Abstract - In many real-time object recognition applications, the system experiences conditions where the classification results are not reliable due to a variety of environmental or object poses where correct classification is difficult or even impossible. We propose that by tracking the motion and orientation of the object of interest, and fusing this information with the classification results, we can greatly improve the classifier performance. This can be achieved firstly by estimating the reliability of the classification, and secondly by using track state estimates to derive additional classification cues. We develop a framework based on Interacting Multiple Model (IMM) Kalman Filtering, Dempster-Shafer evidential reasoning and fuzzy set memberships, for integrating the track and classification information from an incoming video image stream. We demonstrate the performance of our proposed framework in the application of a real-time vision system for smart automotive airbags. We show that our fusion approach improves the final performance to 100% correct classification, the level required for a robust safety system.

Keywords: Image classification, Kalman filtering, Dempster-Shafer, fuzzy set membership.

1 Introduction

Image classification methods generally operate on a single image. Many real-time applications, however, collect and classify a sequence of images over time [2][3]. During operation, the system has the opportunity to collect many images in conditions that provide reliable classification results. Likewise, during these same periods of operation, there are times when the system may experience a variety of situations that may make the correct classification difficult or even impossible. These situations may be due to external environmental conditions (e.g. camera saturation due to bright sunlight) or they may be due to target motion (e.g. a person pulling a sweater over their head). In these difficult to classify cases, where there may be no reliable information from the classification sensor, the system should be able to declare ‘ignorance’ regarding the object classification. Additionally, while the orientation of the object of interest may not be favorable for correct classification, the track information regarding the motion and orientation of the object of interest provides an additional source of information.

The aim of this paper is to present a robust processing framework for fusing a temporal stream of classification results with additional class-related information derived from a tracker based on three technologies: (i) Interacting Multiple Model (IMM) Kalman Filtering, (ii) fuzzy set membership, and (iii) Dempster-Shafer evidential reasoning methods. The datasets we used for our experiments are from a real-world proto-type vision-based airbag suppression system.

2 Automotive Smart Airbag Application

The integration of airbags into passenger vehicles during the 1980’s and 1990’s has been particularly effective in reducing the number of highway fatalities in the United States. Unfortunately, airbags were designed to protect the worst-case scenarios, namely a 95th percentile adult male during a 30 mph crash, which makes them potentially dangerous for smaller occupants [6]. Consequently, between 1986 and 2001, 19 infants and 85 children were killed, where it would have been safer if the airbag had been disabled [6].

Consequently, considerable attention has recently been paid to developing “smart” airbags that can determine if they should be deployed in a crash event, depending on the type of occupant, particularly vision systems [12][13][14]. In our system the occupant is monitored using a real-time monocular computer vision system which is mounted in the roof-liner in the location identified in Figure 1. We chose a monocular vision approach over a stereo vision or multi-sensor approach to develop the lowest possible cost system [11][12][13][14].

There are five operating modes of the system: (i) disable the airbag for infants, (ii) enable the airbag with a low power deployment for properly seated children, (iii) disable the airbag for adults or children that are too close to the airbag, (iv) enable the airbag with full power for
properly seated adults, and (v) disable the airbag for an empty seat. These four classes of occupants yield our application environment, \( \Theta = \{ \text{infant}, \text{child}, \text{adult}, \text{empty} \} \). Images collected with the system for these classes are shown in Figure 2.

![Figure 1. Installation of the camera system within the vehicle showing plan, profile, and forward viewing angles.](image)

The image processing consists of two parallel paths, one for classification processing and one for track processing as shown in Figure 3. The classification processing determines the presence of an empty seat, infant seat or a child, while the track processing determines if the occupant is too close to the airbag for safe deployment.

![Figure 2. Examples of each of the four classes of occupants: (a) infant, (b) child, (c) adult, and (d) empty seat.](image)

### 2.1 Classification Subsystem

We utilize a feature-based image classification approach [7]. The incoming images are initially segmented to remove the occupant and the seat from the background. Examples of segmented images of an adult and an infant are in Figure 4 a) and b) [8]. To reduce the effects of illumination, we then perform edge detection on these segmented images. Examples of these edge detected images for an infant and an adult are likewise provided in Figure 4 c) and d).

![Figure 4. Edge maps of infant and adult segmented images: (a) segmented RFIS, (b) segmented adult, (c) edge image of RFIS, and (d) edge image of adult.](image)

We then compute the Legendre moments up to 45\(^{th}\) order of these edge images and generate a 1081-element feature vector [7]. Off-line feature selection defined the best 160 of these features to use for classification. The incoming reduced feature vectors are then classified using a \( k \)-NN classifier with \( k=5 \). To make the system more robust to classification errors, we integrate the sequence of classifier outputs using evidential reasoning, as defined in [10].

![Figure 3. Processing for airbag suppression.](image)

While past results have shown the system performs well for normally seated adults and infants [8], we also found that specific motions and positions of the occupant regularly confused the classifier. In these instances, the movements of the occupants made their shape resemble an infant seat. Examples of these positions and motions are provided in Figure 5 [8]. We cannot solve these classification errors solely through shape classification and historical evidential reasoning.

### 2.2 Track Subsystem

The tracker subsystem is a parallel processing path to the classification processing, as shown in Figure 3. The track subsystem relies on optical flow motion segmentation to extract the occupant from the background [7]. Optical flow was chosen over template-based methods for the following reasons: (i) it is able to estimate the motion of regions of the image quite accurately, (ii) it does not require any initialization, (iii) it is not sensitive to background clutter or object texture, and (iv) it has no difficulty with large amounts of motion between frames [15]. The trade-off is that optical flow methods do not provide as accurate boundary segmentations as template matching, which we will show does not pose a significant problem for our application.
Figure 5. Example images where the classification subsystem confused an adult with an infant or child seat image: (a) occupant leaning forward, and (b) occupant with arm motions.

In Figure 6, we represent the occupant via the bounding ellipse about the occupant’s head and torso [7]. We have previously shown this bounding ellipse representation to be extremely effective in tracking the position and orientation of the occupant for determining if they are too close to the airbag for a safe deployment [7]. Additionally, this is a very simple representation that facilitates real-time processing [7]. Figure 6 defines these ellipse parameters for the occupant and demonstrates its appearance on an incoming image. Figure 7 provides examples of ellipses fit over the range of occupants.

Figure 6. Human geometry representation for tracking: (a) definition of ellipse parameters, and (b) ellipse fit to the human occupant.

Figure 7. Example ellipses generated from motion segmentation and ellipse fitting for (a) 3 year old, (b) 6 year old, (c) 5th percentile female adult, and (d) 50th percentile male.

The tracker subsystem utilizes two subordinate tracking filters to track subsets of the ellipse parameters, one for shape and one for motion. The motion tracker determines if the occupant’s position relative to the airbag, and hence tracks the x-axis centroid and the lean angle $\theta$. The motion state vector consists of the position, velocity, and acceleration for both the $x_{\text{centroid}}$ and $\theta$. Since the occupant may move either through their own motion or through pre-crash braking, the tracker must handle a broad range of occupant dynamics. Consequently, the motion tracker is an Interacting Multiple Model (IMM) Kalman filter, where the three types of occupant motion are (i) stationary, (ii) normal human motion, and (iii) pre-crash braking motion [7].

The shape tracker determines the size of the ellipse, and tracks the major and minor axes and the y-axis centroid [7]. It uses a traditional Kalman filter, where the state vector consists of the position, velocity, and acceleration for each of these three ellipse parameters.

3 Approaches for Fusing Classification and Track Processing Results

Since the classification subsystem alone cannot meet the government mandated 100% correct classification, we must define a source of additional classification information that can be fused with this image classification information. The use of additional sensors, such as seat weight sensors, [12][13] is not a feasible source of information for two reasons: (i) the added cost, and (ii) the vehicle manufacturers would prefer to not have any sensing hardware in the seats due to complexity of integration, and the fact that the seats are the last element of the vehicle to be designed. We will demonstrate that the tracker subsystem, which was originally designed only for out-of-position detection, provides a rich source of occupant information that can be fused with the classifier information.

There are three data sources available from the track subsystem to fuse with the classification information: (i) motion segmentation cues, (ii) motion-based features to use for classification, and (iii) classifier reliability estimation based on occupant track characteristics. In the first data source, the motion segmentation algorithm and the track state estimates can provide information regarding which regions of the image are experiencing potentially distracting motions. For example, as shown previously in Figure 5, the occupant’s hand motions often prevented correct classification. Therefore, we may mask regions in the image with motion to improve the classification performance.

The second data source, namely track-derived features, augments the features derived through the classification with additional features derived from the motion of the occupant. Here specific track characteristics generate additional features, which may be suitable for occupant classification. The third data source, namely classification reliability estimation, uses characteristics of the occupant track to estimate whether a robust classification is even possible in the current image due to the pose of the occupant. For this paper, we concentrate on methods (ii) and (iii), since they provide direct inputs to the classifier, rather than indirect inputs through the segmentation process.
There are numerous possible data abstraction levels for fusing the track and classifier outputs, and three candidates are shown in Figure 8. In the first option in Figure 8 (a), the outputs from the tracker and the classifier are treated as a single output stream and fused as they are produced. In this approach, however, the tracker tends to dominate since the output rate of the tracker is considerably higher than the classifier (video rate of 40 Hz versus 0.2 Hz for the classifier). Thus, it is justified to integrate the tracker information between classifier updates, as is shown in Figure 8 (b). In approach (b), the system initially integrates the tracker data stream, and then fuses these integrated track outputs with the incoming classifier results. The third method, shown in Figure 8 (c), independently integrates the temporal streams of the tracker and the classifier, and then fuses their results together. The benefit of approach (c) is that it treats each data source symmetrically. It has the added benefit of reducing the random errors present in each channel of information at their original level of abstraction prior to fusion, which allows the fusion process to focus predominantly on managing the more difficult assignable errors.

![Figure 8. Possible approaches to fusing data: (a) directly fuse track outputs with classifier outputs, (b) first integrate track outputs and then fuse with classifier, and (c) separately integrate track outputs and classifier outputs, and then fuse their integrated outputs together.](image)

### 4 Integration of Classification Sequences

Integration of classification sequences requires algorithms that can ‘deal with information that is to be uncertain, imprecise, and occasionally inaccurate’, which is the objective of evidential reasoning [1][3]. There are two distinct mechanisms for managing changes in belief, (i) belief revision and (ii) belief updating, where belief revision integrates information in a static situation, while belief updating integrates information about the world when its state is changing [2]. In the airbag suppression application, we are observing the same occupant while integrating real-time sequences of classification results; hence, we will apply belief revision.

There are many methods for belief revision, such as basic Bayes inference, Transferable Belief Model (TBM), Dempster-Shafer, etc. [1], and both our previous research in [10] and results on classifier combination in [5] have had promising results using Dempster-Shafer. Additionally, our ultimate goal is to integrate the track information with the classifier information, but each of these sources of information are provided at a different level of abstraction. Dempster-Shafer provides a natural mechanism for integrating such data sources through its set-theoretic formulation [1][4].

The basic element of evidence in Dempster-Shafer is the mass or basic probability assignment, defined by \( m \). It has the following two properties [1]:

\[
m(\phi) = 0, \quad \text{and} \quad \sum_{A \in \mathcal{P}(\Theta)} m(A) = 1, \tag{1}
\]

where \( \phi \) is the null set and \( A \) is an element of the power set \( \mathcal{P}(\Theta) \), where \( \Theta = \{ \theta_1, \theta_2, \ldots, \theta_N \} \), a set of mutually exclusive objects.

These probability masses are then used to calculate the belief in any given proposition, and the plausibility of that proposition [1][3]. Mathematically, they are defined to be [1]:

\[
Bel(X) = \sum_{A \subseteq X} m(A) \tag{2}
\]

and

\[
Pls(X) = 1 - Bel(\neg X) = 1 - \sum_{A \subseteq X} m(A), \tag{3}
\]

where \( X \) is an element of the power set \( \mathcal{P}(\Theta) \). The two probability masses \( m_1 \) and \( m_2 \) are combined as follows [1]:

\[
m_1 \odot m_2 (Z) = \frac{\sum_{X \cap Y = Z} m_1(X) m_2(Y)}{1 - \sum_{X \cap \phi = \phi} m_1(X) m_2(Y)}, \tag{4}
\]

where \( X, Y \) and \( Z \) are the elements of the power set \( \mathcal{P}(\Theta) \). The probability masses are derived directly from the classifier outputs.

The nature of the errors that we are trying to address, where there is a sudden change in the environment or the pose of the object being analyzed, results in assignable errors which manifest themselves as sudden changes in the beliefs of the current system. We previously showed that these assignable errors can be detected by monitoring the impact of new information on established beliefs by computing [10]:

\[
|\Delta Bel| = \sum_{\mathcal{P}(\Theta)} |Bel_{\text{temp}} - Bel_{\text{last}}|, \tag{5}
\]

where \( Bel_{\text{temp}} \) is the Beliefs assuming the new information has been integrated, and \( Bel_{\text{last}} \) is the Beliefs up to the inclusion of the new information. In cases where this value becomes large, we must discount the
incoming probability masses [10]. The method we propose for discounting the incoming evidence in the event of conflict parallels Jeffrey’s rule for Bayesian inference and is defined by [10]:

\[
    m(A) = \begin{cases} 
    p \cdot m(A) \lor \forall A \in P(\Theta), \text{where } A \neq \Theta, \\
    p \cdot m(A) + (1-p) \cdot 1, \text{for } A = \Theta 
    \end{cases},
\]

(6)

where \( \Theta = \{\theta_1, \theta_2, ..., \theta_N\} \) is the subset of all atomic elements, which represents ignorance, and \( p \) is the probability of the evidence being accepted. In other words, we rescale the masses of all of the power set elements, except for the last element, which contains all the atomic set members. For this element we scale the existing mass by \( p \) and add \((1-p)\) to this value, which corresponds to adding additional ignorance, based on the consistency of the evidence. The probability \( p \) of the evidence being valid is determined by:

\[
    p = 1 - \frac{|A_{Bel}|}{\sum_{p(\theta)} B_{last}},
\]

(7)

which is a uniformly distributed measure from the relative change in beliefs if the current evidence was used.

Once the incoming classifications probability masses are tested via Equation (5), and discounted as needed through Equation (6), they are integrated with the classification history using Dempster’s rule of combination in Equation (4).

5 Derivation of Occupant Motion Information

As shown in Section 3, we use the tracker subsystem to provide two additional pieces of information: (i) classifier reliability information, and (ii) additional classification features. The motion tracker provides the classifier reliability, and the shape tracker provides additional classification features, as is shown in Figure 9.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure9.png}
\caption{Assignment of ellipse information to the motion and shape tracker components, and outputs used for motion-classification fusion.}
\end{figure}

The motion tracker estimates the classification reliability via two mechanisms: (i) the lean angle of the occupant towards the instrument panel, and (ii) the presence of motion outside of the occupant’s head/torso region. These two mechanisms address the two conditions of the occupant that were demonstrated in Figure 5 (a) and (b) respectively.

The lean angle is provided directly from the motion tracker state vector. When the occupant lean angle is below 70 degrees we find the classifier consistently mis-

classifies the adult as an infant (\( \theta \) is 90 degrees when the occupant is sitting perfectly upright and 0 when the occupant is leaning completely forward). Whenever the lean angle is below this threshold we set the incoming classification probability masses to complete ignorance.

The second mechanism for estimating classifier reliability is the analysis of the motion in the image. We use the binary motion field image, generated by thresholding the optical-flow motion image. We process this image by masking out the region of the occupant’s ellipse, the window region, and the instrument panel and roof-liner regions, as is shown in Figure 10. The remaining image region is then integrated via a moving window of the last \( N \) images analyzed for motion. If the number of pixels with motion in this region exceeds 20% of the occupant’s ellipse area, then we replace the incoming occupant classification with complete ignorance. This threshold is based roughly on the relative size of the arms to the torso of a seated occupant when viewed by our camera system.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure10.png}
\caption{Motion-based classifier reliability estimation, (a) Incoming image, (b) optical flow image, (c) mask defined by interior, window region, and ellipse region(black shows region to consider for motion), and (d) final integrated non-torso motion.}
\end{figure}

The tracker subsystem also provides classification information via the shape tracker. The candidate parameters for this additional classification feature are: (i) the major axis, (ii) the top-most point of the ellipse, (iii) the y-axis centroid (vertical position) of the ellipse, and (iv) the area of the ellipse. We derive the top-most point of the ellipse, \( y_{top} \), via:

\[
    y_{top} = a_{major} \cdot \sin(\theta) - y_{centroid},
\]

(8)

where \( a_{major} \) is the major axis of the ellipse, \( y_{centroid} \) is the y-axis centroid, and \( \theta \) is the lean angle. Of these four candidate measures, we found the best to be the topmost point of the ellipse. It was the least sensitive to arm motions and other occupant motions. Figure 11 provides a histogram showing the frequency of occurrence for the top of the ellipse (measured from the top of the image) for our various occupant types collected from a series of drive tests. We must now map this data into measures of the relative membership into each of our classes. This can be accomplished using a traditional classifier, such as a quadratic Bayes classifier, however, our range of
occupant types maps well to the linguistic notions of the relative sizes of the occupants (small for child, and large for adult, etc.) which is very naturally handled by fuzzy logic set memberships [9].

To use the top of ellipse measure for providing additional classification information, we model the value using the fuzzy membership shown in Figure 12, where we model the occupants as falling into one of three possible sets, \{child\}, \{adult\}, or \{child, adult\}, where we have employed the set-theoretic approach of Dempster-Shafer for membership assignment.

Figure 11. Histogram of top of ellipse for 3 year old (black), 6 year old (cyan), 5th percentile female adult (red) and 50th percentile male adult (blue). Note a larger value means the top of the ellipse is further down from the top of the image.

As with the classification results, we must also compute an estimate of the reliability of the tracker classification features. Note that the tracker shape information can only be considered valid when the occupant is moving. When the occupant is stationary we cannot infer any information since a non-moving adult or an empty seat would both return no motion.

Figure 12. Fuzzy memberships for top of ellipse for adult set (blue) and child set (red).

We estimate the likelihood of human motion by analyzing the probabilities of the IMM tracker being in three motion models: (i) stationary, (ii) human motion, and (iii) pre-crash braking motion. Here each of these models has increasing amount of dynamics. At time \( k \), the probability of occupant track being in each of these motion models is computed by [7]:

\[
\mu_m(k) = \frac{1}{\kappa} \cdot L_m(k) \sum_{s=1}^{N} p(s \mid t) \cdot \mu_s(k - 1)
\]

where

\[
\kappa = \sum_{s=1}^{N} L_s(k) \sum_{t=1}^{N} p(t \mid s) \cdot \mu_s(k - 1),
\]

and

\[
L_m(k) = N[\text{residue}_m(k), 0, \text{cov}_m(k)].
\]

\( L_m(k) \) is the likelihood of the model \( m \) at time \( k \) based on the residue from the incoming measurement, \( \mu_m(k - 1), \mu_m(k) \) are the model probabilities for times \( k - 1 \) and \( k \), respectively, and \( N[.\] is the normal distribution. The term \( \text{residue}_m(k) \) is the difference between the predicted measurements and the actual measurements at time \( k \) and \( \text{cov}_m(k) \) is the covariance matrix that models the process noise for model \( m \) at time \( k \). The likelihood \( L_m(k) \) is the probability of the residue at time \( k \), given the covariance for each of the \( m \) models.

6 Fusion of Classification and Track Information

Based upon the third approach to track-classifier fusion defined in Section 3, the processing flow for applying fusion to airbag suppression is provided in Figure 13. The first step in processing the track information is to determine the reliability of the track information. When there is no motion, the tracker returns ignorance as its classification, despite the fact that the tracker is dead-reckoning the ellipse parameters. If motion is detected, the fuzzy membership function from Figure 12 is then used to determine the probability mass assignment of the occupant into the \{child\}, \{adult\}, or \{child, adult\} subsets.

The integration of the track-based class information is performed using Dempster’s rule of combination. For each incoming track result, we first add ignorance to the mass assignment, as defined in Section 4, and then integrate it with the stream of track inputs. This approach was previously shown to be effective in [10].

The integration of the incoming classification results is performed whenever a classification result is presented to the system. We first test the reliability of the incoming information via the lean angle and occupant hand motion, as defined in Section 5. If the incoming classification result is deemed unreliable it is replaced with ignorance, otherwise, the classification results are used. Again, for each classifier input we add probability mass to the subset
representing ignorance and renormalize the masses. For the classifier we have empirically found adding 0.1 to the ignorance subset, while for the tracker adding 0.3 to account for the fact that in general the classifier is a more powerful estimator of the object classification. We have also found these values are not particularly sensitive, however, maintaining a reasonable ratio between the values is important to ensure the tracker inputs do not dominate the decisions. The integration of the classifier information is performed using another instance of Dempster’s rule of combination.

In the next fusion processing step we integrate the track and classification information streams, as was shown in Figure 8. The accumulated evidence for the tracker is combined with the accumulated evidence for the classifier, using the rule of combination. This fused result of the tracker and the classifier is then integrated into the historical mass assignments by one final application of the Dempster’s rule of combination of the current fused track-classifier data with the data generated up to the previous classification result. After the generation of the probability masses for the power set, we also compute the belief and plausibility for each element of the power set using Equations (2) and (3).

We then load the fused stream of masses, beliefs, and plausibilities into a history cache as one final mechanism to reduce further the effects of longer-duration assignable errors. The cache is a rolling first-in-first-out buffer of the last \(N\) classifications, where the oldest cache element is replaced with the incoming fused classification result. This cache holds the fused probability mass assignments, beliefs, and plausibilities.

These cache elements for the belief and plausibility are averaged at each system output and the final confidence in each subset is computed by [10]:

\[
P(A) = wt_{Bel} \times Bel(A) + wt_{Pls} \times Pls(A),
\]

where \(wt_{Bel}\) and \(wt_{Pls}\) are the weights for combining the beliefs and plausibilities for the subset \(A\), and \(wt_{Bel} + wt_{Pls} = 1\). For our safety application where we need to rely more on evidence that supports the beliefs, rather than simply does not refute the beliefs, we found values of \(wt_{Bel} \geq 0.75\) provided best results. We compute \(P(A)\) for each of the atomic sets of the airbag application, and the atomic subset with the highest value is then the output classification that decides the state of the airbag (disable, low-power enable, fully-enable).

### 7 Experimental Results

We tested the fusion of track information with classification information to improve the final classification results on two very interesting real-time outdoor driving sequence, where there were various effects influencing the classification, including: (i) variable background, (ii) changing illumination, and (iii) dramatically changing occupant pose and orientation. Figure 14 shows examples from the first sequence of a 50\(^{th}\) percentile adult male, which consisted of over 4500 images.

Likewise, Figure 15 shows samples from a sequence of over 2200 images taken in similar driving conditions. Assignable errors are clearly visible in these example image sequences, where the adult image has the shape of an infant seat in certain occupant poses.

We compare the performance of the proposed fusion system on these two sequences, with the original classifier-only performance, and the history processed classification processing defined in [10]. Table 1 and Table 2 provide the performance of the track-classification fusion, measured as correct occupant classification.

The results in Table 1 shows the relative improvements as we add the additional object classification information derived from the tracker, as well as the additional information regarding the reliability of the classification information. Note that the performance of the system is better than for the classification-only data, even when a longer duration data cache was used.

In Table 2, the occupant is the 5\(^{th}\) percentile adult female subject. In her scenario, the leaning motions were typically of shorter duration than for the adult male subject, which allowed the classification-only processing to correct the errors with a cache depth of 15. Importantly, the information from the track subsystem did not degrade the classification information despite the fact that due to her smaller stature, the evidence derived from the top of the ellipse data was not conclusively adult, but often lead to the classification of \{child, adult\} subset.
Table 1. Summary of classification results for image sequence 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input classification sequence</td>
<td>63.8%</td>
</tr>
<tr>
<td>Dempster-Shafer on Classifier only (cache depth = 1)</td>
<td>76.7%</td>
</tr>
<tr>
<td>Dempster-Shafer on Classifier only (cache depth = 15)</td>
<td>87.5%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 1)</td>
<td>88.5%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 5)</td>
<td>91.7%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 15)</td>
<td>91.9%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation + hand-motion tests) (cache depth = 1)</td>
<td>95.6%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation + hand-motion tests) (cache depth = 5)</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2. Summary of classification results for image sequence 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input classification sequence</td>
<td>73.2%</td>
</tr>
<tr>
<td>Dempster-Shafer on classifier only (cache depth = 1)</td>
<td>96.0%</td>
</tr>
<tr>
<td>Dempster-Shafer on classifier only (cache depth = 15)</td>
<td>100%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 1)</td>
<td>100%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 5)</td>
<td>100%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation test only) (cache depth = 15)</td>
<td>100%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation + hand-motion tests) (cache depth = 1)</td>
<td>100%</td>
</tr>
<tr>
<td>Fusion with Tracker (orientation + hand-motion tests) (cache depth = 5)</td>
<td>100%</td>
</tr>
</tbody>
</table>

8 Conclusions

We have demonstrated a processing framework suitable for the integration of temporal stream of image classifications with tracker-derived information. We framed the problem as a belief revision problem, and proposed using evidential reasoning as the general framework. We showed that information from the tracker can be used for two purposes: (i) estimate the reliability of the classification information, and (ii) provide additional information regarding the occupant class. We used a fuzzy-logic membership function to derive classification information from the tracker state estimates. Likewise we derived the classification reliability from a combination of state estimates, and model probabilities from the Interacting Multiple Model (IMM) Kalman filter.

We demonstrated that we were able to correct the assignable errors, resulting from changes in the occupant pose and motions. For two real-world drive sequences of over 6000 combined images, the classification accuracy of the system improved from 63.8% to 100% correct. This approach of successfully fusing track and classification information is a critical step towards a production-ready smart airbag system.

References