NAECON08 Grand Challenge Entry Using the Belief Filter in Audio-Video Track and ID Fusion

Erik Blasch
Wright State University
Dayton, OH 45324

Abstract – *The IEEE NAECON2008 Challenge Problem competition* was to observe, monitor, and determine the cause-effect relations in an audio-video data set. Acoustic-to-visual fusion establishes the relationships of events much as the human brain associates spatial-temporal audio and video feature data. A machine, like a human, can infer from the audio (hear) and the video (see) the cognitive processing relationships between perception and judgment using standard figures of merit (FOM). After downloading the data sets, a series of signal processing, image fusion, and tracking methods were applied to extract salient features for a belief filter (BF) to determine the cause-effect event of ball movement on a suspended object.

Keywords: *Grand Challenge, Audio, Video, Information Fusion, Convolution, Belief Filter*

1 Introduction

The IEEE NAECON Conference resumed with zeal in 2008 with a host of conference activities, including a Grand Challenge competition.[1] The competition supplied audio and video data from disparate views and asked the team entries to assess the data for interesting information. The information contained in the audio tracks and video files (movies) was a toy that had a ball moving with gravity through a series of obstacles as shown in **Figure 1**.

Figure 1. Scene of Challenge Problem Data.

Specifically, the data set was to be analyzed for (1) area of surveillance, (2) dynamic coupling of "cause and effect", and (3) correlation between imaging and acoustics. The initial challenge problem was loosely designed, so it was asked of the team entries to download the data and establish information as to a challenge problem - which left open many questions. The event determined was the *ball rolling sound* (cause) down the incline triggering the suspended object (identified as an *umbrella*) *motion* of sliding down the pole (effect), shown in Figure 1.

1.1 IEEE NAECON08 Grand Challenge

The IEEE NAECON 2008 committee designed a challenge problem to foster innovative competition over avionics sensors technology. [1] The goal for the grand challenge was to observe, monitor, and determine the cause-effect relations of the video images to audio sound files to determine a correlated event. Acoustic-to-video fusion can establish, associate, and verify event relationships. As such, a signal processing solution can extract features from the audio (hear) and the video (monitor) to infer cognitive processing relationships between perception and localization using standard figures of merit (FOM) – detailed in **Figure 2**.

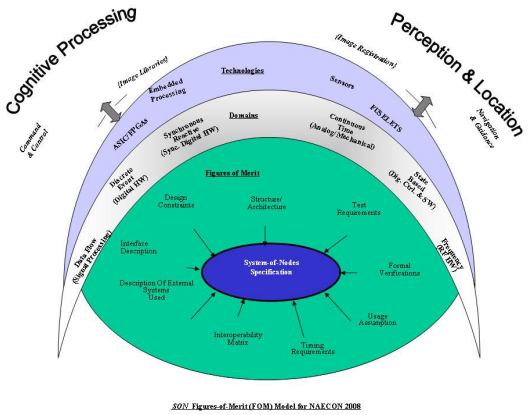


Figure 2. NAECON08 CP Figures of Merit Diagram.

1.2 Engineering Challenge Problems

Engineering challenge problems (CP), like the DARPA Grand Challenge [2], are important for many reasons. A CP definition fosters creativity not envisioned by the designers of the research area as it allows team entries to solve the problem in their own unique way. Second, the competitive spirit drives groups of individuals to derive and implement solutions at a faster rate. Third, there is a defined set of goals and metrics from which the team solutions can be compared to each other. Another aspect is that the fixed competition deadline determines the end

[1] Note: This paper summarizes the NAECON08 winning entry in that the First Grand Challenge Entry did not stipulate a formal paper but only a presentation. Subsequent contests require a submitted entry paper.

solution. Since the competition comes with minimal support, solutions are cheap as the reward is the status of winning the competition. Together CPs enable an efficient and effective solution that adheres to these standard metrics [3]:

Timeliness: Fixed Deadline
Accuracy: Determined by the evaluation criteria
Confidence: Evaluators compare the results
Throughput: Trial and error of different solutions
Cost: Cheap as minimal support is needed

As a quick summary, the following pieces constitute a "Challenge Problem:" [4]

- *Problem Definition*: The scope and significance
- *Data*: Applicable data for the defined problem
  - Tools for reading and processing data
  - Suggestions on training and test sets
  - Characterization of the data
- *Goals*: Research challenge questions and suggested experiments
- *Metrics*: Guidance on reporting results
- *Tools*: Baseline code and results: Shows reproducible minimum performance for the defined problem

Section 2 introduces the team *Belief-Level Association Signals Correlation Handling* (BLASCH) entry. Section 3 presents the biologically-inspired information fusion algorithm. Section 4 shows the utilization of signal-processing techniques and Section 5 presents results. Finally, Section 6 presents our conclusions and discusses future research directions.

2 BLASCH Solution

Team entries for a challenge problem consist of many aspects including designing the team name. The name of this entry is based on the participant; albeit the humor associated with the contributor. To combine the audio and video information follows from signals correlation, audio to video feature association, data filtering (handling), and the belief processing. Information Fusion (IF) is based on association, filtering, and correlation processes evident in the Kalman Filter. The Kalman Filter typically is applied to tracking data (position, velocity, and acceleration). To enable the ability to simultaneously track and identify (STID) objects of interest, the *belief-filter* (BF) was created to not on track, but determine what is being tracked by establishing belief IDs. [5, 6, 7]

The BF mirrors the human cerebral cortex in audio and imagery processing. **Figure 3** shows a bio-inspired object recognition and tracking based on the 1D audio and 2D video analysis of moving and stationary targets. [8]

The *pre-frontal association cortex* is the cerebral answer to associating signals in the audio and visual spectrums coming from the Lateral Geniculate Nucleus (LGN) and the Medial Geniculate Nucleus (MGN). As part of the thalamus, the LGN is the primary processing center for visual information received from the eye's retina and routes information directly to the primary (striate) visual cortex. The MGN is a thalamic relay between the inferior colliculus and the auditory cortex, which is thought to process frequency, intensity, and binaural information.

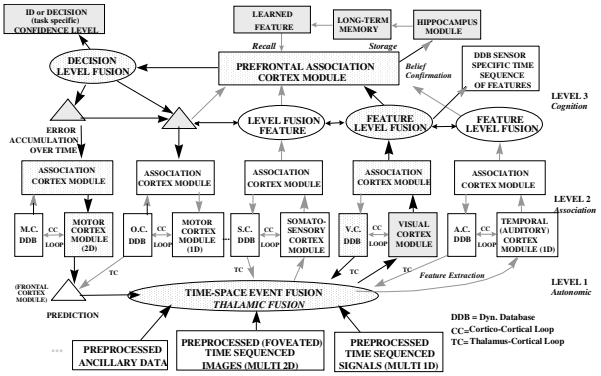


Figure 3. The Physiological-Motivated Fusion Architecture.

Biologically inspired models for sensor fusion are naturally quite common for various applications as robotics and audio visual human computer interaction. Murphy [9] developed a model based on the superior colliculus for mobile robots after reviewing the psychophysical and neurophysical literature. There has been a long history to psychophysically model the human auditory system [10] for such things as cochlear implants as well motivations for neurophysical computational models for vision [11] for designing image processing capabilities such as security surveillance.

The fusion of auditory and visual information also has parallel developments in auditory augmented perception through experiments such as those by Wallach [12] and man-machine cognitive signal processing developments for multimedia [13].

Recent trends in biologically inspired information fusion analysis include : (a) *biological understanding* such as visual modulation of auditory signals [14], (b) *cognitive developments* over user coordination (i.e. high-level fusion) [15], and (c) *distributed processing* models between sensors and agents [16].

For this entry, we correlate the movement (or changes) in the audio signal to that of the video. For the audio, we assume that a sound frequency is detected from the moving object and that a change in the image correlates to the movement of the object. Associating the audio and video information determines the cause-effect relationship. Next, we show the biologically-based signal processing solution based on feature analysis.

3 Bio-Inspired Solution

The biological solution is based on the ability of the

human brain to integrate sensed signals. Specifically, the *thalamus* receives, processes, and routes sensed information to cortex. In the thalamus, the video and audio data is processed and might have computational functions of time stamping, correlation, and filtering.

3.1 Video Solution

For example, in *visual processing*, the motion (from the magnocelluar pathway) and object recognition (i.e. color and shapes from the Parvocellular pathway) are processed. **Table 1** provides the functions of the pathways (as adapted from Wikipedia). [17]

M: Magnocellular (Layer 1 and 2)
P: Parvocellular (Layer 3, 4, 5 and 6)
K: Koniocellular cells (or "interlaminar")

Table 1: Thalmo-Cortical Pathways

	Size	Function	Info Type
M	Large cell bodies	Short time to process information. This system operates quickly but without much detail.	Transmits rod information necessary for ***perception of form, movement, depth***, and brightness differences.
P	Small cell bodies	Use a relatively long time to process information. This system operates more slowly and with lots of information about details. For example, these cells carry color information	Transmits information from the long- and medium-wavelength cones ("red" and "green" cones) for the perception of color and fine details.
K	Very small cell bodies	Linked with integrating Somatosensory/proprioceptive information with visual and color perception.	Transmits information from the short-wavelength "blue" cones.

http://en.wikipedia.org/wiki/Lateral_geniculate_nucleus

Blasch utilized collected thalamic evidence and theories to establish an information fusion spatial-temporal biological processing model to incorporate such signal processing techniques as signal registration, Kalman Filtering estimation, data association tracking, object fusion, pulse-code neural network (PCNN) feature analysis [18], and belief reasoning. **Figure 4** shows an illustration of the pathways used as a model for the signal processing methods. Note that it is hypothesized a linking between V3 and V4 for simultaneous tracking and ID.

Figure 5 shows the processing steps modeled of the human thalamus for signals processing while **Figure 6** shows the columnar structure associated with the processing functions of the thalamus links. The axons that leave the LGN go to V1 visual cortex. Both the magnocellular layers 1-2 and the parvocellular layers 3-6 send their axons to layer 4 in V1. Within layer 4 of V1, layer 4cβ receives parvocellular input, and layer 4cα receives magnocellular input. However, the koniocellular layers (in between layers 1-6) send their axons to layers 4a in V1. Axons from layer 6 of visual cortex send information back to the LGN.

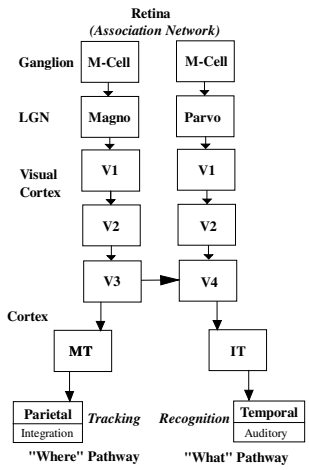


Figure 4. Track and ID Thalamic Pathways.

Visual area V5, also known as visual area MT (middle temporal), is a region of extrastriate visual cortex that is thought to play a major role in the perception of motion, the integration of local motion signals into global percepts, and the guidance of some eye movements to ID objects.

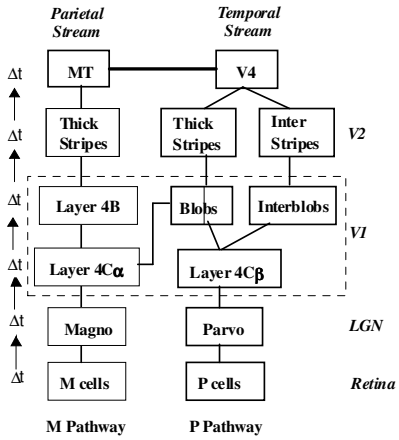


Figure 5. Track and ID Thalamic Pathways expanded.

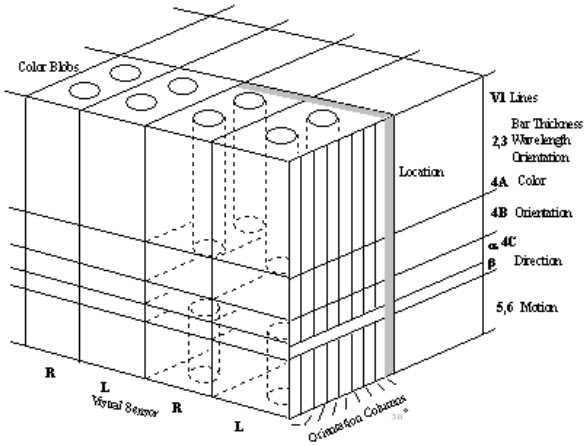


Figure 6. Columnar Structure of Thalamic Links.

Figure 7 shows the integrated visual processing of video feature extraction which is inspired by the thalamic processes. These features were incorporated into a physio-spatial-temporal simultaneous track and ID model of object tracking utilized for the Grand Challenge entry.

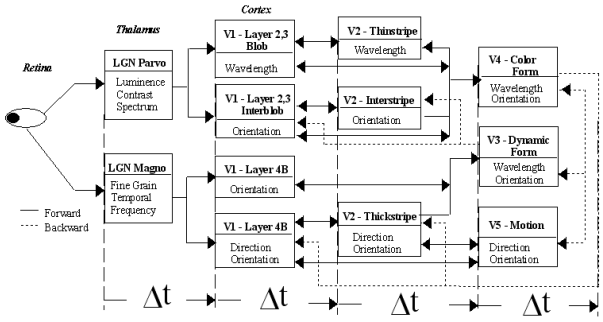


Figure 7. Visual Pathway Feature Extraction.

3.2 Audio Solution

The human brain fuses information from a variety of modalities to locate, track, and identify targets. Vision-based tracking, which uses a 2D signal, is able to accurately identify and locate objects, however it requires more processing time than 1-D auditory systems. Auditory data can cue of an impending event. Auditory systems can locate and identify objects based on interaural time difference (ITD) and interaural intensity difference (IID) fusion. [15] We utilize the ITD and IID to locate the distance of the target from the different camera locations and the direction of movement; however, the accuracy of the data was not detailed enough to refine the event perception from a general cause-event relationship. We utilized the audio intensity to determine potential causes.

4 Signal Processing Techniques

The BLASCH CP entry detected signal changes within a movie file as well as linking the audio data to the images. The analysis consisted of determining changes occurring in frequency, light/intensity, space, and time. Using concepts of convolution, filters, signal energy, and information fusion; various preprocessing techniques were used prior to audio-to-video association and correlation for cause-event determination.

Two analyses were performed to detect the presence of a ball moving down the incline which causes the umbrella to slide down the pole as shown in **Figure 8**. From two streaming audio and video data sets preprocessing was needed to determine the combined perspective of the event (i.e. stereoscopic analysis) and audio-to-video registration. The first was the analysis of the audio feed synchronized with the video, and the second was the analysis of the video images for features used for object tracking.

The file that was input to our system was 3a.mwv (audio) and video1p.avi, (video). Within this file, the audio and video were recorded at slightly different rates and estimation of the audio kHz and video frame per seconds (fps) was needed. An empirical conversion rate was used to process the audio and video at the same time with *N* audio samples per video sample. Various values of *N* were used for FOM assessment; however, further analysis would be needed as a close conversion was established to determine and localize the cause-event relationship. The audio was recorded in stereo and Doppler results could also be considered for movement analysis.

Figure 8. Scene of Challenge Problem Data.

The following analysis was performed:

AUDIO
- A. FFT of each synchronized frame
- B. Spectogram of each frame
- C. Pwelch of each frame
- D. Lowpass filter and Bandpass filter each frame

VIDEO
- A. Convolution filters for scene content
- B. Sobel operator for segmentation
- C. Image subtraction for event detection

ADUIO-VIDEO Fusion
- A. Cause-to event correlation and association
- B. Belief filtering for Simultaneous Track and ID

4.1 Audio Processing

The audio portion analysis required conversion from a stereo signal into a mono signal for audio processing. The ***time-response plot*** of this mono signal can be seen in **Figure 9**.

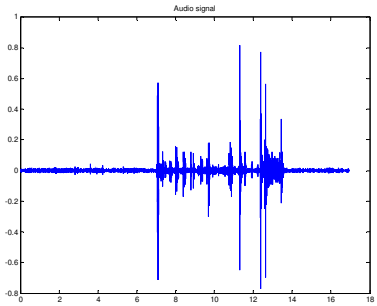


Figure 9. Audio Signal.

Using a ***Fourier analysis*** over the input audio and video files, we were able to link the audio (cause) with the video

(effect). The FFT was performed per frame to isolate frequencies that would be present during the cause rather than not. After analyzing the results, it became evident that during the strongest portions of the ball movement, fewer harmonic frequencies are present. **Figure 10** details the FFT before and during the ball movement showing the harmonic changes while Figure 11 is a plot of the global FFT change.

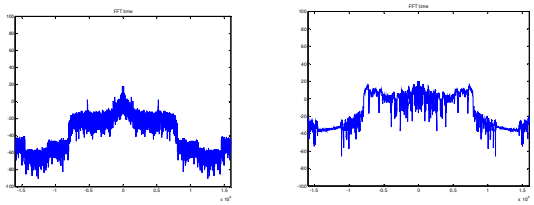


Figure 10. FFT.

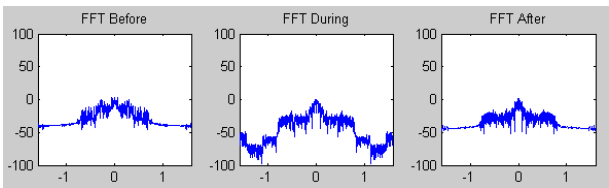


Figure 11. FFT.

Also, a ***log magnitude FFT*** of the audio signal analyzed on a frame by frame basis yielded the same properties as the FFT for event detection. To refine the audio processing, we looked at the spectrogram.

A **spectrogram** is an image that shows how the spectral density of a signal varies with time. The spectrogram was performed to compute the Short Time Fourier Transform (STFT) of each audio segment. STFT analysis is useful for an incomplete or streaming signal set to detect via color gradients sudden changes in the frequency pattern of the signal.

Creating a spectrogram using the STFT is usually a digital process. [19] Digitally sampled data, in the time domain, is broken up into chunks, which usually overlap, and Fourier transformed to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image; a measurement of magnitude versus frequency for a specific moment in time. The spectrums or time plots are then "laid side by side" to form the image or a three-dimensional surface. The spectrogram is given by the squared magnitude of the STFT of the function:

$$\mathrm{Spec}(t, \omega) = |\,\mathrm{STFT}\,(t, \omega)\,|^2$$

The *spectrogram*, as seen in **Figure 12**, was obtained by first normalizing the signal amplitudes before applying the spectrogram function. The spectrogram plot displays the color-based visualizations of the evolution of the signal power spectrum as this signal is swept through time. Using the spectrogram, we were able to validate the time periods of the ball movement.

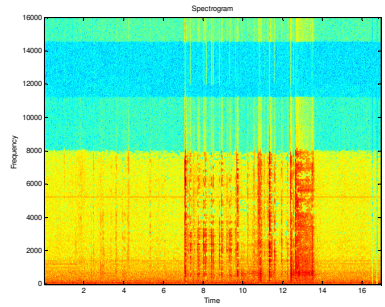


Figure 12. Spectrogram of Noise Time Sequences.

Figure 13 details the spectrogram before, during, and after the ball movement (cause).

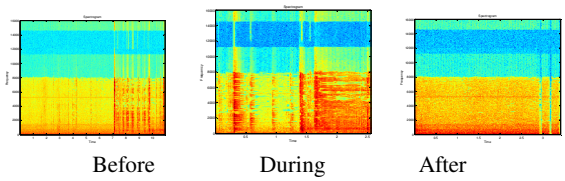


Figure 13. Spectrogram Around Event.

The ***pwelch analysis*** performs a power spectral density plot of the audio frame. The pwelch was useful in displaying a large overall gain of the signal as the ball movement began and ended as well as threshold for feature extraction. **Figure 14** details the pwelch which correlated an event cause.

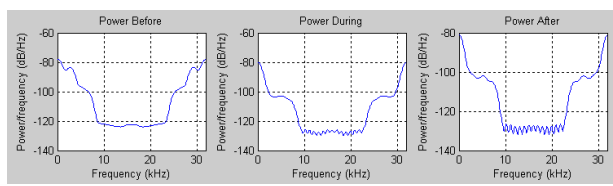


Figure 14. PWELCH.

Other methods of analysis included the power analysis and signal processing filtering methods. The ***power*** of the signal was plotted frame by frame as a detection threshold by taking the average of each point to obtain an average power. Power analysis was compared to a set threshold which would trigger an *alarm* (i.e. hypothesized event).

The results from the FFT analysis indicated a window of frequencies could be detected at the incident of the ball movement. Performing a ***lowpass filter*** was a starting point to isolate the signaler frequency. A ***bandpass analysis*** was used to isolate a frequency window to estimate the ball movement starting and verifying stopping. The ***FIR band pass filter*** designed to pass a window of frequencies emphasizing frequencies when the ball starts and terminates. In summary, there was ample evidence to detect a cause and cue video analysis.

4.2 Video Processing

In order to process the video, a ***convolution filter*** was designed to extract the vertical lines from the movie, shown in **Figure 15a**. To get the horizontal edges, the same image data was convolved with the transposed filter vector as seen in **Figure 15b**. Note that the *pole* and other objects are easily detected in the vertical features image and the horizontal lines reveal the *incline*.

It was determined that the convolution filter gives the geo-spatial scene analysis for locating the incline (cause) and the pole (effect). With *a priori* knowledge of the cause-event we are trying to detect, the location of the environmental objects assists in determining the start of the ball, the incline from which the ball roles, and the location of the umbrella (Figure 1 versus Figure 8) sliding down the pole evident from the Figure 15. It was noted that the sound of the ball (cause) has an result of the umbrella sliding down the pole (effect). The information evident in the video was used as parts of the challenge problem goals – to utilize the sound of the ball movement and its subsequent event triggering of the umbrella movement.

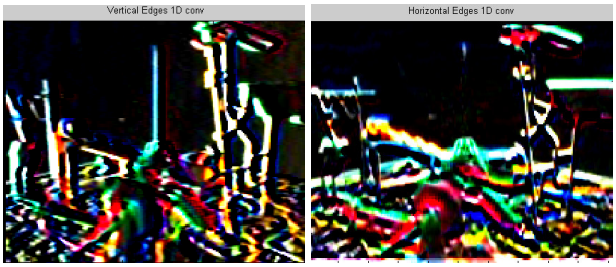


Figure 15. Convolution filtering for (a) vertical edges and (b) horizontal edges for scene content. Note the pole (blue) in the left image and the incline (white) in the right image.

Shown in **Figure 16**, the ***Sobel Operator*** performs a 2D convolution for edge detection on the video frame displaying the amplitude changes of movements. The Sobel operator was utilized for intensity changes.

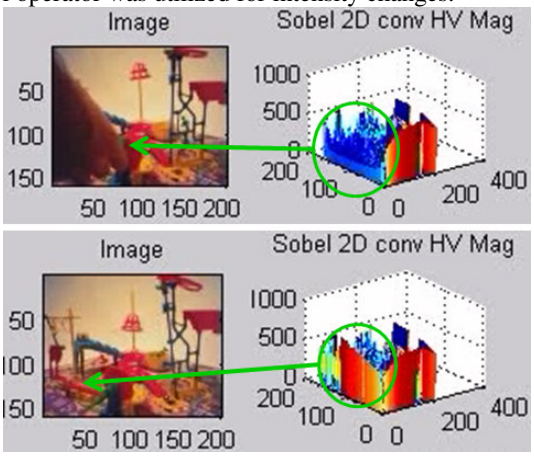


Figure 16. Sobel Operation features for ball initiation.

Using the Sobel Operator, an event window was created from the horizontal and vertical edge features. After the image was converted from an NTSC to RGB signal, a red filter (umbrella was red) was applied to the image. Also, the average image pixel value was plotted in a frame by frame basis over the length of the video. In a chipped section of the image, the results showed that there was a change in the total red pixel count when the event occurred, as seen in **Figure 17**.

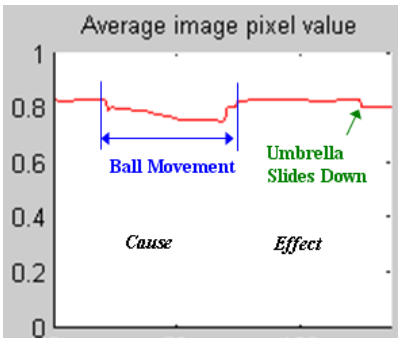


Figure 17. Movement analysis.

5 Combined Audio-Video Fusion

Once the audio and video signals were separated, we were able to register them frame by frame. The audio cued the video processing. In the video signal, we were able to establish scene content, feature extraction, and filter initialization through convolution.

Additionally, we conducted an FFT of each frame for both audio and video. We noticed dramatic changes in the FFT graph when the ball starts (audio) and when umbrella moves (video). Although the FFT could not be predicted, we did utilize them for event detection. Figures 18 and 19 show the signal processing feature analysis.

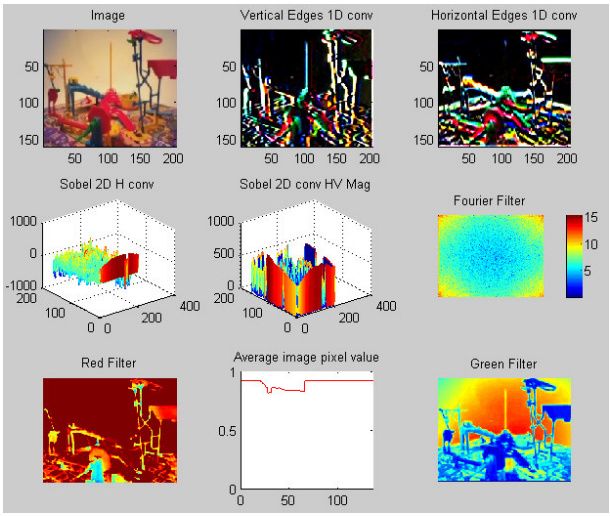


Figure 18. Video Processing.

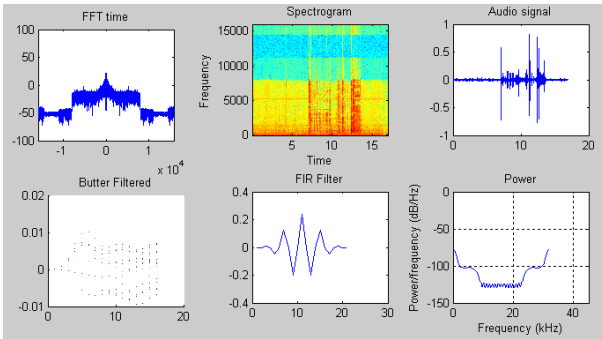


Figure 19. Audio Processing.

For the video, we explored feature analysis and object tracking. We isolated the scene change through image subtraction. Image subtraction and threshold-detection design aided track initiation [20] for the ball and umbrella for audio/video processing for perception (ID) [21] and localization (tracking) [22].

With the signal processing and biologically-inspired feature analysis, we incorporate the features into the belief filter to enable event prediction and target analysis.

5.1 Belief Filtering

The *belief filter* was developed in response to: (a) typical processing of only kinematic data of the *Kalman filter*, (b) standard non-physical state meanings in the *Information filter*, and (c) the biological motivated development of an integrated target track and identity processing versus a sequential computational approach. Table 2 formulates a taxonomy of these filtering techniques noting that evidential analysis (e.g. Dempster-Shafer theory) defaults to Bayes when there is no conflict in the data.

Table 2. Filter Relationships

ID / Track	**Probability Theory** (Bayes)	**Evidence Space** (Dempster-Shafer)
Kinematic	Kalman Filter	Belief Filter
Information States	Information Filter	Augmented Belief Filter

For the analysis, we utilized the audio-video feature measurements of (x, y, B) for position and beliefs; respectively,:

$$z(k) = H\,x(k) + v(k) \tag{1}$$

where $v(k) \in \Re^{2+B}$ is a *white noise* vector $v(k) \sim N(0, R)$ and $H \in \Re^{2+B}$ is the measurement matrix; where B affords multi-feature $B = \{1, \ldots. f_B\}$ processing.

The updates for the filter models are computed via the Belief filter equations (which are essentially the Kalman filter equations with augmentations for using the mixed estimate and mixed covariance between the kinematic and feature identity representations) :

$$x_j(k \mid k-1) = \Phi \hat{x}_j^0(k-1|k-1) + \Gamma\, a_j(k) \tag{2}$$

$$P_j(k \mid k-1) = \Phi \hat{P}_j^0(k-1|k-1)\,\Phi^T + \Gamma\, Q_j\, \Gamma \tag{3}$$

$$S_j(k) = H\,P_j(k \mid k-1)\,H^T + R_j \tag{4}$$

$$v_j(k) = z(k) - H\,\hat{x}_j(k \mid k-1) \tag{5}$$

$$K_j(k) = P_j(k \mid k-1)\,H^T S_j^{-1} \tag{6}$$

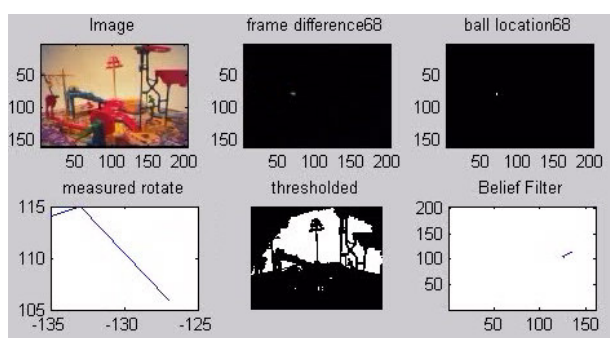
$$\hat{x}_j(k \mid k) = \hat{x}_j(k \mid k-1) + K_j(k)\,v_j(k) \tag{7}$$

where $x_j(k \mid k) \in \Re^{4+B^2}$ is the state estimate, $P_j(k \mid k) \in \Re^{(4+B^2)\times(4+B^2)}$ is the covariance of the state estimate error, $v_j(k) \in \Re^{2+B}$ is the innovation process, $S_j(k \mid k) \in \Re^{(2+B)\times(2+B)}$ is the covariance of the innovation process, and $K_j(k \mid k) \in \Re^{(4+B^2)\times(2+B)}$ is the Belief filter gain.

The mixed kinematic-to-ID feature information corresponds to spatial pose, frequency changes, and temporal predictions. [23, 24] For example, *kinematic pose* estimates dictate ID feature extraction and ID length-to-width feature estimates determine the target velocity direction.[25] Other examples are Doppler processing in radar [26] and illumination changes in EO [27].

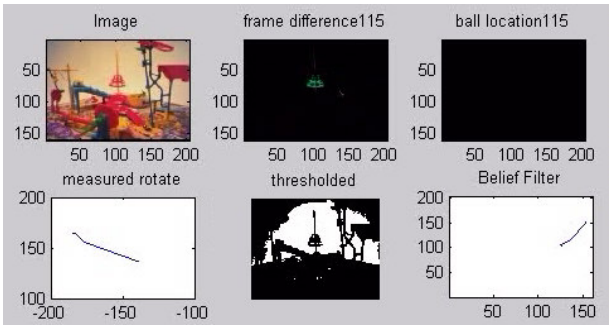
The belief filter is overkill for the data set as the ball is determined from a single color (frequency) feature where as the umbrella has color and 2D shape features. The belief filter demonstrated conducts ball detection, movement estimation, and termination using features extracted from the audio and video data. The audio information was used for determining a "Cause". The "Cause" was utilized to *cue* for an impending movement of the umbrella – "Event". As the ball was moving a second belief filter was created to initiate, monitor, and detect and umbrella movement estimate which is the "Effect". Further events could be grouped for analysis. [28]

Presented in **Figures 20 and 21** is the detection of the ball movement and the initiation of the track belief filter.



(a) Ball Detection (Belief Filter Start)

(b) Ball at incline end to cue event



(c) Belief Filter Track from cause to event
(ball ends and umbrella motion starts).

Figure 20. Tracking Results.

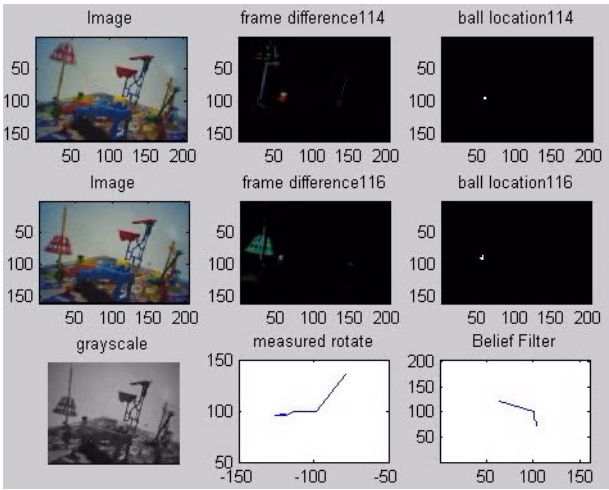


Figure 21. Tracking Results from the second perspective.

6 Conclusion

The paper described the BLASCH entry for the NAECON08 Grand Challenge. Given the open exploratory nature of the problem description, we formulated details associated with a "cause-event" cognitive processing solution. We utilized the belief filter to process the audio and video data by way of signal processing feature extraction, estimation of feature and target changes, and cognitive processing of events (given some analysis of the context – scene and hypothesized events).

To expand challenge problems, IEEE NAECON is working with the ATR Center to expand and clarify future Challenge Problems. [29] The IEEE NAECON09 Challenge Problem is "Signals of Opportunity".

7 References

[1] R. Ewing, *IEEE 2008 NAECON Grand Challenge Guidelines*, www.naecon.org.

[2] G Seetharaman, A. Lakhotia, E. Blasch," Unmanned Vehicles Come of Age: The DARPA Grand Challenge," *IEEE Computer Society Magazine*, Dec 2006.

[3] E. Blasch, I. Kadar, K. Hintz, J. Salerno, C. Chong, J. Salerno, and S. Das, "Resource Management Coordination with Level 2/3 Fusion", *IEEE AES Magazine* , Mar. 2008.

[4] E. Blasch, C. Yang, L. Hong, and L. Bai, "Performance Assessment of Multitarget Tracking Algorithms in the Real World," *SPIE08*, April 2008.

[5] E. Blasch, "Biological Sensor Fusion for HCI Security Identification," *Joint Conference on Information Sciences*. 1998.

[6] E. Blasch, "Biological Data Association for Combat Identification Using Belief Filters," *Joint Service Combat Identification Systems Conference,* JHU APL, Laurel, MA, Nov 1998.

[7] E. P. Blasch and L. Hong, "Sensor Fusion Cognition using belief filtering for tracking and identification," *SPIE Vol. 3719,* 1999.

[8] E. Blasch and J. Gainey, Jr., "Physio-Associative Temporal Sensor Integration," *SPIE Int. Symp.*, April 1998.

[9] R. R. Murphy, "Biological and cognitive foundations of intelligent sensor fusion," *IEEE Trans. On Systems, Man, and Cybernetics*, Vol. 26, No. 1, 1996.

[10] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, Inc., 1982.

[11] M. S. Landy and J. A. Movshon (Eds.), *Computational Models of Visual Processing*, MIT Press, 1991.

[12] H. Wallach, *On Perception*, Quadrangle–NY Times Book, 1976.

[13] J. W. Fisher. III and T. Darrell. "Learning cross-modal appearance models with application to tracking." In *Proc. 2003 IEEE Int'l Conf. on Multimedia and Exposition*, 2003.

[14] J. F. Bergand and E. I. Knudsen, "Visual modulation of auditory responses in the owl inferior colliculus," *J. Neurophysiology*, March 2009.

[15] E. Blasch, "Neurophysiologically-based signals analysis for Level 5 User Refinement for target localization, " *Proc. SPIE 4731*, 2002.

[16] M. Cetin, C. Lei, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed Fusion in Sensor Networks," *IEEE Signal Processing Mag*., Vol. 23, No. 4, 2006.

[17] http://en.wikipedia.org/wiki/Lateral_geniculate_nucleus

[18] E. P. Blasch, "Biological Information Fusion Using PCNN and Belief Filters," *Int. Joint Conf. on Neural Networks,* 1999.

[19] http://en.wikipedia.org/wiki/Spectrogram

[20] S. Mori, C-Y. Chong, E. Tse, and E. P. Wishner, "Tracking and Classifying Multiple Targets Without A Priori Identification," *IEEE Trans. On Automatic Control*, AC-31 (5), 1986.

[21] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House Publisher, Boston, 1999.

[22] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS, New York, 1995.

[23] E. Blasch, J.J. Westerkamp, J.R. Layne, L. Hong, F. D. Garber and A. Shaw, "Identifying moving HRR signatures with an ATR Belief Filter," *SPIE* 2000.

[24] E Blasch and C Yang, "Ten methods to Fuse GMTI and HRRR Measurements for Joint Tracking and ID", *Fusion 04*, 2004.

[25] C. Yang, E. Blasch, W. Garber, and R. Mitchell, "A Net Track Solution to Pose-Angular Tracking of Maneuvering Targets in Clutter with HRR Radar," *Asilomar Conf. on Signals,* 2007.

[26] C. Yang and E. Blasch, "Moving ATR Analysis Using a Simultaneous IMM Track/Belief ID Analysis," *Proc. SPIE Great Lakes Photonics Symp* , 2004.

[27] E. Blasch and B. Kahler, "Multi-resolution EO/IR Tracking and Identification" *Fusion 05*, July 2005.

[28] T. Connare, E. Blasch, J. Schmitz, *et. al*., "Group IMM tracking utilizing Track and Identification Fusion," *Proc. of the Workshop on Estimation, Tracking, and Fusion*, 2001.

[29] G. Arnold, T. Ross, L. Westerkamp, L. Carin, and R. Moses, "The ATR Center and ATR Pedia," *Proc. SPIE 6978,* 2008.