

## Video Occupant Detection for Airbag Deployment

John Krumm and Greg Kirk  
Intelligent Systems & Robotics Center  
Sandia National Laboratories  
Albuquerque, NM 87185



**Figure 1: Empty, infant, and occupied seats as seen by camera inside vehicle. Airbag should not deploy on empty nor infant seat.**

### Abstract

When an airbag deploys on a rear-facing infant seat, it can injure or kill the infant. When an airbag deploys on an empty seat, the airbag and the money to replace it are wasted. We have shown that video images can be used to determine whether or not to deploy the passenger-side airbag in a crash. Images of the passenger seat, taken from a video camera mounted inside the vehicle, can be used to classify the seat as either empty, containing a rear-facing infant seat, or occupied. Our first experiment used a single, monochrome video camera. The system was automatically trained on a series of test images. Using a principle components (eigenimages) nearest neighbor classifier, it achieved a correct classification rate of 99.5% on a test of 910 images. Our second experiment used a pair of monochrome video cameras to compute stereo disparity (a function of 3D range) instead of intensity images. Using a similar algorithm, the second approach achieved a correct classification rate of 95.1% on a test of 890 images. The stereo technique has the advantage of being less sensitive to illumination, and would likely work best in a real system.

---

For correspondence, contact first author at:  
Microsoft Corporation  
One Microsoft Way  
Redmond, WA 98052  
jkrumm@microsoft.com

This work was performed at Sandia National Laboratories and supported by the U.S. Department of Energy under contract DE-AC04-94AL85000.

### 1. Introduction

The increased safety afforded by automobile airbags (2505 lives saved to date) has produced government regulations and consumer demand that will have half of all vehicles on the road in 2000 equipped with driver- and passenger-side airbags[3]. All cars made after August 31, 1997 must have dual airbags[4].

The increased number of airbags will also magnify their problems: airbags wastefully deploying on empty passenger seats and dangerously deploying on rear-facing infant seats (RFIS). The average cost of replacing an airbag is \$700, which has in part fueled more airbag thefts[4]. A more serious unwanted airbag deployment occurs on RFIS's. Airbags deploy at speeds up to 200 mph. This force is blamed for 12 infant deaths since 1990[3].

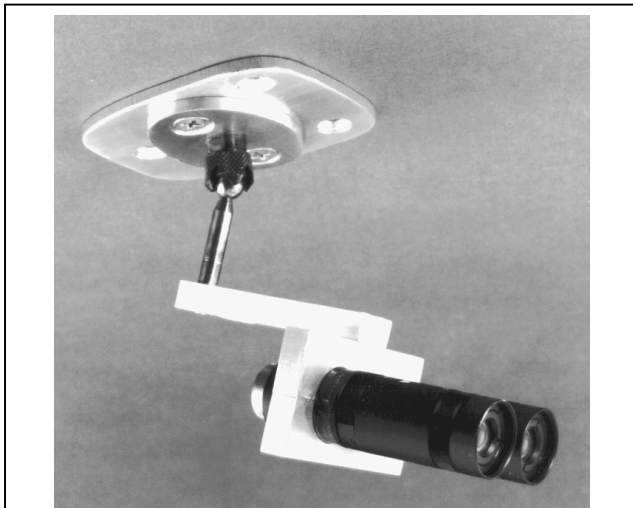
This paper describes a research effort at Sandia National Laboratories to develop a video sensor mounted inside a vehicle to solve these problems. This occupant detection system reliably classifies a vehicle's passenger seat as either empty, occupied by a RFIS, or occupied by a regular person. The camera's view of these three situations is shown in Figure 1. Section 2 describes a system based on principle components of images from a single black-and-white video camera along with conventional nearest-neighbor image classification. In order to make the system less sensitive to illumination and color, we implemented a simple dense stereo algorithm that is described in Section 3. Section 4 describes how these range images can be classified using an algorithm similar to the one used for monocular intensity images. On tests in a real vehicle, the intensity-based algorithm

correctly classified about 99% of test images, while the range-based algorithm correctly classified about 95% of a more challenging test set.

Alternative technologies for occupant sensing[5] include a system from Mercedes Benz that senses the presence of a RFIS with a special resonating device built in. They also use a weight sensor to prevent the airbag from deploying if the passenger seat is empty. Tests have shown, however, that any non-zero reading from a weight sensor in the seat can indicate a wide variety of ambiguous situations, from a bag of groceries to a tightened-down RFIS. Technologies for measuring the presence of an occupant and his/her position include infrared spots, ultrasound, capacitance, and a piezoelectric sheet embedded in the seat to measure pressure distribution. Although the ultimate occupant sensing system will likely use multiple sensors, vision is attractive because it is passive and can provide a multitude of cues for determining how to deploy the airbag, *e.g.* RFIS/empty/occupied, occupant size, and position. This paper shows how we used video to classify the state of the passenger seat.

## 2. Intensity Image Classification

Our first approach to the problem of occupant detection was to use black & white intensity images taken from a single video camera mounted inside a vehicle. We gathered hundreds of images over several days from a stationary vehicle parked outside our laboratory building. Some of the images were used to train our program to explicitly recognize the empty and RFIS classes based on a nearest neighbor algorithm. In order to reduce the amount of computation required, we used eigenvector principle components to compress the image data. This section describes the experimental setup, theory, algorithm, and results of image classification using intensity images.



**Figure 2: Cameras used for monocular and binocular images of vehicles' interiors**

### 2.1. Experimental Setup

We parked our test vehicle near our laboratory building such that it would be shaded for part of the day. We mounted a single video camera near the top of the driver's side "A" pillar using the driver's-side sunvisor mount points for attachment. The camera itself appears in Figure 2 along with a companion camera used for stereo described in Section 3. Typical black & white images from the camera are shown in Figure 1. The images were digitized, stored, and processed on a general-purpose workstation computer inside the laboratory.

Images were taken every five minutes during daylight hours for six days. Three of the days were devoted to images of the empty seat, with the three separate days having the passenger seat adjusted to its most rearward, middle, and most forward positions respectively. The seat was similarly adjusted for the next three days of images of a doll baby in a RFIS. Full days of imaging gave a good variety of illumination as the sun moved overhead on the typically cloudless days in Albuquerque, NM. We also took ten images each of ten adult volunteers as they sat in the passenger seat. In all, we took 638 images of the seat empty, 576 of the RFIS, and 101 images of the seat occupied.

In order to simulate an inexpensive video camera such as might be used in real production, we reduced the resolution of the images by averaging square regions of pixels in the original images into single pixels in their lower resolution counterparts. We varied the amount of resolution reduction for testing.

After reducing the resolution, each image was histogram equalized to help reduce the effects of illumination variations. Histogram equalization was particularly good at recovering acceptable images taken in the darker conditions near sunrise and sunset.

Finally, each image was normalized by dividing each pixel by the square root of the sum of the squares of all its pixels. Mathematically, this means that the sum of the squares of the pixels in each normalized image is one. Practically, this helps factor out overall illumination differences in images that are otherwise similar.

### 2.2. Theory of Image Matching

We classified the test images into three categories: empty, RFIS, or other. We chose not to create an explicit class for occupied seats, since the appearance of an occupied seat is so variable. Any image that was not explicitly classified as either empty or RFIS was considered a case of an occupied seat.

In order to do the classification, we extracted every sixth image from the "empty" and "RFIS" image sets to make a set of prototype images taken 30 minutes apart. Spacing the prototype images evenly over the day helped the system work in spite of changing illumination and shadows. The remaining 5/6 of the images were used as tests, and they were classified by comparing each of them to all the prototype images. If a test image was deemed

similar enough to a prototype image, it was given the same class as the prototype.

To make the image comparison faster, we compressed all the images using principle components computed from the preprocessed (resolution reduction, histogram equalization, normalization) prototype images. For a given set of prototype images (either empty or RFIS), we raster scan each image into a column vector  $\bar{p}_i$ . (In our notation, a bar over a variable indicates a vector.) We form a matrix  $P$  containing all the column vectors  $\bar{p}_i$  side-by-side in no particular order. The sample covariance matrix is

$$P = [\bar{p}_0 - a \quad \bar{p}_1 - a \quad \bar{p}_2 - a \quad \dots \quad \bar{p}_{n-1} - a], \quad (1)$$

where  $n$  is the number of prototype images in the prototype set, and  $a$  is the overall mean of all the elements of all the  $\bar{p}_i$  of the prototype set. For the empty and RFIS classes,  $n$  had the value 106 and 96, respectively. The sample covariance matrix is  $Q = PP^T$ . The eigenvectors of  $Q$  are  $\bar{e}_j$ . Any of the  $\bar{p}_i$  can be reconstructed from the  $\bar{e}_j$  using

$$\bar{p}_i = \sum_{j=0}^{n-1} (c_{ij} \bar{e}_j) + a, \quad (2)$$

where the  $c_{ij}$  are computed from

$$c_{ij} = (\bar{p}_i - a) \cdot \bar{e}_j. \quad (3)$$

These  $c_{ij}$  coefficients serve as a representation of the images. The values of the first two coefficients  $c_{i0}$  and  $c_{i1}$  for the empty and RFIS classes are shown in Figure 3. Each dot in the plots represents one image in the prototype set.

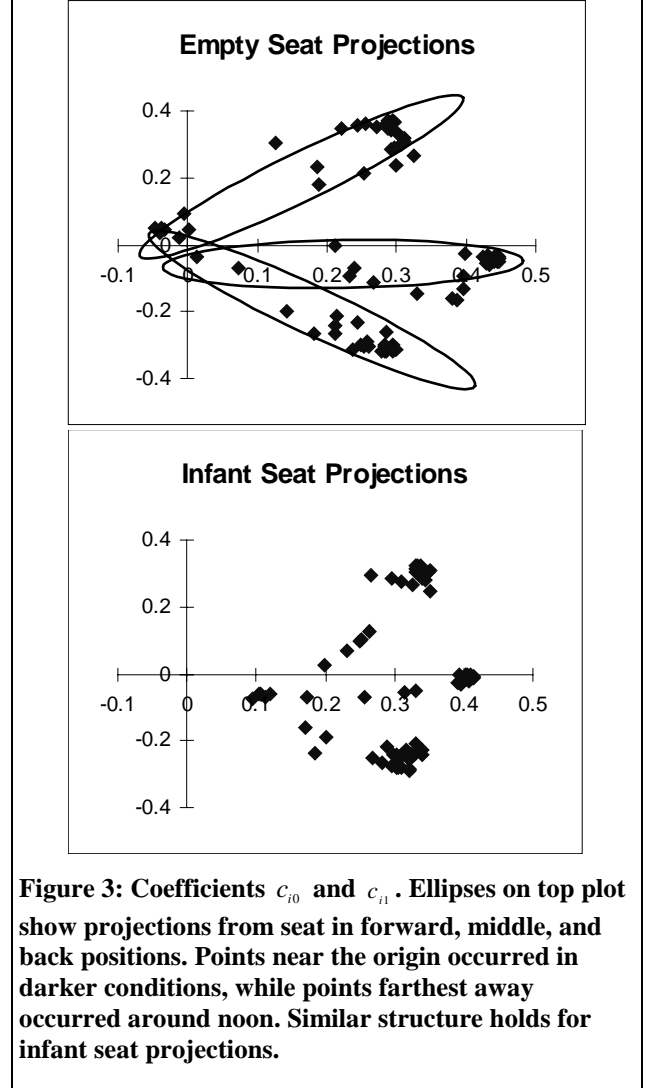
A preprocessed image  $\bar{v}$  with an unknown class can be decomposed with the same eigenvectors into coefficients  $d_j$  using

$$d_j = (\bar{v} - a) \cdot \bar{e}_j. \quad (4)$$

It can be shown that the sum of squared differences (SSD) between a prototype image  $\bar{p}_i$  and the unknown image  $\bar{v}$  is

$$\|\bar{v} - \bar{p}_i\|^2 = \sum_{j=0}^{n-1} (d_j - c_{ij})^2 \quad (5)$$

in terms of the two images' coefficients. To the extent that the images can be approximately reconstructed from the first  $n'$  coefficients (with  $n' \leq n$ ) and corresponding eigenvectors, the SSD can be approximated as



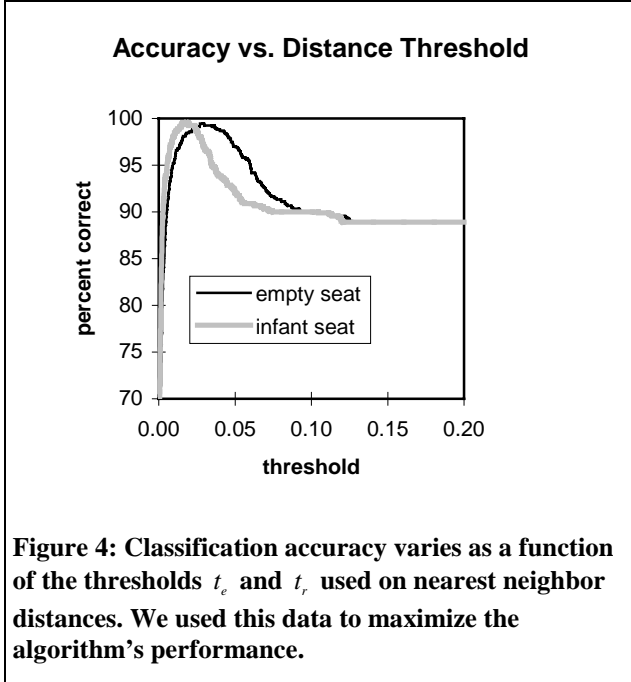
**Figure 3: Coefficients  $c_{i0}$  and  $c_{i1}$ . Ellipses on top plot show projections from seat in forward, middle, and back positions. Points near the origin occurred in darker conditions, while points farthest away occurred around noon. Similar structure holds for infant seat projections.**

$$\|\bar{v} - \bar{p}_i\|^2 \approx \sum_{j=0}^{n'-1} (d_j - c_{ij})^2. \quad (6)$$

Based on our experiments, we used  $n' = 18$ . We justify this choice in Section 2.4.

Once we have the  $c_{ij}$  (which are precomputed) and the  $d_j$  (which are easy to compute from Equation (4)), Equation (6) gives a fast way of approximating the SSD between an unknown image and each of the prototypes.

These ideas of principle components and nearest neighbor classification can be found in standard textbooks such as Fukunaga's [1].



**2.3. Image Matching Algorithm**

We compare each new unknown image to all the prototypes as described above using Equation (6). For a given unknown image,  $\Delta_e$  and  $\Delta_r$  are the SSD's between the image and the nearest neighbor in the empty and RFIS prototype sets, respectively. We classify the image as empty if the image is close enough to an empty prototype, and likewise for RFIS. Specifically, we decide what to do with the airbag according to the following decision table and experimentally determined thresholds  $t_e$  and  $t_r$ :

	$\Delta_e \leq t_e$	$\Delta_e > t_e$
$\Delta_r \leq t_r$	tiebreaker (retain airbag)	RFIS (retain airbag)
$\Delta_r > t_r$	empty (retain airbag)	occupied (deploy airbag)

Unless the seat is explicitly recognized as either empty or RFIS, the airbag is deployed. In the tiebreaker case, the unclassified image looks similar to an image in both prototype classes. By luck of the problem, however, the action of the airbag should be the same in both cases (retain airbag), so it makes no difference which of the two classes the unknown image actually belongs in. We discuss our choice of the thresholds  $t_e$  and  $t_r$  in the next section.

**2.4. Experimental Results**

In assessing the accuracy of our algorithm, we were free to vary several parameters. We adjusted the thresholds  $t_e$  and  $t_r$ , the resolution of the images, and the number of eigenvectors used ( $n'$ ). Our best results were

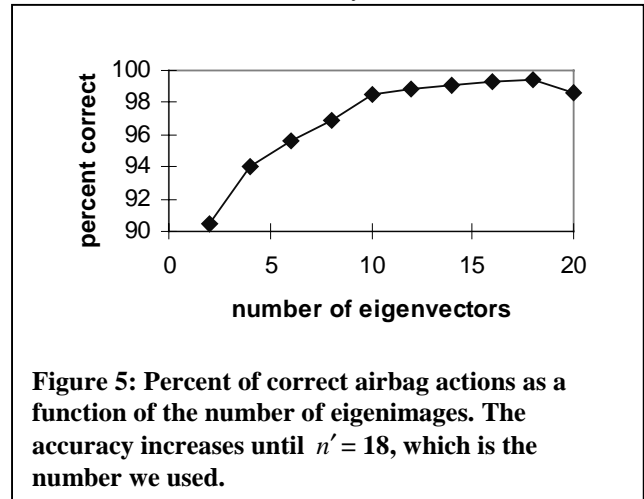
achieved with  $t_e = 0.029$  and  $t_r = 0.019$ , an image resolution of  $96 \times 102$ , and  $n' = 18$  eigenvectors.

The classifier was actually tested as two separate classifiers, one for empty seats and one for RFIS's. The empty seat classifier failed to recognize three of 413 empty seat images as empty seats, and it misclassified one of 101 occupied seats as empty seats. The RFIS classifier failed to recognize one of 396 RFIS images as a RFIS, and it misclassified none of the 101 occupied seats as a RFIS. From the point of view of airbag actions, the results are:

Airbag action	Computed action/ correct action	Percent
Correct overall	905/910	99.5%
Fatal retention (on occupied seat)	1/101	1.0%
Fatal deployment (on RFIS)	1/396	0.3%
Unneeded deployment (on empty seat)	3/413	0.7%

We refer to the 99.5% figure as the "accuracy" of the system. This is the percentage of images on which the system directs the airbag to take the correct action. We note that there were no cases where the system achieved extra accuracy by merely confusing an empty seat with a RFIS or vice versa.

We chose the thresholds by computing the percentage of correct airbag actions as a function of the thresholds. This data is plotted in Figure 4. By picking the thresholds to correspond to the maxima of these plots, we maximized the accuracy. Ideally, the plots would show broad peaks near their respective maxima, which would indicate relative insensitivity to the value of the thresholds. As the thresholds increase, the accuracy reaches a constant value.



At this point, the net action of the system is to retain the airbag in every case, and the accuracy percentage simply reflects the relative number of the three classes of images in the test set.

We found that accuracy was not a strong function of resolution in the range of resolutions that we tested. From a minimum image size of 68x73 to a maximum size of 120x128, the accuracy varied by only 0.2%.

The final adjustable parameter was the number of eigenvectors used,  $n'$ . We optimized performance by computing the accuracy as a function of  $n'$ . The results of this computation are shown in Figure 5, and the best value was  $n' = 18$ .

### 3. Stereo Vision

One potential problem with classifying intensity images, as described in the previous section, is that a class with large intensity variations may be difficult to characterize with a limited number of prototype images. For instance, we would not expect our classifier to work well if the seats of the test vehicle were temporarily covered with a towel. Even more important, this limitation prevented us from establishing a separate “occupied” class, because the appearance of an occupied seat varies significantly with what clothes the occupant is wearing. We could not hope to capture enough variations in the appearance of an occupied seat with a reasonable number of prototype images.

This problem prompted us to consider using images whose pixels represent range (distance to surface) rather than intensity. Our justification is that range images are ideally insensitive to the lightness of objects in the scene, and that prototype range images of a given class will be more similar to each other than prototype intensity images. This is especially true for occupied seats, where the range image is ideally independent of the color of the occupant’s clothes.

Our technique for getting range images is binocular stereo, which we describe in the next subsection. We used essentially the same techniques for classifying range

images as we did for intensity images. Classification of the range images is described in Section 4.

#### 3.1. Experimental Setup

We used two cameras, mounted side-by-side, as shown in Figure 2. The camera mount held the cameras nearly parallel. We aligned the rows of the two cameras by pointing the cameras at a horizontal edge. We rotated them each around their respective roll axes until the edge was as close as possible to being horizontal near the center row of both images. This made the epipolar lines correspond approximately to rows in the image, meaning that a match for a point in the left image would fall on a known row in the right image. Given that we reduced the image resolution by four times before stereo matching (480x512 down to 120x128), approximate alignment was sufficient. A typical stereo pair from inside the vehicle is shown in Figure 6.

#### 3.2. Binocular Stereo

Measuring range from a stereo pair such as ours reduces to measuring the disparity (shift) between corresponding points in the left and right images. The range is inversely proportional to disparity. In fact, we did not compute range at all, using just the raw disparities for classification.

For each point in the left image, we find a match in the right image using a stereo correlation method described by Matthies in [2]. This method extracts a small window around each point in the left image and finds the best match in the right image using correlation (SSD) search along the epipolar line. For our reduced resolution stereo images of size 120x128, we used windows of size 5x7. Based on the geometry of the cameras and scene and the resolution of the images, we limited the disparities to the range [2,5] pixels. Following Matthies’ algorithm, we computed subpixel disparities by fitting a parabola to the SSD values at the minimum SSD and its neighbors on either side. The subpixel disparity was taken as the location of the minimum of this parabola. A typical disparity image is shown in Figure 6. Note that we have masked out the pixels on the window of the vehicle, as



**Figure 6: Stereo (left/right) images taken inside test vehicle. Rightmost image shows computed disparity, with lighter points having more disparity.**

they give no indication of the state of the passenger seat.

#### 4. Disparity Image Classification

Our procedure for classifying disparity images is nearly the same as that for classifying intensity images, as described in Section 2. Besides the obvious difference of using disparity instead of intensity, the other difference was that we classified into three classes (empty, RFIS, occupied) rather than just two (empty, RFIS) as we did with intensity images. We felt that the intensity invariance of the disparity images justified a separate class for occupied seats that would be compact enough to give accurate classifications.

We collected stereo images over a period of seven days. For the empty and RFIS cases, we collected data in the same way as for the first experiment using monocular images (every five minutes, set in rearward, middle, and forward positions for one day each for both empty and RFIS). We also took ten images of a different RFIS ("minority RFIS) for testing. For the occupied class, we took 439 stereo pairs of 22 regular occupants (20 images of each person, with one bad image thrown out). The occupants were asked to change positions during the imaging. Of all the images, we used 76 empty seat pairs and 68 RFIS pairs for training, taken at 30-minute intervals. None of the 10 "minority" RFIS were used for training. We used 220 of the occupied seat pairs for training. All the stereo pairs that were not used for training were used for testing. For the occupied seat, 11 pictured individuals were used for training, and 11 others were used for testing. All the stereo pairs were subjected to our stereo algorithm, and all subsequent processing was done on the real-valued disparity images.

We processed the disparity images in the same way as the intensity images, eliminating histogram equalization and normalization. We approximated the SSD using the top 20 eigenvectors and classified an unknown disparity image with its nearest neighbor out of all the images in the prototype sets. This method achieved a classification accuracy of 93%.

We modified the classification program to give weights to the three SSD's for each of the three classes. After computing the SSD between the projection of an unclassified image and the projection of a training image, it was scaled by a weighting factor for that prototype's class. The optimal weights for the empty, RFIS, and occupied classes were 1.000, 1.016, and 0.792, respectively. Using these weights brought the classification accuracy up to 95%.

The following table shows the specific types of classification errors. We note that two of the fatal deployments (on RFIS) were due to misclassifying the minority RFIS as an occupied seat. The minority RFIS was correctly classified in the remaining eight images. The weights could be adjusted to decrease the number of

fatal errors (fatal retention and fatal deployment) at the expense of unneeded deployments.

Airbag action	Computed action/ correct action	Percent
Correct overall	846/890	95.1%
Fatal retention (on occupied seat)	1/219	0.5%
Fatal deployment (on RFIS)	16/318	5.0%
Unneeded deployment (on empty seat)	27/353	7.6%

#### 5. Conclusions

We have shown that video images can be successfully used to determine whether or not to deploy the passenger-side airbag. Images of the passenger seat taken from a video camera mounted inside the vehicle can be used to classify the seat as either empty, containing a RFIS, or occupied. Our first experiment used a single video camera. The system was automatically trained on a series of test images. Using a principle components (eigenimages) nearest neighbor classifier, it achieved a correct classification rate of 99.5% on a test of 910 images. Our second experiment used a pair of video cameras to compute stereo disparity (a function of 3D range) instead of intensity images. Using a similar algorithm, the second approach achieved a correct classification rate of 95.1% on a test of 890 images. The stereo technique has the advantage of being insensitive to illumination, and would likely work best in a real system. In addition, range data from stereo images could be used to estimate the position of the occupant, giving important information on how to deploy airbags in an advanced system with multiple airbags and variable inflation rates.

#### References

1. Fukunaga, Keinosuke. Introduction to Statistical Pattern Recognition, Second Edition. Academic Press, 1990.
2. Matthies, Larry. *Dynamic Stereo Vision*, (Ph.D. Thesis), Carnegie Mellon University School of Computer Science, Technical Report CMU-CS-89-195, October 1989.
3. McGinn, Daniel and Daniel Pedersen. "A Life-or-Death Choice?" *Newsweek*, October 20, 1997, 41-45.
4. O'Donnell, Jayne. "Insurers find some accident costs actually increase," *USA Today*, September 29, 1996, Sec. B, p. 4.
5. Paula, Greg. "Sensors Help Make Air Bags Safer," *Mechanical Engineering Magazine*, 119(8), August 1997.