Data Mining in Widespread Video and Other Sensors

MOUNTAIN STATE INFORMATION SYSTEMS, INC.

MISSION MOUNTAIN TECHNOLOGY ASSOCIATES

27 April 2010

Systems & Software Technology Conference

Paul D. Garnett, President
619 W. Columbia Ave., Sunset #1
P.O. Box 3738
Telluride, CO 81435

(304) 367-0500: Fairmont, WV Office
(970) 728-1972: Colorado Office
(970) 708-7723: Cell
pgarnett@msisinc.com

Fred A. Palma
10240 Melbenji Ct.
Elk Grove, CA 95757

916 501-6051
Mountain State
Information Systems, Inc.
fpalma@msisinc.com

Cody A. Benkelman
100 Second St. East
Suite 204
Whitefish, MT 59937

406 270-1176
Mission Mountain
Technology Associates
cody@MissionMtnTech.com

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

PREFACE

- SSTC Conference Challenge:
  - "What are these new game changing technologies going to be?"
  - *… look into the future*
- Our Objectives:
  - Ask "*What if?*"
  - Facilitate dialogue regarding potential risks
  - Raise awareness
  - No definitive answers

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

PRESENTATION OUTLINE

- Introduction (Why are we here?)
- Objectives
- General Applications (Threats, Positive Uses)
- Scenario #1
- Scenario #2
- Full Body Biometrics
- So What? Where do we go from here?

MOUNTAIN STATE

INFORMATION SYSTEMS, INC.

INTRODUCTION

- What are the new risks in the exploitation of available data sources, given current/emerging technologies, and someone with a hostile intent?
  - Public (e.g. web cams, air traffic data, weather stations, others)
  - Private – behind firewalls (security cameras in stores, airports, etc.)
- What are the new business opportunities for someone with a non-hostile intent?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

EXAMPLE SCENARIOS

- Non-hostile applications
  - "Big Brother" Auto Insurance
  - Coordinate resources during an "Amber Alert"
- Threats
  - Use public sensor data to plan / execute a terrorist attack, or coordinate a robbery
  - Use public sensor data to search for missing children – for human trafficking

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

PRESENTATION OBJECTIVES

From Initial Abstract

1. Summarize applicable technologies
2. Review statistics
3. Discuss firewall vulnerabilities
4. Explore applications enabled by access to data from a variety of sensors

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

#1: APPLICABLE TECHNOLOGIES

- Data mining, Feature extraction & Pattern recognition
  - ***(text)*** Content analysis, conflation
  - ***(imagery)*** Segmentation, template matching
  - ***(video)*** Object-based methods – HTM [Hawkins]
- Change detection
  - Image differencing, change vector analysis, correlation change image
- Artificial intelligence
  - Machine learning, genetic algorithms / evolutionary programming, neural networks, support vector machines
  - Low power, noise tolerant hardware (NeuroGrid [Boahen])

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

#1: APPLICABLE TECHNOLOGIES

- Data correlation & Data fusion
  - Normalized Cross Correlation (NCC), Fast Fourier Transform (FFT), Mutual Information [Cole-Rhodes]
- Data compression
  - Discrete Cosine Transform (DCT), Wavelets, MPEG4
- Real-time data processing
  - CloudShield “deep packet inspection”
  - IBM InfoSphere Streams

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

#2: STATISTICS

- Current number of web cams and other sensors:
  - Not easily obtained – unknown –
- Annual Growth Rate
  - Again unknown – exponential?
- Type
  - Streaming video
  - Still-frame web cams
  - Air traffic data
  - Weather stations
- Mode
  - Real-time - Really real-time? Or varying latency?
  - On line data archive

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

#3: FIREWALL VULNERABILITIES

Compromising Sensors on Private Networks

Many Methods

- Automated password attack
- Phishing
  - Kneber botnet
- Security lapses
  - Predator UAV hack via SkyGrabber software
- Bottom Line: We need to presume that any system may be vulnerable to compromise

MOUNTAIN STATE

INFORMATION SYSTEMS, INC.

#4: THREATS AND APPLICATIONS

The previous 3 objectives set the stage for the rest of the briefing

There are hostile threats and opportunistic applications enabled by access to data from a variety of sensors

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

THREATS / OPPORTUNISTIC APPLICATIONS

COMMON TASKS

- Evaluation (scoping)
  - Identifying the data sources
  - Accessing the data
  - Extracting information / Processing data
  - Acting on the information
- Countering potential threats / opportunistic applications
  - Known vs. unknown
  - Today's briefing purpose: Raising awareness, not proposing a full set of solutions

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

IDENTIFYING THE DATA SOURCE

- What type & where are the sensors?
  - "Where" = both IP address & Lat / Long
  - Summarized on web sites
  - Automated "spider"
- Is the data quality (e.g. resolution of a web cam) adequate for the application?
  - Perhaps not today, but constantly improving
- Is the data rate adequate?
  - Can the data support the application?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

ACCESSING THE DATA

- Publicly accessible vs. those behind firewall
- Connecting
  - Bandwidth & number of simultaneous connections
  - Storing and/or processing
  - Disk space, processing speed
  - Will our connection be detected? Covert connection required?
- Real-time threats vs. storing data for weeks (or mining archive data) and then acting on it
  - Real-time threat – no need to store data

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

EXTRACTING INFORMATION

- Automated vs. Manual
  - E.g., Coordinating a robbery or preplanning a terrorist attack, manual monitoring of web cams may be adequate
  - For facial recognition (e.g., supporting an “Amber Alert”), automation would be critical
- Automated Feature Extraction:
  - What features do we need?
  - What methods do we use?
  - How quickly do we need an answer?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

EXTRACTING INFORMATION

- Automated Feature Extraction
  - Object Detection
    - *Change detection, segmentation*
  - Preliminary Classification (? Of interest / Ignore ?)
    - *Template matching, clustering*
  - Tracking
    - *MPEG-4 object-based compression (I, B frame)*
  - Detailed Classification / Identification
    - Method depends on application & metrics

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

ACTING ON THE INFORMATION

Carry out the Attack

- Depends on the Application / Threat
- Timeliness of the Data
  - "Take into Custody" the target person
    - You must be prepared to act FAST
  - Some scenarios do not require "instantaneous" action
    - "Big Brother Auto Insurance"
    - Can do analysis at your leisure

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

COUNTERING THREATS / OPPORTUNISTIC APPLICATIONS

- Threats identified in advance
  - Prevent the threat before it is started
  - Stop the attack in real-time once it has started
- Identifying unknown threats
  - Can previously unknown threats - attempts to exploit data sources - be detected?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

EXAMPLE SCENARIOS

- Non-Hostile Applications / Opportunism
- Hostile Threats

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

HYPOTHETICAL OPPORTUNITIES (NON-HOSTILE)

- "Big Brother" Auto Insurance Company
  - Capture and store traffic camera data, then replay to identify fault after an accident
  - Map severe weather data to corroborate or invalidate crop damage claims
- Search for or Follow Individuals
  - Build profile of target based on photos, videos, other data
  - Software provides an alarm when individual sighted, or a report of where the individual has been – time/date stamped
  - Monitor a spouse, child, etc.

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

HYPOTHETICAL THREATS (HOSTILE)

- Terrorism, Theft, Personal Vendetta
  - Traffic web cams or "site of interest" video for planning or coordinating an attack/theft
  - Monitoring known targets (ref: www.PleaseRobMe.com)
- Identity Theft / Targeted Phishing:
  - Monitor video in Dept. stores for key shoppers (e.g. Mx. Bigspender)
  - Targeted phishing the next day "your purchase at Bloomingdales was rejected… Enter account info…" → Capture credit card #
  - *Ignores question of "how do they know how to match her face and email address?"*
- How would a terrorist exploit live video feeds or real-time airport traffic data from a major US city?
  - *Let's discuss privately*

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

BIG BROTHER AUTO INSURANCE

The Application:

- An auto insurance company records all available traffic video for 14 days
- Upon notification of any traffic accident, the historic video is reviewed to locate the participants
  - Can fault be proven?
  - Evidence of reckless / drunk driving or road rage?
  - Other scenarios (drivers switch after accident)?
  - Automated methods? Minimal…

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

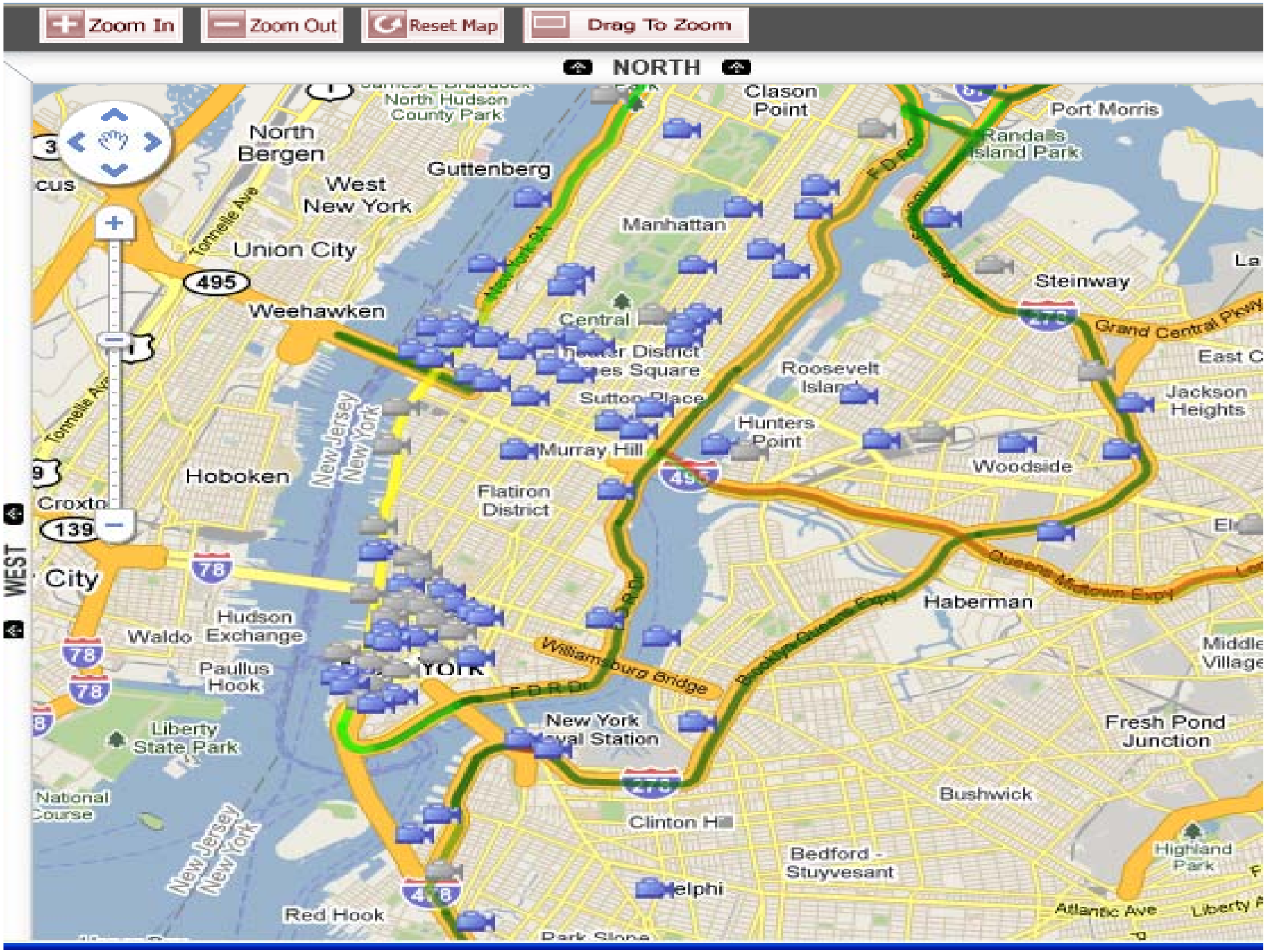
BIG BROTHER AUTO INSURANCE

Scope

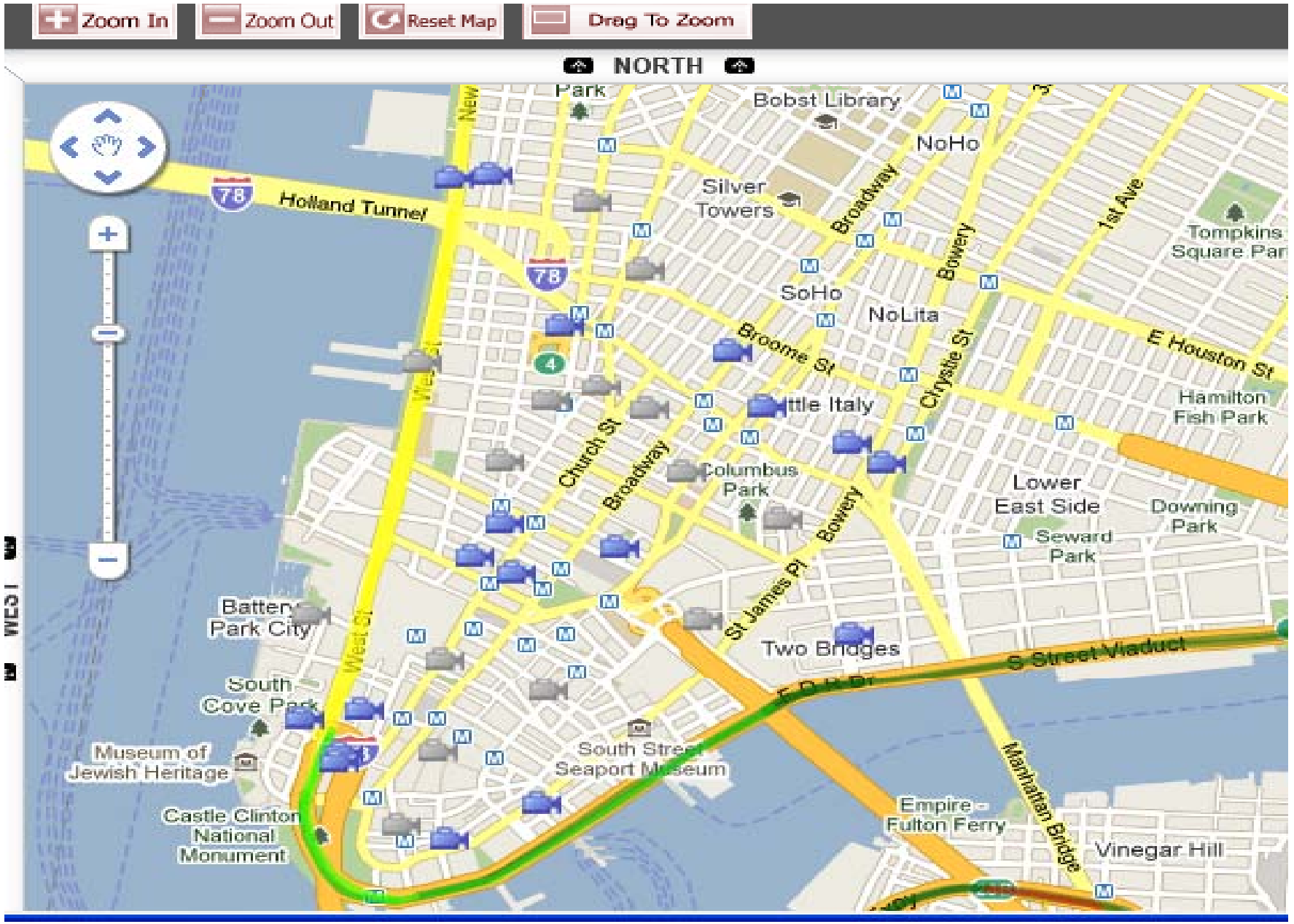
- 85 traffic cameras in Manhattan
- Bandwidth required = 256 kb/sec/camera
  - Disk storage for 14 days = 3.6 TB
- Assumes full motion (> 5 frames/second)
  - Ambiguity in exactly what happened?
- Size of car in image = 15 - 150 pixels
  - Positive ID possible?
- Traffic lights, pedestrians, other pertinent details visible?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

Data Sources – NYC web cams



Data Sources – NYC web cams



BIG BROTHER AUTO INSURANCE

- Data Analysis
  - Map web cams to Lat/Long
  - Locate proper time sequence in video stream, at accident site & prior inbound traffic points
  - Automated methods for identifying target of interest, accidents, risky behavior?
- Other Issues
  - Web cam coverage and resolution (HD) increasing
  - Legal to use as evidence?
  - Can insurance companies already get this data from the government without having to set up their own data center?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

BIG BROTHER AUTO INSURANCE

Video Examples

- MD-DOT – 128 kbps

  http://www.chart.state.md.us/TravInfo/trafficCams.asp

- MD-DOT – 4 cameras

  http://www.chart.state.md.us/video/video.asp?feed=n

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

SEARCH FOR INDIVIDUALS

- Applications
  - Casino security – big spenders, suspicious people
  - Amber Alert – find missing children
  - Human Trafficking – find missing children
  - Search for Terrorists / Escaped Prisoners
- Can individuals be identified in public web cams?
  - Facial identification metrics
    - Interpupillary distance, etc.
    - Do web cams provide adequate detail?
    - If not today, is it conceivable in the near future?
  - Full body biometrics
    - Body dimensions
    - Gait, other dynamic metrics

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

SEARCH FOR INDIVIDUALS

Scope

- Difficulty
  - "*Amber Alert! Find this face anywhere in Manhattan*." VERY HARD
  - "*Follow my husband, ensure he takes the Jerome Avenue line, exits at 176th, goes to Murphy's for lunch*." A LITTLE EASIER
- X number of cameras in Manhattan
  - Restaurants, Department Stores, others on known routes?
- Bandwidth "required" = 256 kb/sec/camera
- Full body biometrics "requires" continuous motion (> 10 - 30 frames/second)
- Height of person in image = 15 - 150 pixels

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

SEARCH FOR INDIVIDUALS

Data Analysis

- *Automated methods for identifying targets of interest*
- Build geospatial map of web cam locations (if necessary)
- Locate proper time sequence in video stream (for possible later detailed analysis)
- Not completely realistic today… Too few data sources, but web cam coverage and resolution (HD) increasing

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

SEARCH FOR INDIVIDUALS

Other Issues

- Non-public data sources
- Police could build a private network of cameras on every road in/out of a city
  - Monitor for people based on stolen vehicle ID and/or facial recognition
  - Expand to capture every street, sidewalk, doorway, etc. in a city?
- Cameras to monitor every entrance to a school?
- Personal Freedom? Is this George Orwell's 1984?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

SEARCH FOR INDIVIDUALS

Video Examples

- Times Square Video

  http://www.earthcam.com/usa/newyork/timessquare/

- Times Square HD Video

  http://www.earthcam.com/usa/newyork/timessquare/?cam=lennon_hd

- Video tools from Vitamin D
  - Follow objects in motion
  - Trigger on objects / people crossing thresholds

Think about Full Body Biometrics

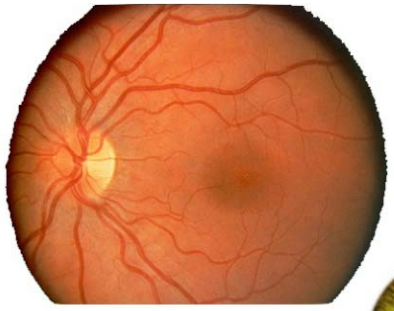
MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

FULL BODY BIOMETRICS

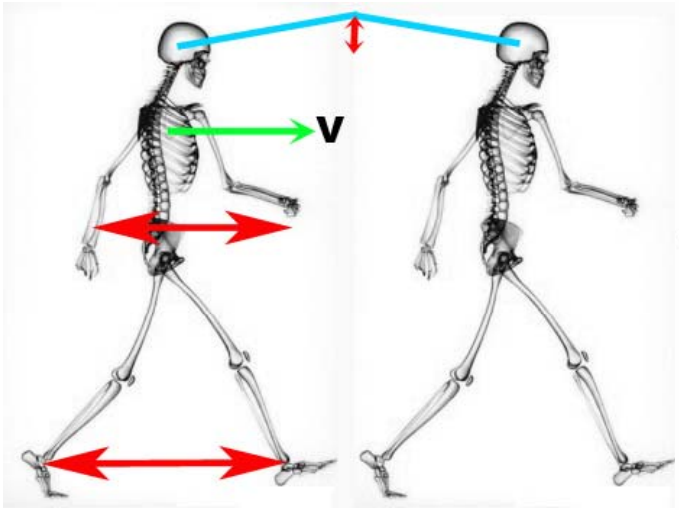
- Objective: Unique individual identification based on multiple views and dynamic analysis extracted from video subsets, using a multi-resolution data model compiled over time
- Applicable to other animals and objects (automobiles etc.)
- Full individual recognition based on a composite of numerous biometrics, all of which may be partially complete

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

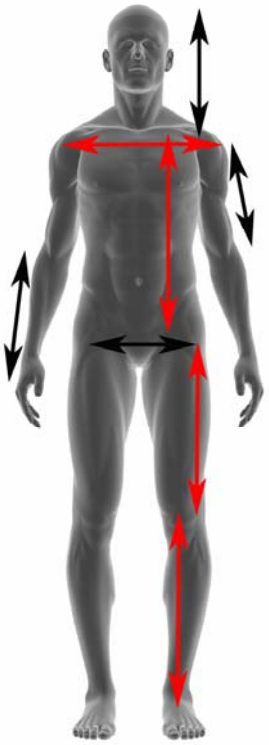
EXAMPLE METRICS



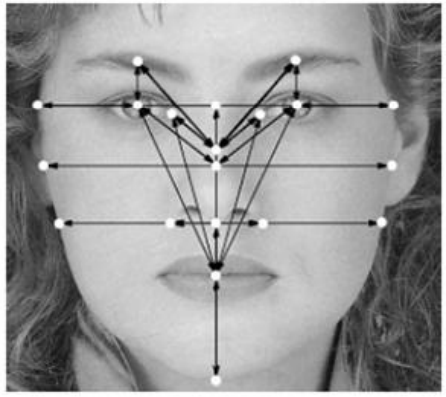
MICRO



MACRO



MID



SUPER MACRO

MOUNTAIN STATE

INFORMATION SYSTEMS, INC.

DATA MODEL

- Four spatial scales
  - Micro, mid, macro, and supermacro
  - Network (Tree) data structure
  - Wavelets / Fractals for data compression
- Dynamic motion analysis integrated at macro/supermacro scales
  - Eigenvectors / Eigenvalues used to extract unique features at each scale
  - Populate predefined data models for high value targets (people on a watch list)

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

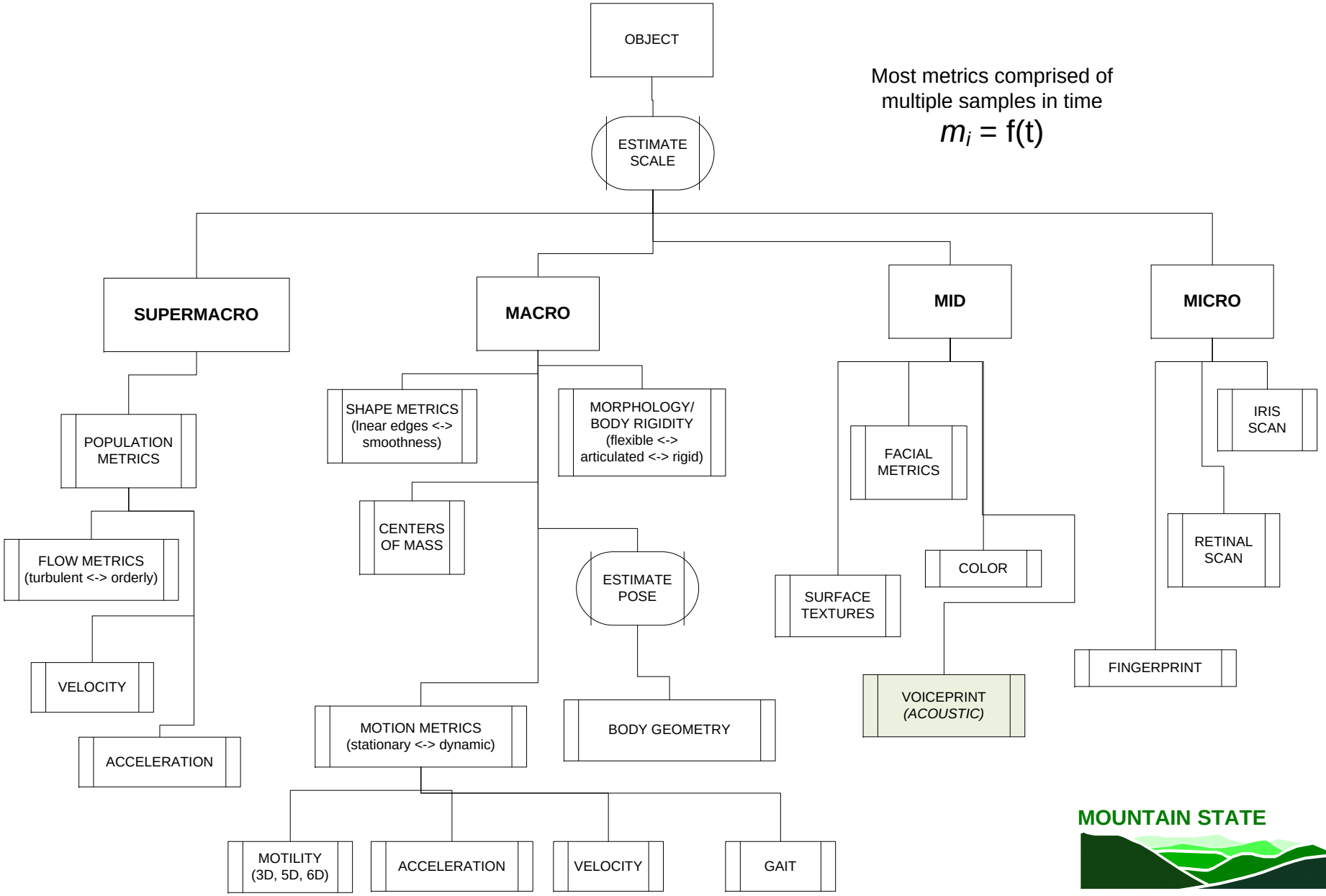
METRICS AS A FUNCTION OF SCALE

Scale \ METRIC	Texture	Specular reflection	Corners	Edges	Color	Speed	Acceleration	Pose	Gait	Motility (3D, 5D, 6D)	Rigidity	Centers of mass	Traffic / Flow	Population
MICRO	*	*	*	*	*									
MID		*	*	*	*						*	*		
MACRO		*	*	*	*	*	*	*	*	*	*	*		
SUPERMACRO						*							*	*

This is in addition to classic biometrics, e.g. facial metrics, iris/retina scan, etc.

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

DATA MODEL



MOUNTAIN STATE INFORMATION SYSTEMS, INC.

SUMMARY

- A Glimpse into Our Vision of the Future wrt Data Mining of Video
- Scenarios: Opportunities and Threats
  - Big Brother Insurance Company
  - Search for Individuals
- Common Tasks
  - ID Data Sources
  - Access the Data
  - Extract / Process the Data
  - Act on the Information
- A lot possible today. More coming.
- Will there eventually be so many video cams that almost everything is captured? Available? Ripe for analysis?
- Will Full Body Biometrics provide an important augmentation to other methods, especially in scenarios that require covert monitoring?

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

TAKEAWAY MESSAGES

- Technology is available – or nearly ready – to act on these threats/applications
- Data quality and data coverage are lacking, but gaining fast
- Technology cost is high, but dropping
- We in the tech community need to be careful about what we offer to the world, and think about risks
- Full body biometrics offers great promise
- Continuing studies need to be performed
- Continue research to serve the needs of the defense / intel communities

Paul Garnett
970 708-7723
pgarnett@msisinc.com

Cody Benkelman
406 270-1176
cody@MissionMtnTech.com

Fred Palma
916 501-6051
fpalma@msisinc.com

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

REFERENCES

- "*U.S. Warned of Predator Drone Hacking*", CBS news, 12/17/2009 http://www.cbsnews.com/8301-504383_162-5988978-504383.html?tag=mncol%3btxt
- "*Kneber botnet catches 2,500 companies worldwide*" , the Guardian, 18 February 2010. http://www.guardian.co.uk/technology/2010/feb/18/kneber-botnet-netwitness-cybercrime
- K Boahen, **Neuromorphic Microchips**, *Scientific American*, vol 292, no 5, pp 56-63, May 2005.
- E Culurciello, R Etienne-Cummings, and K Boahen, **A Biomorphic Digital Image Sensor**, *IEEE Journal of Solid State Circuits*, vol 38, no 2, pp 281-294, 2003.
- "Hierarchical Temporal Memory: Concepts, Theory, and Terminology", Jeff Hawkins and Dileep George, Numenta Inc., http://www.numenta.com/Numenta_HTM_Concepts.pdf
- M. Turk and A. Pentland. **Eigenfaces for Recognition**. *Journal of Cognitive Neuroscience*, 3(1), 1991
- Cole-Rhodes, Arlene, Kisha Johnson, Jacqueline Le Moigne, "Multiresolution registration of remote sensing images using stochastic gradient," Proc. SPIE Vol. 4738, p. 44-55, Wavelet and Independent Component Analysis Applications IX; Harold H. Szu, James R. Buss; Eds. Mar 2002
- "Casino insider tells (almost) all about security", ComputerWorld, 10 March, 2008
- IBM **InfoSphere Streams:** http://www-304.ibm.com/jct01003c/software/data/infosphere/streams/
- CloudShield: http://www.cloudshield.com/

MOUNTAIN STATE

INFORMATION SYSTEMS, INC.

DATA SOURCES

- Object tracking video
  - Numenta
    - http://www.numenta.com
  - Vitamin D
    - http://www.vitamind.com/demo.php
  - General Electric
    - http://ge.geglobalresearch.com/technologies/imaging/

- Traffic
  - NYC:
    - http://www.earthcam.com/usa/newyork/timessquare/
    - HD: http://www.earthcam.com/usa/newyork/timessquare/?cam=lennon_hd
  - NYC: http://nyctmc.org/
  - California: http://www.video.dot.ca.gov/

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

DATA SOURCES

- Air Traffic
  - http://www.passur.com/airportmonitor-locations.htm
  - NY
    - http://www4.passur.com/jfk.html
    - http://www4.passur.com/lga.html
  - CA
    - http://www4.passur.com/sna.html
    - http://www4.passur.com/bur.html
    - http://www4.passur.com/san.html

- Weather
  - www.anythingweather.com/
  - wxqa.com/stations.html
  - www.space.com/php/weatherbug/
  - www.met.utah.edu/mesowest/
  - Summary at http://www.dslreports.com/faq/13342

MOUNTAIN STATE
INFORMATION SYSTEMS, INC.

