of speed and of the automobile interior was the main research issue addressed in this paper. This paper presented the systematic investigation, development, and evaluation of using thermal long-wavelength infrared, stereo, and multicamera video-based real-time vision systems. This work involved the design of several test beds, including instrumented vehicles for systematic experimental studies that allow independent and comparative system evaluation in automobiles. Results of extensive experimental trials suggest basic feasibility of video-based occupant position and posture analysis.

Acknowledgment

The authors would like to acknowledge valuable input and assistance from Dr. K. Schaff and Dr. A. Stoschek of VW Research Laboratory, Palo Alto, CA. They are also thankful for the assistance and support of their colleagues from the University of California, San Diego, Computer Vision and Robotics Research Laboratory, especially Dr. I. Mikic, T. Schoenmackers, K. Huang, R. Chang, P. Putthividhya, and J. Wu. Finally, they would like to thank the anonymous reviewers for their insightful comments.

References

[1] National Highway Transportation and Safety Administration, "Occupant Crash Protection Standard," Federal Motor Vehicle Safety Standard 208, Oct. 2002.

[2] K. S. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," presented at the IEEE Intelligent Vehicle Symp., June 2003.

[3] A. Dhua, F. Cutu, R. Hammoud, and S. J. Kiselewich, "Triangulation based technique for efficient stereo computation in infrared images," presented at the IEEE Intelligent Vehicle Symp., June 2003.

[4] C. Koch, T. J. Ellis, and A. Georgiadis, "Real-time occupant classification in high dynamic range environments," in *Proc. IEEE Intelligent Vehicle Symp.*, vol. 2, June 2002, pp. 284–291.

[5] J.-M. Lequellec, F. Lerasle, and S. Boverie, "Car cockpit 3D reconstruction by a structured light sensor," presented at the IEEE Intelligent Vehicles Symp., Oct. 2000.

[6] F. Lerasle, J. M. Lequellec, and M. Devy, "Relaxation vs. maximal cliques search for projected beams labeling in a structured light sensor," presented at the 15th Int. Conf. Pattern Recognition, vol. 1, Sept. 2000.

[7] H. Nanda and K. Fujimura, "Illumination invariant head pose estimation using single camera," presented at the IEEE Intelligent Vehicle Symp., June 2003.

[8] Y. Owechko, N. Srinivasa, S. Medasani, and R. Boscolo, "Vision-based fusion system for smart airbag applications," in *Proc. IEEE Intelligent Vehicle Symp.*, vol. 1, June 2002, pp. 245–250.

[9] P. Faber, "Seat occupation detection inside vehicles," presented at the 4th IEEE Southwest Symp. Image Analysis and Interpretation, Apr. 2000.

[10] J. Krumm and G. Kirk, "Video occupant detection for airbag deployment," presented at the IEEE Workshop Applications of Computer Vision, Oct. 1998.

[11] R. Reyna, A. Giralt, and D. Esteve, "Head detection inside vehicles with a modified SVM for safer airbags," presented at the IEEE Conf. Intelligent Transportation Systems, Aug. 2001.

[12] M. Devy, A. Giralt, and A. Marin-Hernandez, "Detection and classification of passenger seat occupancy using stereovision," presented at the IEEE Intelligent Vehicles Symp., Oct. 2000.

[13] M. E. Farmer and A. K. Jain, "Occupant classification system for automotive airbag suppression," presented at the IEEE Conf. Computer Vision and Pattern Recognition, Madison, WI, 2003.
[14] A. Eleftheriadis and A. Jacquin, "Face location detection for model-assisted rate control in H.261-compatible coding of video," *Signal Process.*, vol. 7, no. 4–6, pp. 435–455, 1995.
[15] S. J. Krotosky and M. M. Trivedi, "Occupant posture analysis using reflectance and stereo images for 'smart' airbag deployment," presented at the IEEE Intelligent Vehicles Symp., June 2004.
[16] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland, "Illumination invariant face recognition using thermal infrared imagery," presented at the IEEE Conf. Computer Vision and Pattern Recognition, Dec. 2001.
[17] A. Laurentini, "How many 2D silhouettes does it take to reconstruct a 3D object?," *Comp. Vision Image Understand.*, vol. 67, no. 1, pp. 81–89, 1997.
[18] D. E. Small, "Real-time shape-from-silhouette," M.S. thesis, Dept. Comp. Sci., Univ. New Mexico, Albuquerque, June 2002.
[19] S. Y. Cheng and M. M. Trivedi, "Human posture estimation using voxel data for "smart" airbag systems: Issues and framework," presented at the IEEE Intelligent Vehicles Symp., June 2004.
[20] C. Eveland, D. Socolinsky, and L. Wolff, "Tracking human faces in infrared," presented at the CVPR Workshop Computer Vision Beyond the Visible Spectrum, Kauai, HI, Dec. 2001.
[21] E. Hjelmas and B. K. Low, "Face detection: A survey," *Comp. Vision Image Understand.*, vol. 83, no. 3, pp. 236–274, 2001.
[22] I. Mikic and M. M. Trivedi, "Vehicle occupant posture analysis using voxel data," presented at the 9th World Congr. Intelligent Transport Systems, Oct. 2002.
[23] T. Schoenmackers and M. M. Trivedi, "Real-time stereo-based vehicle occupant posture determination for intelligent airbag deployment," presented at the IEEE Intelligent Vehicle Symp., June 2003.
[24] I. Mikic and M. M. Trivedi, "Articulated body posture estimation from multi-camera voxel data," presented at the IEEE Conf. Computer Vision Pattern Recognition, Dec. 2001.
[25] I. Mikic, M. M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *Int. J. Comput. Vision*, vol. 5, no. 3, pp. 199–223, 2003.
[26] G. K. M. Cheung and T. Kanade, "A real-time system for robust 3D voxel reconstruction of human motions," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, vol. 2, June 2000, pp. 714–720.
[27] M. H. Henry, S. Lakshmanan, and P. Watta, "Optical flow preprocessing for pose classification and transition recognition using class-specific principle component analysis," presented at the IEEE Intelligent Vehicle Symp., June 2003.
[28] H. Z. Tan, L. A. Slivovsky, and A. Pentland, "A sensing chair using pressure distribution sensors," *IEEE/ASME Trans. Mechatron.*, vol. 6, pp. 261–268, Sept. 2001.

Mohan Manubhai Trivedi (S'76–M'79–SM'86) was born on October 4, 1953, in Wardha, India. He received the B.E. (Honors) degree in electronics from the Birla Institute of Technology and Science, Pilani, India, in 1974 and the M.E. and Ph.D. degrees in electrical engineering from Utah State University in 1976 and 1979, respectively.

He is a Professor of Electrical and Computer Engineering at the University of California, San Diego (UCSD), La Jolla. He has a broad range of research interests in the areas of intelligent systems, computer vision, intelligent (smart) environments, intelligent vehicles and transportation systems, and human–machine interfaces areas. He established the Computer Vision and Robotics Research Laboratory, UCSD. Currently, he and his team are pursuing systems-oriented research in distributed video arrays and active vision, omnidirectional vision, human body modeling and movement analysis, face and affect analysis, and intelligent vehicles and interactive public spaces. He serves on the Executive Committee of the California Institute for Telecommunication and Information Technologies [Cal-(IT) 2] as Leader of the Intelligent Transportation and Telematics Layer, UCSD. He also serves as a Charter Member of the Executive Committee of the University of California System-Wide Digital Media Innovation Program (DiMI). He regularly serves as a consultant to industry and government agencies in the U.S. and abroad. He was Editor-in-Chief of the *Machine Vision and Applications Journal* from 1997 to 2003. He has served on the editorial boards of journals and program committees of several major conferences.

Dr. Trivedi is a Fellow of the International Society for Optical Engineering (SPIE). He served as Chairman of the Robotics Technical Committee, IEEE Computer Society. He was elected to serve on the Administrative committee (BoG) of the IEEE Systems, Man and Cybernetics Society and has received the Distinguished Alumnus Award from Utah State University, Pioneer Award (Technical Activities), and Meritorious Service Award from the IEEE Computer Society.

Shinko Yuanhsien Cheng (S'98) was born in Taipei, Taiwan, R.O.C. in 1978. He received the B.S. and M.S. degrees in electrical engineering from the University of California at San Diego, La Jolla, in 2001 and 2003, respectively, where he is currently working toward the Ph.D. degree in computer vision.

He was with the Radio Systems Division, Northrop Grumman, Rancho Bernardo, CA, as a Systems Engineer for two summers, in 2001 and 2002. In the summer of 2000, he was an Engineering Program Manager Intern at Electronics for Imaging Inc., Foster City, CA. He currently is a Graduate Student Researcher at the Computer Vision and Robotics Research Laboratory, San Diego, CA. His research interests include intelligent systems, image processing, computer vision, statistical shape modeling, and pattern recognition.

Edwin Malcolm Clayton Childers received the B.S.M.E degree in machine design from the University of Virginia, Charlottesville, in 1988 and the M.S.E.E degree in intelligent systems, robotics, and control from the University of California at San Diego (UCSD), La Jolla, in 2003.

He has spent the majority of his career consulting for various concerns in the San Diego technical community, in the fields of medical and industrial instrument and product and process design, in which he is presently employed. He has also held the positions of Machine Design Engineer at Jamar Medical Systems, Senior Mechanical Engineer at Four Pi Systems, Member of Technical Staff at Hewlett Packard, and Director of Systems Engineering at Fiberyard, all in San Diego. He held several fellowships and teaching assistantships during his graduate work at UCSD.

Stephen Justin Krotosky (S'05) was born in Vineland, NJ, in 1979. He received the B.C.E. degree from the University of Delaware, Newark, in 2001 and the M.S.E.E. degree from the University of California at San Diego (UCSD), La Jolla, in 2004, where he is currently working toward the Ph.D. degree, focusing his research on computer vision.