Information Fusion: the Concepts, the Technology, the Community, and Modern Research Challenges

Dr. James Llinas
Research Professor (Emeritus), Executive Director
Center for Multisource Information Fusion
State University of New York at Buffalo, USA
llinas@buffalo.edu
About the “University at Buffalo”

• “University at Buffalo” aka State University of NY at Buffalo
• New York State’s largest and most comprehensive public university
• Member of the Association of American Universities
• Funded research activity in the range of US$350M per year
• Selected Research Centers:
  – Center for Multisource Information Fusion
  – Natl Center for Geographic Information and Analysis
  – National Center for Ontological Research
  – Virtual Reality Laboratory
  – Center for Information Systems Assurance
  – Lab for Advanced Network Design, Evaluation and Research
  – Wireless and Networking Systems Lab
  – Semantic Network Processing Systems Research Group
  – The Center for Unified Biometrics and Sensors
  – Center for Computational Research (Supercomputing)
  – Center for Document Analysis and Recognition
• Current enrollment approx 30K+ students, 18K undergrad, 12K grad
Around the Campuses: Suburban-South-Downtown
CMIF OVERVIEW

- **Mission:** Information Fusion and related areas primarily but not exclusively for defense and homeland security applications

- **Basic and Applied Research in:**
  - Multiple-sensor and instrumented systems
  - Synergistic Human-Multisensor systems
  - Real-time Decision-making using Hierarchical Fusion
  - Graph Theory and Optimization for Level 2/3 Fusion
  - Multi-modal/spectral information environments (speech+text+imagery+RF sensor+human input)

- **Applications:**
  - Defense: Intelligence/Surveillance/Reconnaissance; Tactical Applications; Homeland Security
  - Non-Defense: Robotics; Conditioned-Based Maintenance; Medical; Transportation; Geology; Natural Disasters/Crisis Mgmt

- **History and Funding:**
  - Started in 1996 with Air Force Research Lab Contract
  - Typical funding activity ~US$4M/year

- **Scholarly:**
  - Long-standing member of “JDL” fusion group and First President of Intl Society for Info Fusion
  - Extensive publishing by CMIF PI Team including books, Jl papers, conference papers and review boards
  - “Critical Issues” Workshops—8 years
  - CMIF is unique in American Universities as a research activity focused on IF technology for DHS/DoD
  - Consortium development to include other universities (SU, RIT and PSU) and industrial partners and development of a Graduate-level pgm in Data Fusion
The Concepts
Everyday Data Fusion

- Sound
- Smell
- Taste
- Images
- Touch
- Pain
- Balance
- Temperature
- Body Awareness (Proprioception)

Multinodal Fusion

Augmented Sensing

Robotic Multisensor Fusion
History of Information Fusion

- **Dates to circa 1970’s** — fairly young in the sense of technological history—a maturing technology/field of study

- **Driven by defense and intelligence needs**
  - Originally as a “data compression” device to digest huge amounts of sensed data as sensors advanced in capability (a “push” requirement)
  - Later as an important element for decision support (a “pull” requirement)

- **Matures to very broad range of application**
  - Robotics, medicine, imagery/remote sensing, intelligent transportation, conditioned-based maintenance, biometrics, etc
What is Information Fusion?

"Information fusion is an [Information Process] dealing with the:

- \{Association, correlation, and combination\} of data and information from
- \{Single and multiple sensors or sources\} to achieve
- \{Refined estimates\} of parameters, characteristics, events, and behaviors for observed entities in an observed field of view

- It is sometimes implemented as a Fully Automatic process or as a [Human-Aiding process] for Analysis and/or Decision Support
Most Simply--

Observation System

Multiple types of data
--various types of information
--redundant
--and complementary)

“Associated” or “Correlated” to :
--the same object
or event
or behavior

So that estimation algorithms (mathematical techniques)—or—automated reasoning methods (artificial intelligence techniques) can produce better estimates (than based on any single type of data)

Real World

Multiple types of data Related to things To improve estimates about those things

Observations Association Estimation
(Multiple) of Observations

These Basic Ideas are Transferable to Many Types of Problems
Basic Role of Fusion: Adaptive, Recursive Estimation

Six Informational / Knowledge Components:
- Observational Data
- A Priori Dynamic World Knowledge Model (Deductive)
- Contextual Information
- Runtime Learned Knowledge
- Tacit and Explicit Human Knowledge
- (Network) External Obsvns and Estimates
### Information Fusion Exploits Sensor/Source Commonalities and Differences

#### Unknown Moving Object

<table>
<thead>
<tr>
<th>DETECTION</th>
<th>KINEMATICS</th>
<th>CLASSIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONFIDENCE</td>
<td>COVERT COVERAGE</td>
<td>RANGE</td>
</tr>
<tr>
<td>RADAR</td>
<td>FAIR</td>
<td>POOR</td>
</tr>
<tr>
<td>EO/IR</td>
<td>FAIR</td>
<td>FAIR</td>
</tr>
<tr>
<td>C3I</td>
<td>FAIR</td>
<td>GOOD</td>
</tr>
</tbody>
</table>

#### Data Preparation
- GOOD

#### Data Association
- GOOD

#### State Estimation
- GOOD

#### Classification
- GOOD
Operational Benefits of Multiple Sensor Data Fusion

- **Multiple Sensors**
  - Reliability
  - Improved Detection

- **Multiple Platform Sensors**
  - Extended Coverage (spatial and temporal)
  - Improved Spatial Resolution

- **Diverse Sensors**
  - Robustness (Weather/visibility, Countermeasures)
  - Improved Detection

**State Estimates of Reduced Uncertainty And Improved Accuracy**
The Technology
The Technology: Scientific Foundations of the Data Fusion Process

- Sensing Technologies
- Signal Processing
- Combinatoric Optimization
- Mathematical And Symbolic Estimation Techniques
- Human Computer Interfacing
- Human Factors and Human Engineering
- Visualization
- Virtual Reality
- Decision Science
- Control Theory
- Sensor Networks
- Modeling Tactical Phenomena
- Signal Propagation

A Process to ESTIMATE conditions in the Real World from Observational Data

Broadly Multidisciplinary
**Choices in Fusion Approaches**

<table>
<thead>
<tr>
<th>Fusion Approach</th>
<th>Nature of Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>* No Fusion</td>
<td>* Best single-source approach</td>
</tr>
<tr>
<td>* Non-fused but adaptive</td>
<td>* Phased application of single sources:</td>
</tr>
<tr>
<td></td>
<td>-- Multiresolutional</td>
</tr>
<tr>
<td></td>
<td>-- Cueing</td>
</tr>
<tr>
<td>“Real” Fusion</td>
<td>Synergistic; Adds Information, Reduces Uncertainty</td>
</tr>
<tr>
<td>-- Limited Fusion</td>
<td>-- Single Source + A Priori Info</td>
</tr>
<tr>
<td></td>
<td>-- Occasional in time, or</td>
</tr>
<tr>
<td></td>
<td>-- On demand</td>
</tr>
<tr>
<td>-- Moderate-Level Fusion</td>
<td>-- Few Sources</td>
</tr>
<tr>
<td>-- Aggressive-Level Fusion</td>
<td>-- ”All Source”</td>
</tr>
</tbody>
</table>
Some Research Strategies

• **Estimation-process-centric**
  – Signal processing (e.g., detection fusion), intersource registration/alignment, estimation algorithms
  – Input: given; Output: mathematical estimate

• **System-centric**
  – Process architecture, standards, integration
  – Process mathematics
  – Process control, estimation/decision-making interdependencies, dynamic resource mgmt
  – Human-system design
  – Input: controllable; Output: usable by a human
Architecting Systems
--Architectural Elements
--Dealing with Uncertainties
--Data Association, a core function
Data Fusion Tree Node

Now exploit the multiple observational data for a fused estimate

Things that can cause expected observations

How it is that observations are related to the entities or objects
(A notion of “closeness”—a “score”)

Optimally assigning the observations to an estimation process which is estimating a parameter of interest for the entity/object

Prior Data Fusion Nodes & Sources

Data Preparation (Common Referencing)

Source/Sensor Status

• Detect and resolve data conflicts
• Convert data to common time and coordinate frame
• Compensate for source misalignments

Data Association

Hypothesis Generation

Hypothesis Evaluation

Hypothesis Selection

State Estimation & Prediction

User or Next Fusion Node

Data Fusion Tree Node

• Gating and generation of feasible and consistent association hypotheses
• Scoring of data associations
• Select, delete, or feedback data associations

Source/Sensor Status

- Time and coordinate frame (e.g., blue, red)
- Aggregate states (e.g., blue, red)
- Source/sensor misalignments
- Feed forward source/sensor status

- Feasible and consistent association hypotheses
- Score data associations
- Select, delete, or feedback data associations
## Architectural Elements

<table>
<thead>
<tr>
<th>Data Fusion Node</th>
<th>Common Representation for all Data Fusion Processes</th>
<th>Data Fusion Network</th>
<th>Common Representation for all Data Fusion Architectures</th>
<th>Integration of Data Fusion and Resource Management Networks</th>
<th>Common Representation for all Information System Architectures</th>
</tr>
</thead>
</table>

**Data Fusion Node**
- Common Representation for all Data Fusion Processes

**Data Fusion Network**
- Common Representation for all Data Fusion Architectures

**Integration of Data Fusion and Resource Management Networks**
- Common Representation for all Information System Architectures

![Diagram](image-url)
Dealing with Uncertainty

Second Order Uncertainty and Imprecise Probability

- Theoretical aspects of Second Order Uncertainty
  - Focus on Epistemic Uncertainty (limitations in knowledge)
  - Aspect of degrees of satisfaction of the Kolmogorov Axioms, especially the Additivity Axiom
- Walley (1991) shows that imprecise probabilities satisfy the principles of coherence—relaxes need to satisfy Additivity
- Leads to range of alternatives, each of which satisfy “relaxed” Additivity Axioms

Most problems involve both Aleatory and Epistemic Uncertainties
Data Association Basics
--What measurement goes with what entity?

Note: All Msmts are Inside Feasibility Gates

Leads to the formulation of a classic OR Assignment problem with usual repertoire of solutions
Association and Assignment Optimization

Need some type of Mapping that determines a good way to allocate Observations To Tracks

M Observations From N Sensors

DATA ASSOCIATION

“Assigned” Observations Resulting from some “Best” way to decide which Observations should be “given” to each State Estimator

Tracks “T”

STATE ESTIMATION & PREDICTION
One Taxonomy of Assignment Problem Solutions

- Single Scan
  - Nearest Neighbor
    - Sequential NN
    - Munkres or Auction NN
    - Standard NN
    - Most Likely Event NN
  - All Neighbor
    - Probabilistic Data Association (PDA)
    - Joint Probabilistic Data Association (JPDA)
- N-Scan
  - MHT
    - Branching
    - Detection Oriented
    - Structured Branching
  - Dynamic Programming (Viterbi)
    - Target Oriented
    - Track Oriented
Categories of Data Association

“Report-to-Report”

Associating Data from Multiple Sources at a Fusion Node
(“Measurement-to-Measurement Association”)

“Report-to-Track”

Associating Data and Estimates at a Fusion Node
(“Measurement-to-Estimate Association”)

“Track-to-Track”

Associating Multiple Estimates at a Fusion Node
(“Estimate-to-Estimate Association”)

- SOURCES
- DATA ALIGNMENT
- DATA ASSOCIATION
- STATE ESTIMATION & PREDICTION
The Community
The Early Community

• Early 1980’s
  – Mostly US, UK, Australian, Canadian Defense related (in UK eg Royal Signals and Radar Establishment, DRA, etc before DERA)
  – 1985: First US “National Symposium on Sensor and Data Fusion”, NSSDF
    • US only (attempts at NATO integration fail)
    • Classified
    • Ongoing today
  – US: Joint Directors of Laboratories
    • Aids in unifying terminology and concepts
  – 1990: First unified text published
Second Generation

• Early 1990’s to early 2000’s
  – Still rather ad hoc through early 90’s
  – Mid 90’s sees evolving structure

• Mid-late 90’s
  – IEEE Conference on Multisensor Fusion and Integration, from 1995 on
  – International Conference on Information Fusion, annually, from 1998 on
  – International Society of Information Fusion, established in 1999

• Early 2000’s
  – International Journal of Information Fusion, 2000
  – Journal of Advances in Information Fusion, 2003
  – Many other conferences (e.g., in SPIE)
  – Textbooks begin to flow
Current Status

• Community
  – Stable but needing a broader technological view as capability for L1 matures and challenges of L2, L3 are addressed
  – Structured outreach required

• Operating Domain
  – Too defense-oriented; multi-domain outreach also required

• Fusion process and concepts
  – Need structured extension eg to distributed, networked case and L2, L3 processes
  – Control-theoretic aspects to be addressed
  – Frameworks for cost-effective development
Modern Research Challenges
One List – No particular order

• Achieving scalability and robustness
  – Beyond one-algorithm/process solutions
• Structured, standardized strategies for contextual exploitation
  – Extends, as a basic research topic, to hybrid deductive/inductive systems
• Holistic strategies for distributed fusion processes
  – Eg linking Information-sharing strategies with network fusion operations
• Dealing with weak knowledge problems (world dynamics poorly understood)
  – Second-order uncertainty, response-based balanced designs
  – Extends to the case of Situation Management
• Overall Hard and Soft Fusion process designs and methods
• Improved techniques for Test and Evaluation, Metrics
A Taste of the Hard + Soft Data Fusion Problem
The Network Enterprise, Irregular Warfare, and Fused Situational Awareness

Net Enterprise Services

“Hard” Sensor Data
Calibrated, Precise

“Soft” Sensor/Contextual Data
Uncalibrated Human Observers/Uncalibrated Sources
Observations expressed in (inherently) ambiguous language

Extensive Data and Information-Sharing Enabled by Network Infrastructure
Gives Rise to a New Challenge in Information Fusion:
“HARD” + “SOFT” INFORMATION FUSION
Some Distinctions in Hard and Soft Observational Data

<table>
<thead>
<tr>
<th>Data Characteristic</th>
<th>Hard</th>
<th>Soft</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation sampling rate</td>
<td>High</td>
<td>Low</td>
<td>Imputes requirements for adaptive, retrodiction-type processing (i.e. “Out-of-Sequence Measurement” type processing), as well as agile Temporal Reasoning</td>
</tr>
<tr>
<td>Semantic Content</td>
<td>Limited to specific, usually</td>
<td>Can be conceptually broader than single</td>
<td>Imputes requirements to design an automated Semantic Labeling process, coupled to a rich Domain Ontology</td>
</tr>
<tr>
<td></td>
<td>singular Entities</td>
<td>Entities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limited to Entity Attributes</td>
<td>Can include Judged Relationships</td>
<td>Requires ability to associate and infer at multiple levels of abstraction</td>
</tr>
<tr>
<td>Accuracy, Precision</td>
<td>Relatively high, good</td>
<td>Broadly low accuracy in attributes, high</td>
<td>Imputes requirements for robust Common Referencing and Data Association</td>
</tr>
<tr>
<td></td>
<td>repeatability (Precision)</td>
<td>at the conceptual level</td>
<td></td>
</tr>
</tbody>
</table>

Humans can also judge intangibles --emotional state

This line of thought suggests *that relations are the result of a process* of some type of comparison, ie [Brower, 2001], “an act of reasoning”.

Categories of Human Input

Passive Observation

Report Credibility

Direct Interaction

Source Credibility

Third-party Reporting

Source Credibility

Audio - Text Unit

Other (Hard) Source Information

TOC
Source Characterization—very difficult to generalize

**Soft Data**

Real World Truth

**Hard Data**

Calibration (Truth) Target

**Average Human Soldier**

- Perceptual and Cognitive Errors in observation
- Error in oral expression
- Error in audio capture
- Error in audio - to text Conversion
- Error in text extraction

- Some Obsvnl Data types qualified and Generalized

- Some errors specific to obsvnl conds (need context)

- Some errors will go unlabeled, unknown

**Auto Text Extract & Semantics**

- To Common Ref, Data Association

- To Common Ref, Data Association

- $\epsilon_1$
- $\epsilon_2$
- $\epsilon_3$
- $\epsilon_4$
- $\epsilon_5$

- $d' = 1$ (lots of overlap)
- $d' = 3$ (not much overlap)

- ROC curves

- Pd (Obs Params)
Initial Prototype Approach

Soft Data Source 1

Observer 1
Source Characterization A
Linguistic Framing
Utterance (Audio)
Audio-to-Text

Observer 2
Source Characterization A
Linguistic Framing
Utterance (Audio)
Audio-to-Text

Human Filter

RDF Triples
Text Extractor

Word-sense Disambiguation
Latent Semantic Analysis

Common Referring
Inter-source Data Association
Multi-source Soft Data Estimation

Context-based Reasoner
Consistency Checking

Relevant Context Filter
Knowledge Base

WEB
Newswire
Web Harvester

Soft Data Source 2

Multiple (3) Hard Data Sources

Visual
Segmentation
Feature Extraction

Acoustic
Measurement Processing
Feature Extraction

RF
Measurement Processing
Feature Extraction

L1 Hard Fusion
Classifier
Location, Kinematics
Estimation

Fused Hard Estimate

Network

Hard+Soft Common Referencing
Hard+Soft Data Association
Hard+Soft State Estimation

Low Sampling Rate
Abstract Entities
Low Accuracy

High Sampling Rate
Specific Entities
High Accuracy

Truth

Truth