Dynamic Sensor Fusion

Bir Bhanu

bhanu@cris.ucr.edu

Center for Research in Intelligent Systems

University of California, Riverside, CA 92521, U.S.A.

November 17, 2005

Research in Intelligent Systems

- Computer Vision and Pattern Recognition
  - Object recognition and image segmentation
  - Performance theory and modeling
  - Dynamic scene and motion analysis
  - Physics-based models and sensor fusion
  - Sensor networks
  - 3-D model acquisition and refinement from video
  - VLSI implementations
  - Applications with a variety of sensors

Research in Intelligent Systems

- **Machine Learning and Data Mining**
  - Synthesizing recognition system architectures, features, strategies
  - Closed-loop object recognition
  - Learning segmentation, feature extraction, concepts, recognition strategies
  - Behavior learning
  - Multi-strategy learning techniques
  - Intent recognition
  - Complex Pattern Learning

Research in Intelligent Systems

- Image and Video Databases
  - Relevance feedback for feature selection, online indexing, content-based image retrieval
  - Exploitation of meta knowledge for online indexing, perceptual partitioning of databases
  - Semantic concept learning from video
  - Personalized information synthesis from multi-media data
  - Uncertainty handling in dynamic databases

Research in Intelligent Systems

- **Biometrics**
  - **Multi-modal human ID**
    - Face (3D) and face profile recognition
    - Ear recognition in 3-D
    - Gait recognition
    - Fingerprint recognition
    - Performance prediction and modeling
    - Multi-modal biometrics
    - Integrated recognition at various levels of access

Patterns for Network Security

Network Intrusion

- Intrusion detection
- Pattern matching
- Monitoring
- Online corrective actions

Research in Intelligent Systems

• Sensor Networks

- A wireless networked environment with several hundred PTZ color and infrared cameras
- Real-time motion detection at a node
- Dynamic multi-objective optimization for tradeoffs between processing, actuation and communication
- Environmental invariance for distributed detection and recognition at a distance
- Learning to recognize usual and unusual human activities and individuals
- Registration of heterogeneous data and sampling
- Dynamic sensor (imaging and non-imaging) fusion
- 3-D model building of objects
- Active vision in a networked environment
- Architecture of the system

Outline of the Presentation

- **Moving Object Detection**

- **Introduction**
  - **Background Work**
  - **Statistical Background Model Representation**

- **Evolutionary Sensor Fusion**
  - **System Architecture**
  - **Physical Models (Visible, Thermal)**

- **Technical Approach**

- **Experimental Results**

- **Conclusions**

Objective and Contributions

Objective: Develop new fusion techniques based on sound physical, statistical and evolutionary models to detect moving objects in outdoor 24/7 under variety of illumination and environmental conditions.

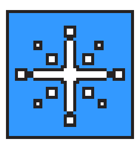
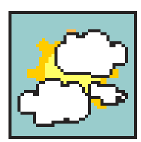
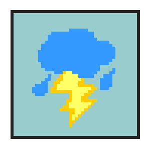
Contributions:

1. It integrates thermal and reflectance physical models into a uniform approach for sensor fusion.
2. It develops a novel sensor fusion technique based on cooperative coevolutionary computational model, which incorporates physical and environmental conditions into a new evolutionary dynamic sensor fusion model.
3. It provides analysis and results of moving object detection using color and IR video for a full diurnal cycle.

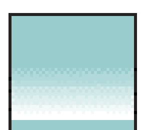
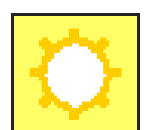
Moving Object Detection in Video

Many problems make the detection difficult:

- ***Complexity of the scene***
  - simple uniform vs. highly textured background
  - moving background (e.g., swaying trees)
- ***Lighting conditions***
  - moon light, sun light, fluorescence, colored light
- ***Weather phenomenon***
  - rain, snow, sudden cloud cover,

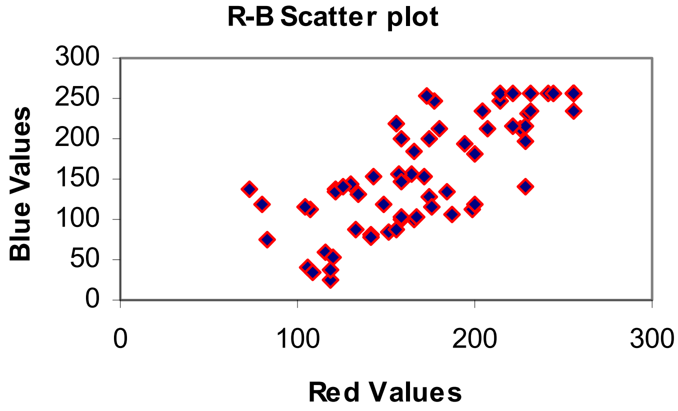
- ***Camera noise and sensitivity***
- ***Time of day***
  - Sunset, sunrise

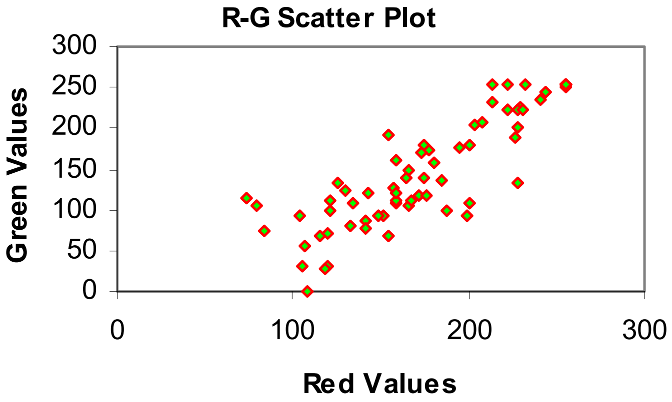



Moving Object Detection in Video

Example of a moving background:

A pixel





Moving Object Detection in Video

Statistical-based approach

Each pixel is viewed as an independent process and its history is tracked over time by an independent statistical method.

Examples:

PFINDER, [Pentland, et. al 1997]. designed for indoor, controlled lighting illuminations and uses one Gaussian pdf to model the background.

VSAM [Stauffer, and Grimson, 1998], for tracking people and cars in outdoor environment. Uses mixture of Gaussians pdf to model background, requires significant statistics and sufficient illumination.

Multistrategy Fusion [Nadimi, and Bhanu, 2001], applies multistrategy fusion rules including OR, AND and Dempster Shafer method to the mixture of Gaussian model. Performs robust fusion, indoor/outdoor, requires sufficient illumination.

Multisensor Fusion

Multisensor fusion attempts to combine the information from all sensors into a unified representation.

Why Sensor Fusion?

Some of the advantages to multisensor fusion are improved detection, increased accuracy, reduced ambiguity, robust operation, extended coverage.

Fusion Methods:

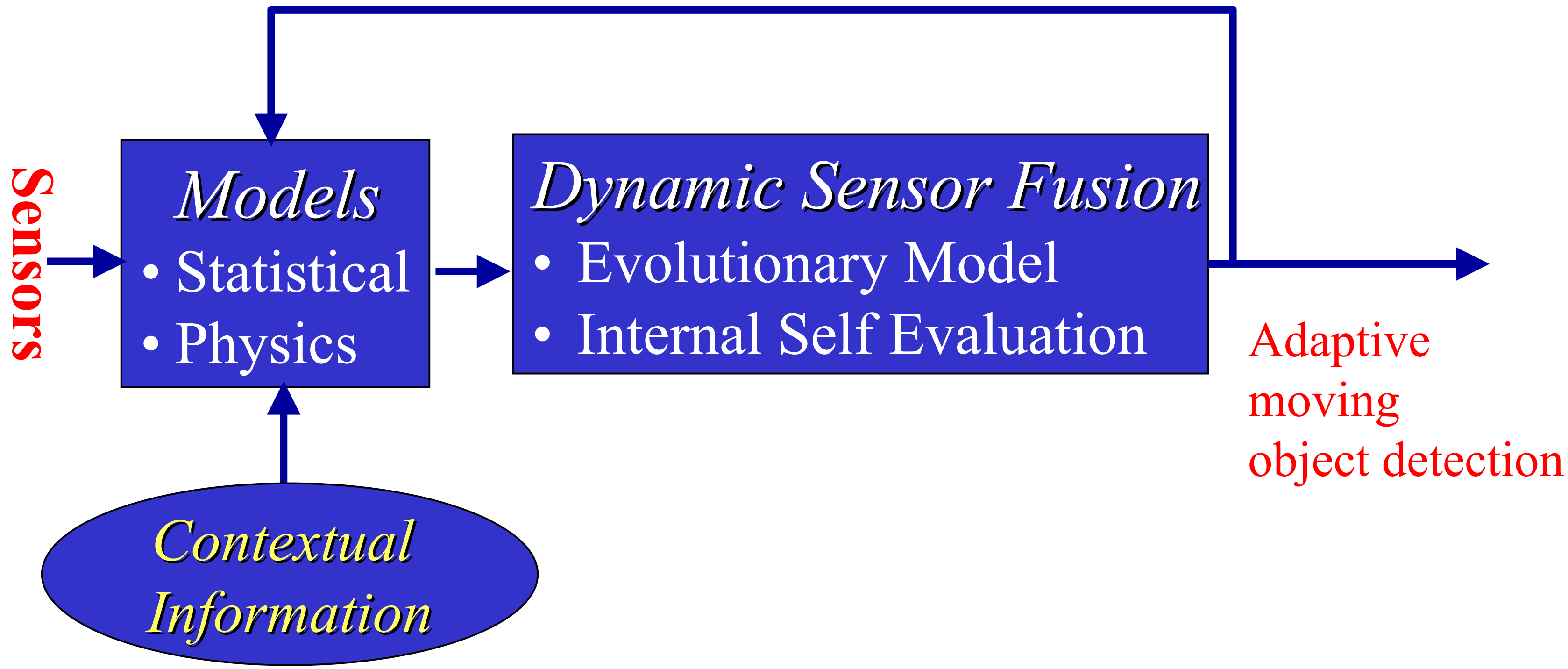
Statistical based: ***Bayesian*** – ***Dempster-Shafer*** – ***Fuzzy***

AI: ***Knowledge-based, Rule-based, Information Theoretic***

Algorithmic: ***Graphs, Trees, Tables, Hough transform***

Physics-based: Physics-based approaches utilize the sensor phenomenology and are based on sound physical models that can interpret the signal.

Adaptive Sensor Fusion Model



Model Representation

Recent history of each pixel $\{X_1, \ldots X_t\}$ is modeled by a mixture of K Gaussians :

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} \times \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

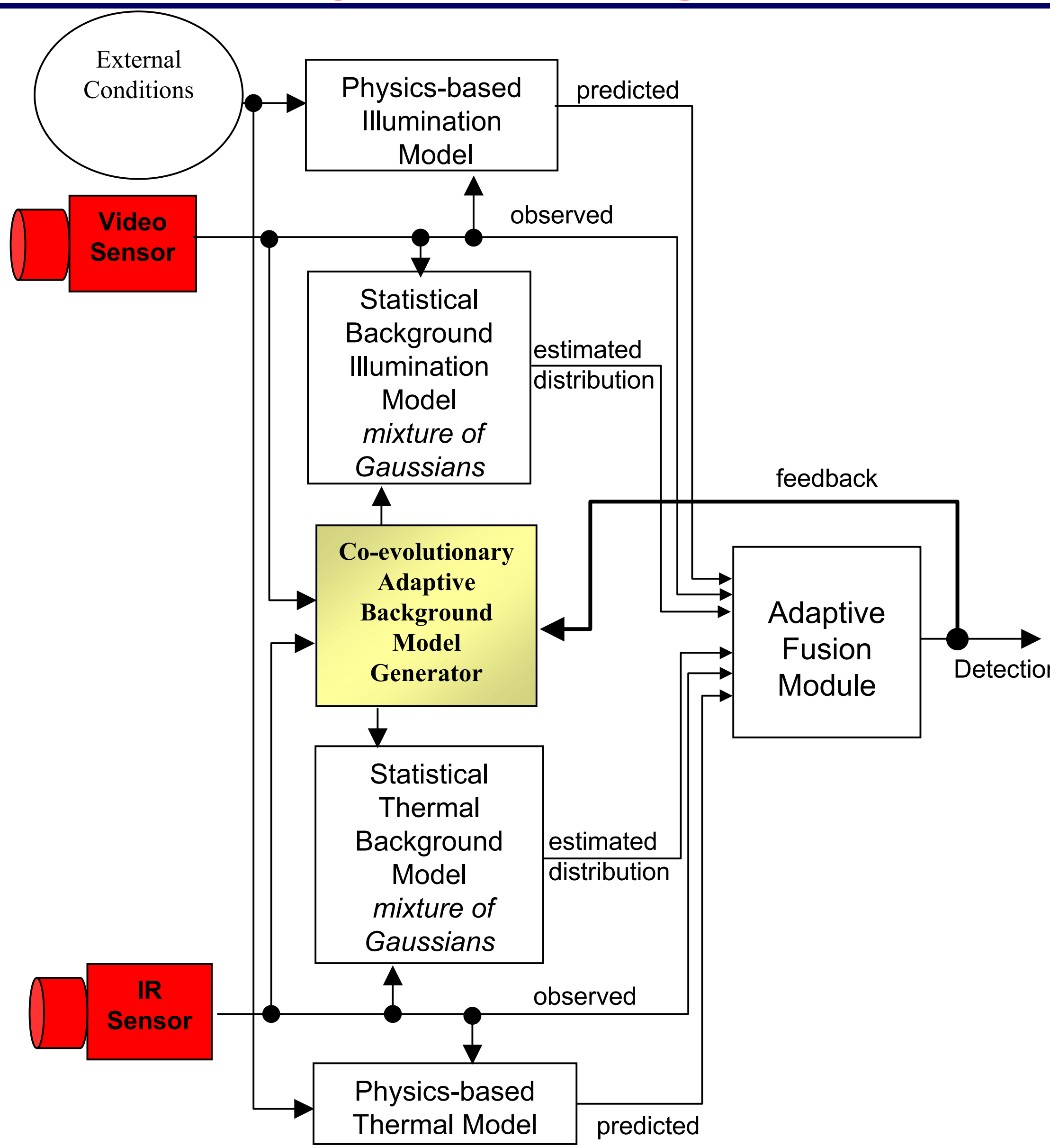
$$\eta(X_t, \mu, \Sigma) = \frac{1}{2\pi^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

We extend the idea of mixture model in the color to both color and infrared (IR). A pixel is then represented by a mixture of Gaussians for 4 channels.

$$\begin{aligned}
<& w_{R1}, \mu_{R1}, \sigma_{R1}, \ldots, w_{Rk}, \mu_{Rk}, \sigma_{Rk}, \\
& w_{G1}, \mu_{G1}, \sigma_{G1}, \ldots, w_{Gk}, \mu_{Gk}, \sigma_{Gk}, \\
& w_{B1}, \mu_{B1}, \sigma_{B1}, \ldots, w_{Bk}, \mu_{Bk}, \sigma_{Bk}, \\
& w_{T1}, \mu_{T1}, \sigma_{T1}, \ldots, w_{Tk}, \mu_{Tk}, \sigma_{Tk}>
\end{aligned}$$

W = Prior (weight), μ = Mean, σ = Standard Deviation, R,G,B = Red, Green or Blue channels, T = Thermal channel, k = number of Gaussians

Evolutionary Sensor Fusion System Diagram



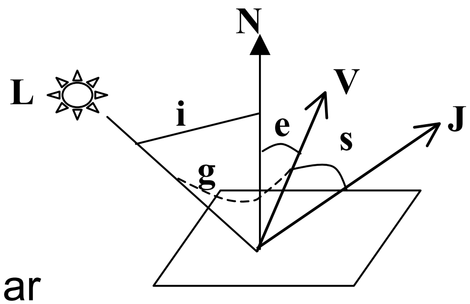
Predictive Physical Models (Illumination)

Lambertian $I = K \; L \; \cos(i)$

Phong $I = K \; L \; \cos(i) + \omega \; \cos^n(s)$

I = Intensity, **K** = Coefficient of reflection

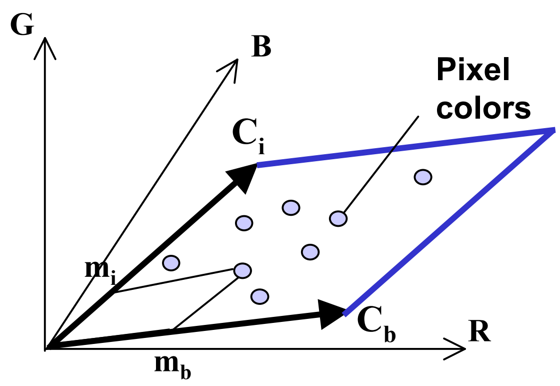
Directions: L = Illumination, N = Normal, V = Viewing, J = Specular
Angles: (s, e, g, i)



Dichromatic

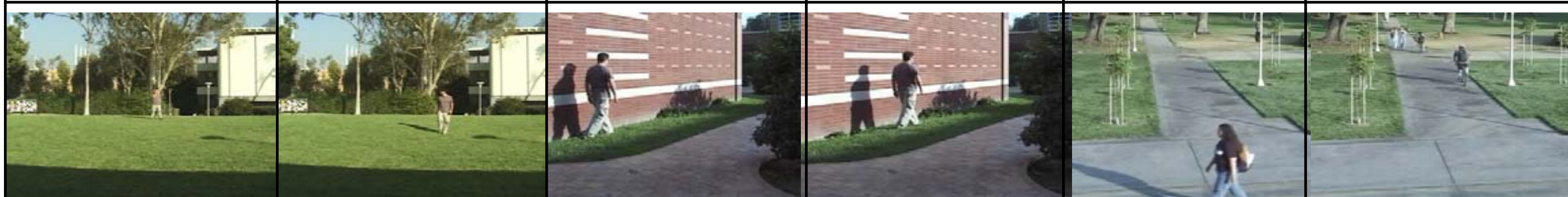
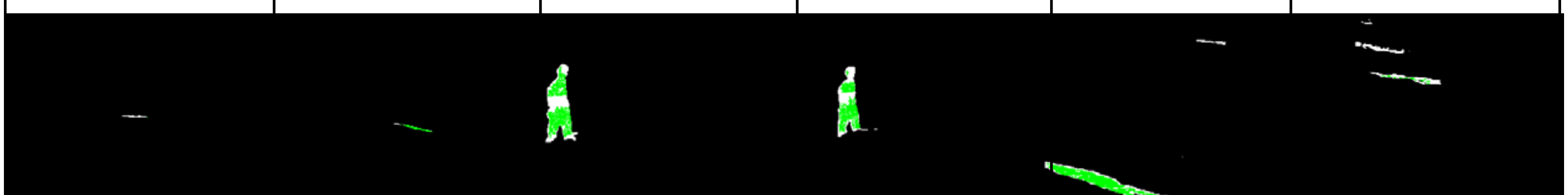
$$L(\lambda,I,e,g) = L_i(\lambda,I,e,g) + L_b(\lambda,I,e,g) = m_i(I,e,g)\, c_i(\lambda) + m_b(I,e,g)\, c_b(\lambda)$$

L = total reflected radiance,
L_i = reflected radiance at the surface,
L_b = reflected radiance from the body (subsurface),
m_i and m_b = geometric terms for the surface and body,
C_i, and C_b = relative spectral power distribution due to surface, and body reflections.



We use c_b as the surface color invariant
[Nadimi and Bhanu, MFI 2003, PAMI Aug. 2004]

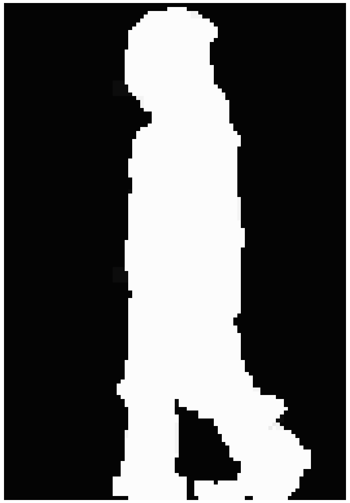
Moving Object and Shadow Detection (IEEE TPAMI 2004)

Example: *Date*: 11-01-2001; *Time*: 3:30 p.m.; *Sun Direction*: Right to left; *Surface Orientation*: Inclined, Curved; *Surface Type*: Grass.		**Example:** *Date*: 11-01-2001; *Time*: 4:30 p.m.; *Sun Direction*: Right to left; *Surface Orientation*: Flat, Horizontal, Vertical; *Surface Type*: Grass, Red Tile.		**Example:** *Date:* 10-26-2001; *Time*: 4:30 p.m.; *Sun Direction*: Right to left; *Surface Orientation*: Flat, Horizontal; *Surface Type*: Grass, Textured Concrete.		
Frame # 1076	1252	312	360	765	4300	
						**Input Video**
						% of shadow detected (green)
						% of object detected (red)

GAIT - DARPA HumanID Database

An example of color image sequence

An example of extracted binary silhouette sequence



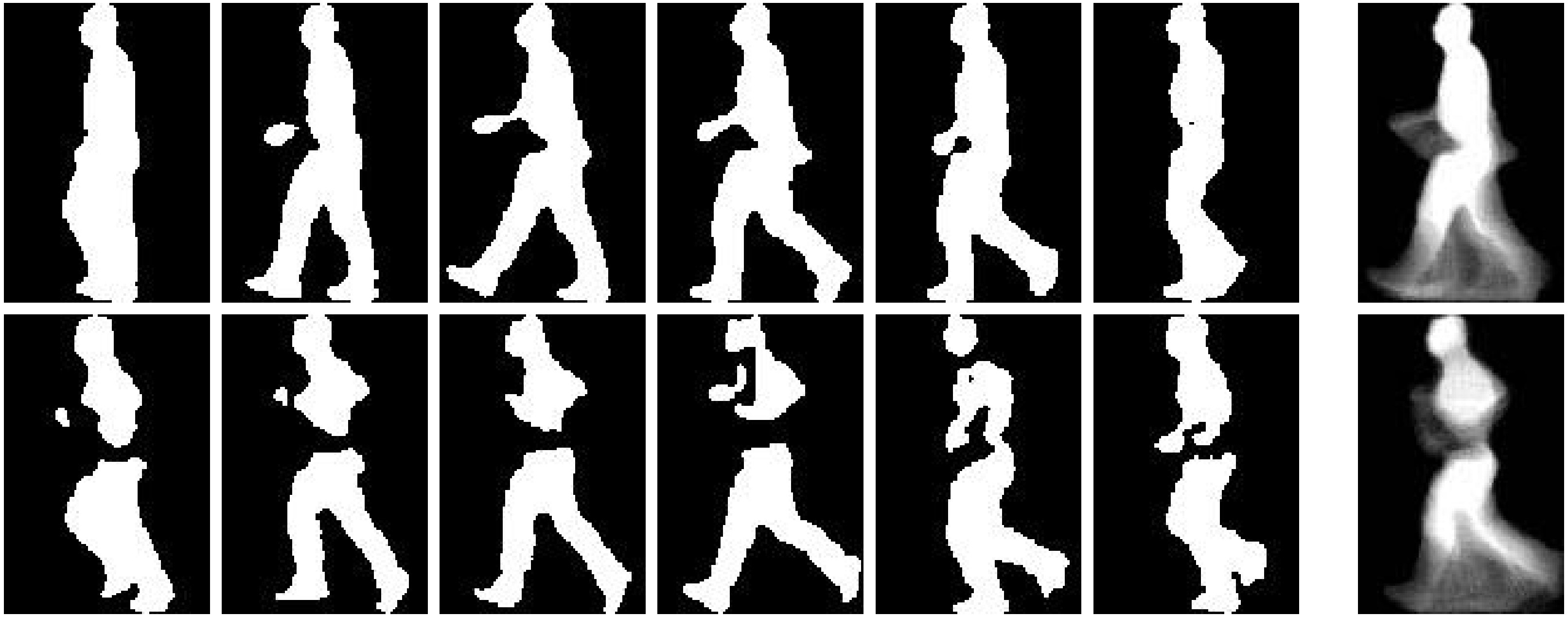
- Database contains 1870 sequences from 122 subjects
- Database and the extracted silhouette sequences are developed by University of South Florida http://marathon.csee.usf.edu/GaitBaseline/

Gait Energy Representation (IEEE TPAMI, in Press)

Normalized binary silhouette sequence GEI

Slow Walking

Run



- **Representation construction**
  – Silhouette extraction, normalization, alignment and averaging
- **Representation properties**
  – Represents human motion sequence in a single image while preserving some temporal information
  – Saves both storage space and computation time for recognition
  – Less sensitive to silhouette noise in individual frames

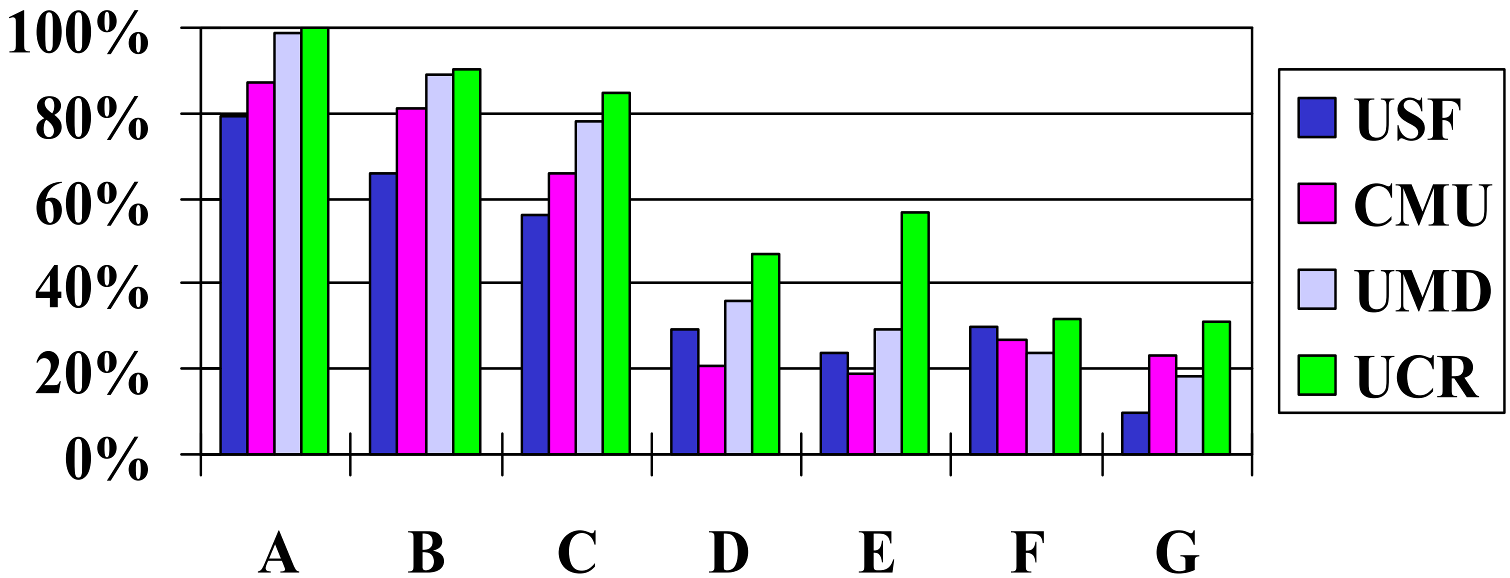
DARPA's HumanID Database

Version 1.7

Label of Dataset	Size of Dataset	Data Recoding Conditions
Gallery	71	Grass (Surface), A (Shoe), Left (camera view)
Probe A	71	Grass, A, Right
Probe B	41	Grass, B, Left
Probe C	41	Grass, B, Right
Probe D	70	Concrete, A, Left
Probe E	44	Concrete, B, Left
Probe F	70	Concrete, A, Right
Probe G	44	Concrete, B, Right

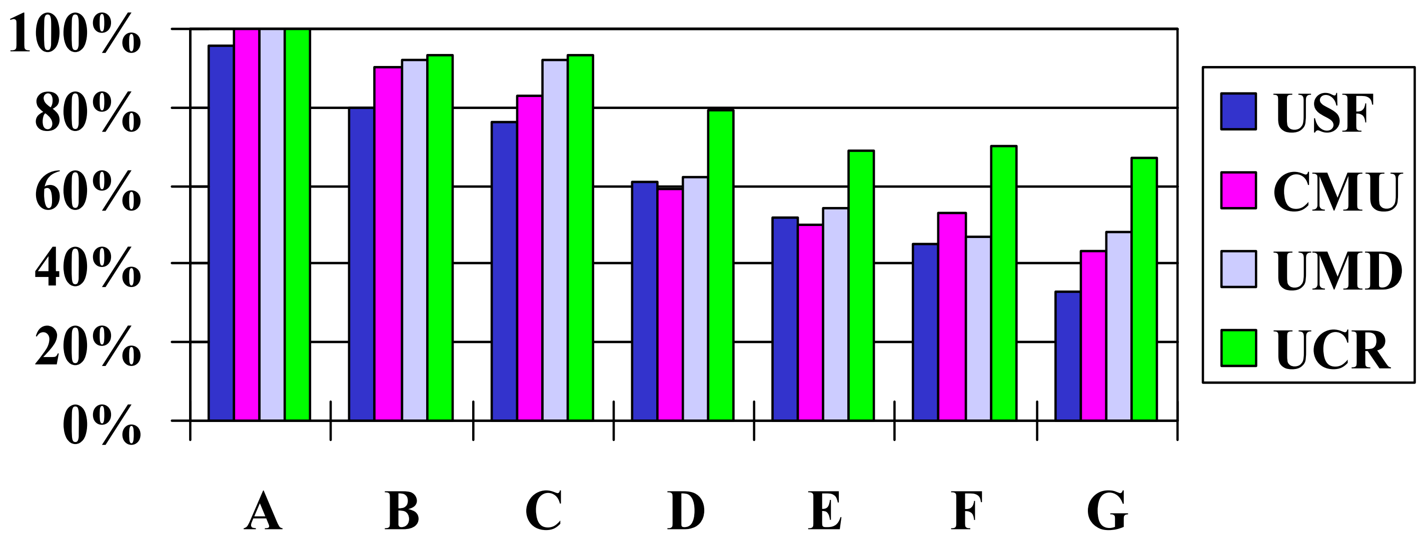
- Evaluate the performance of gait recognition approaches
- Evaluate the effect of environmental condition changes

Rank1 Performance



- **Legends**
  - Rank1: only the first candidate in the rank list is considered
  - USF: direct frame shape matching approach (USF website)
  - CMU: key frame shape matching approach (AVBPA 03)
  - UMD: HMM approach (ICIP 03)
  - UCR: our approach (PAMI in Press)
- **Observations**
  - Our approach is better than other approaches on all experiments
  - The surface change in A-G results in dramatic performance drop (introducing more distortion in the silhouette)

Rank5 Performance

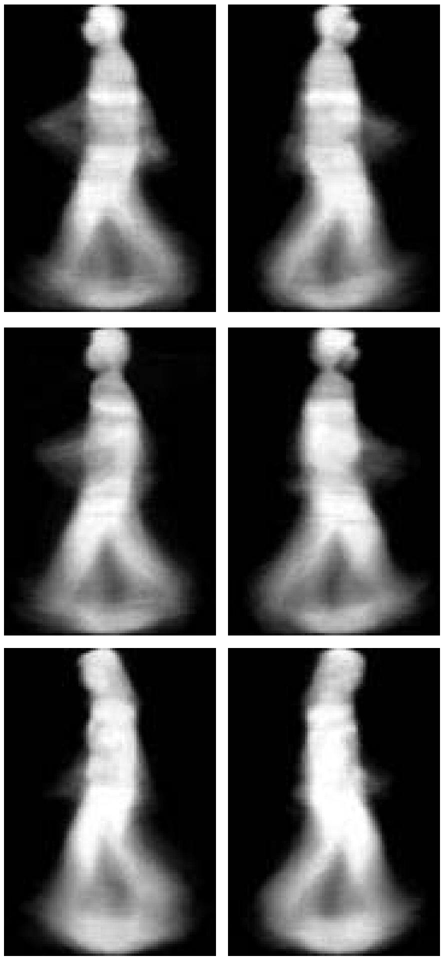


- **Legends**
  - Rank5 – all the first five candidates in the rank list are considered
  - USF – direct frame shape matching approach (USF website)
  - CMU – key frame shape matching approach (AVBPA 03)
  - UMD – HMM approach (ICIP 03)
  - UCR – our approach (PAMI in Press)
- **Observations**
  - Our approach is better than other approaches in all experiments (silhouette distortion has been considered in synthetic training data)

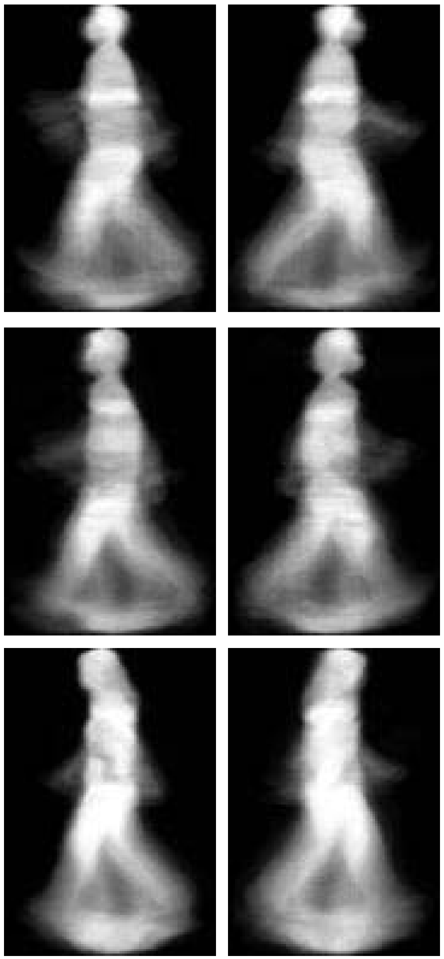
Activity recognition in IR video (IEEE CVPR OTCBVS 2005)

Walking (noon)

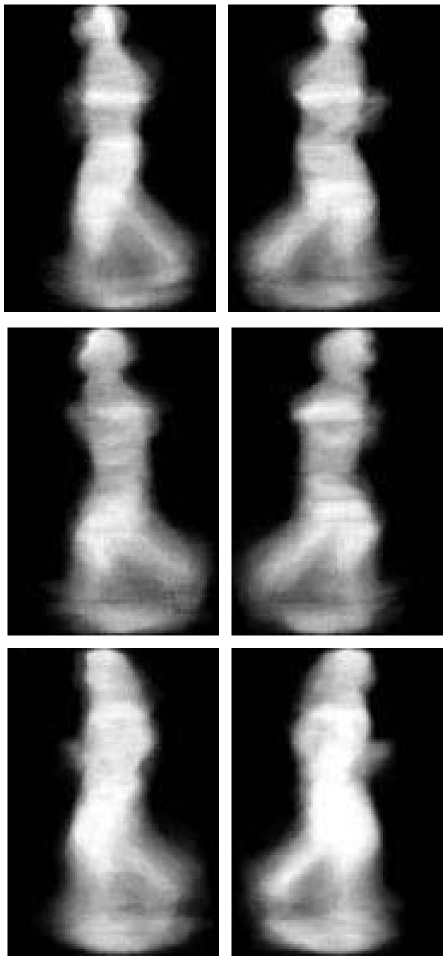
Slow Walking



Fast Walking



Run (noon)



Experiment I (IR): Testing Data II

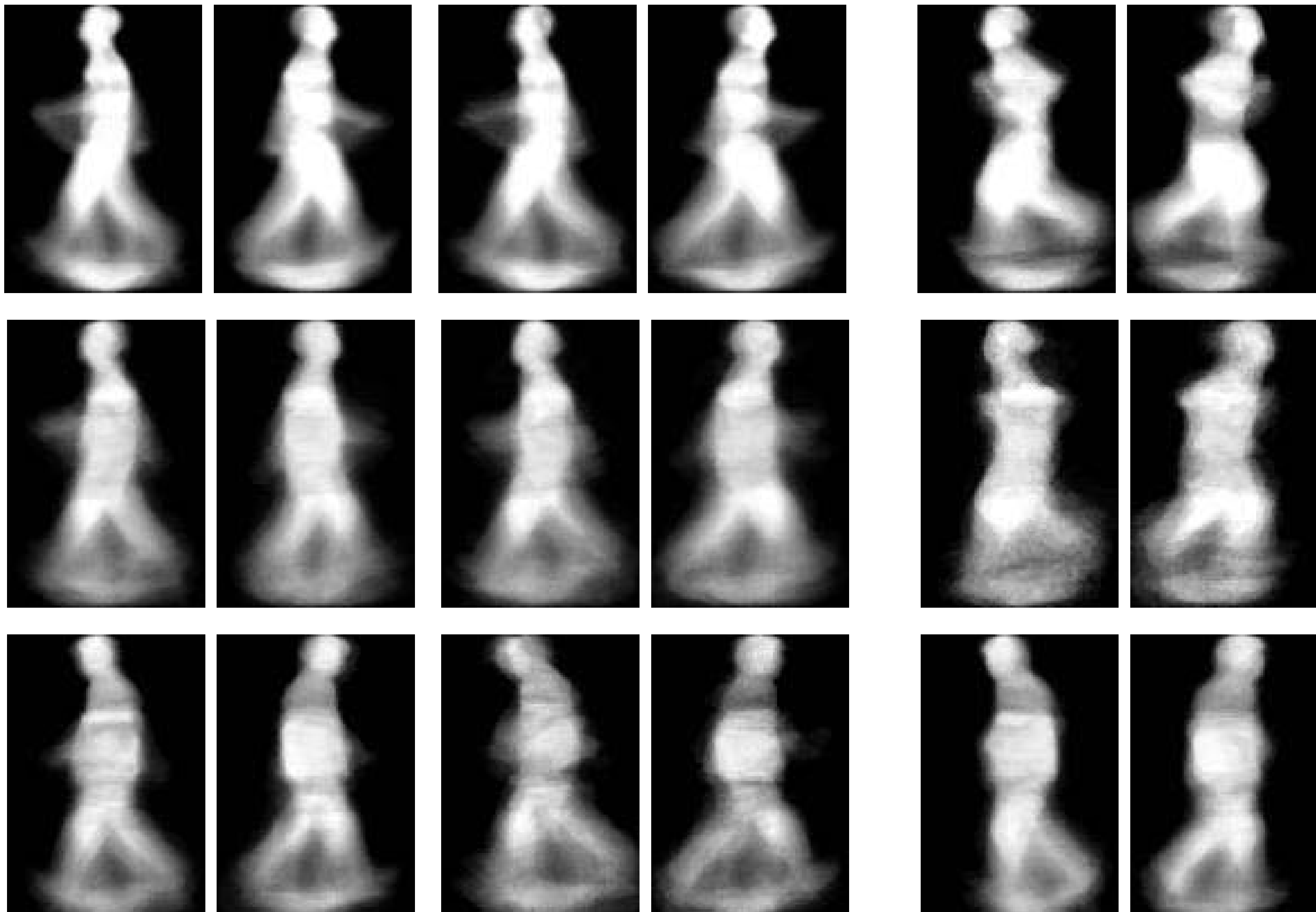
Walking | Run

Slow Walking | Fast Walking

Late Afternoon

Night

Night



Center for Research in Intelligent Systems

Experiment I (IR): Testing Data I

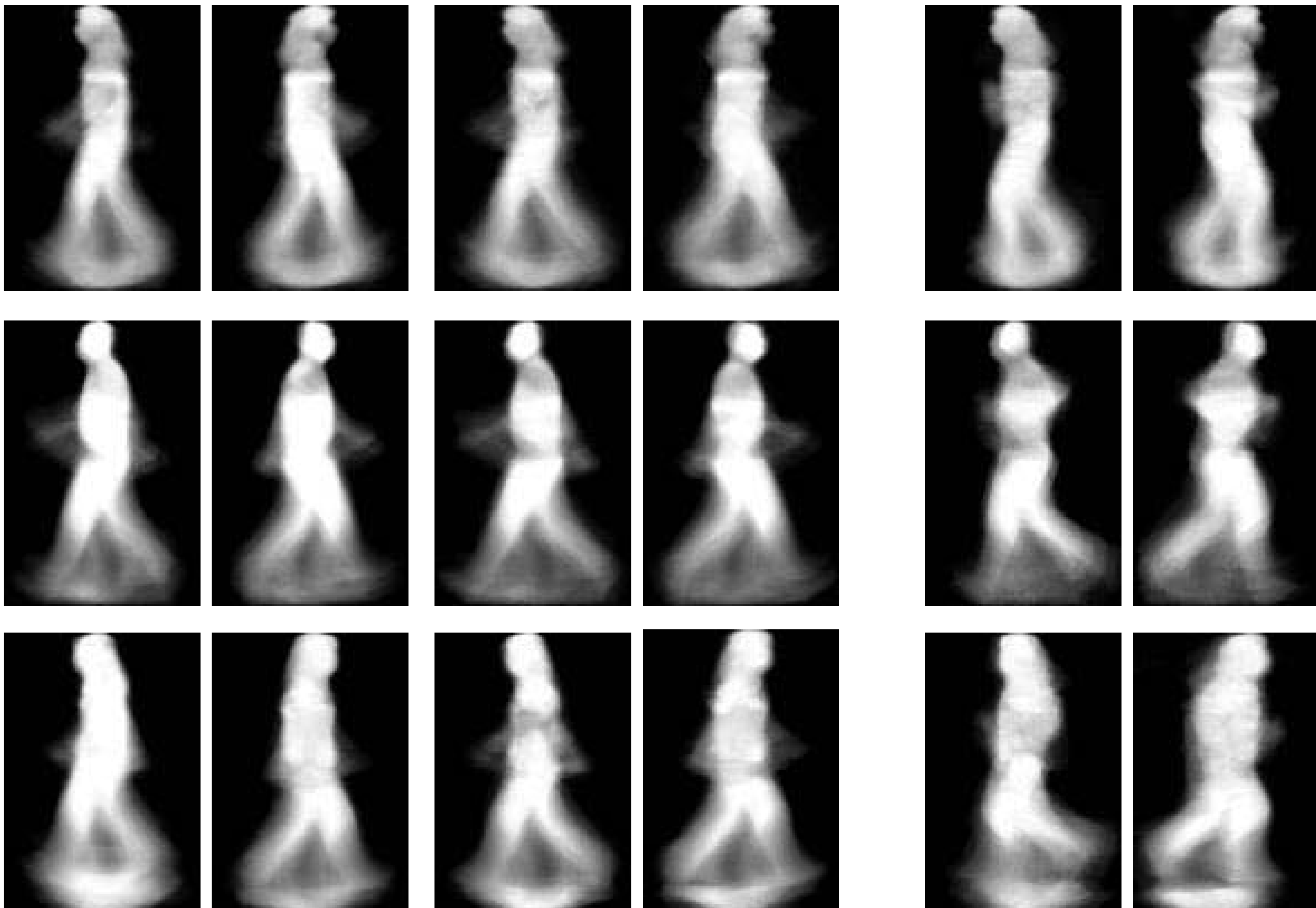
Walking | **Run**

Slow Walking | **Fast Walking**

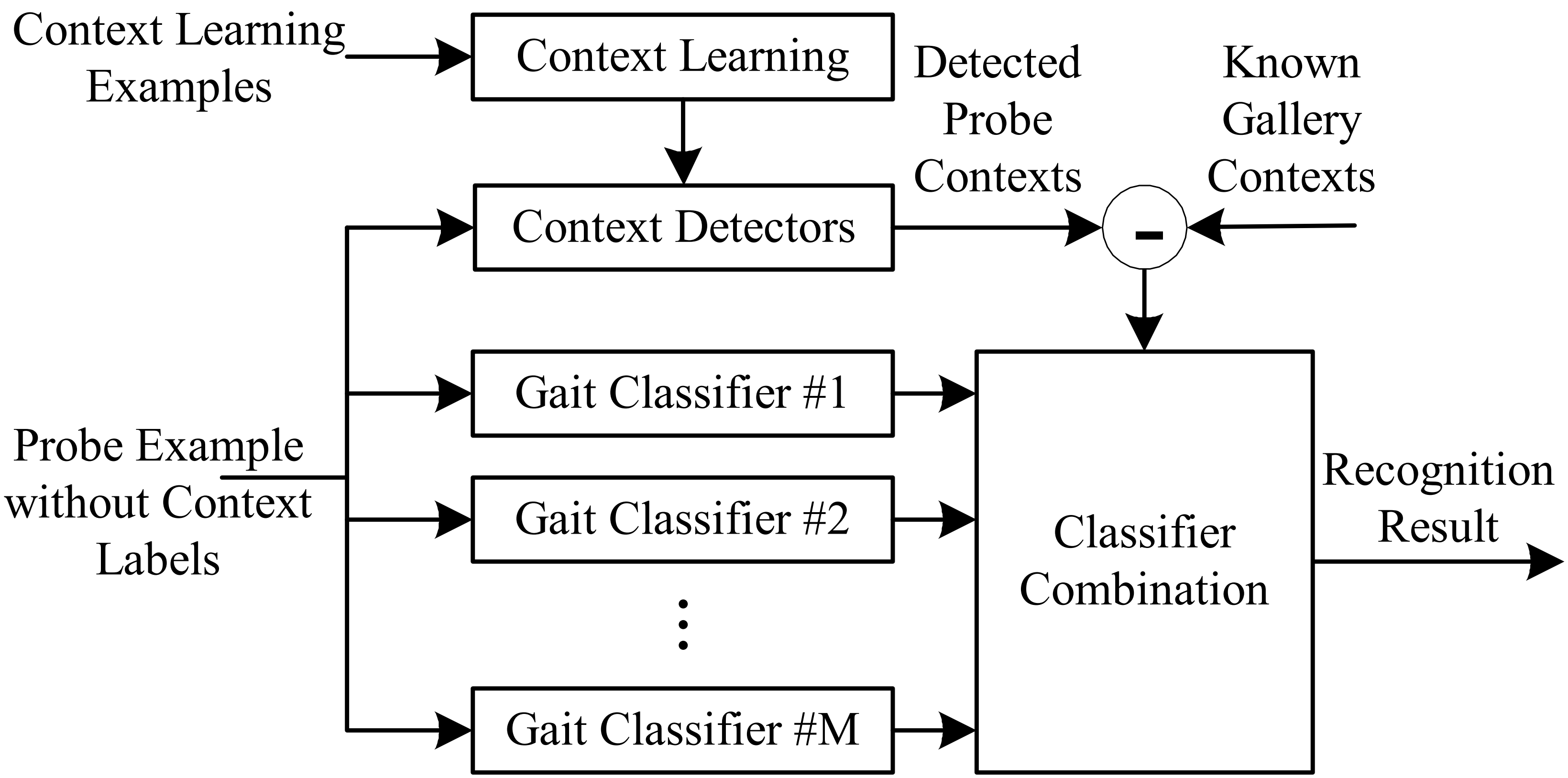
Noon

Late Afternoon

Late Afternoon



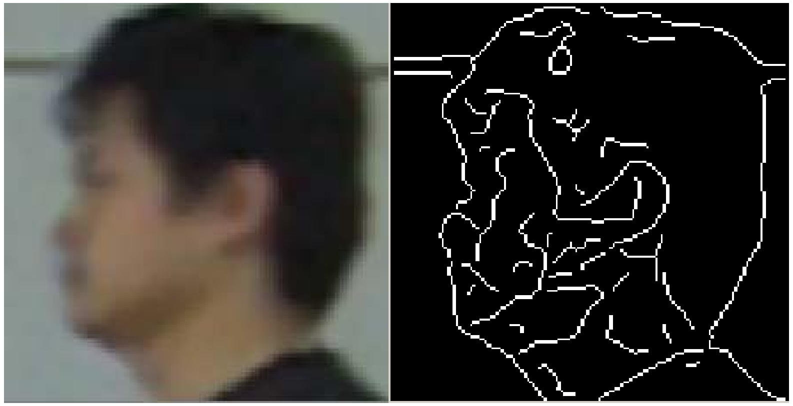
Context-based Recognition (AVBPA 2005)



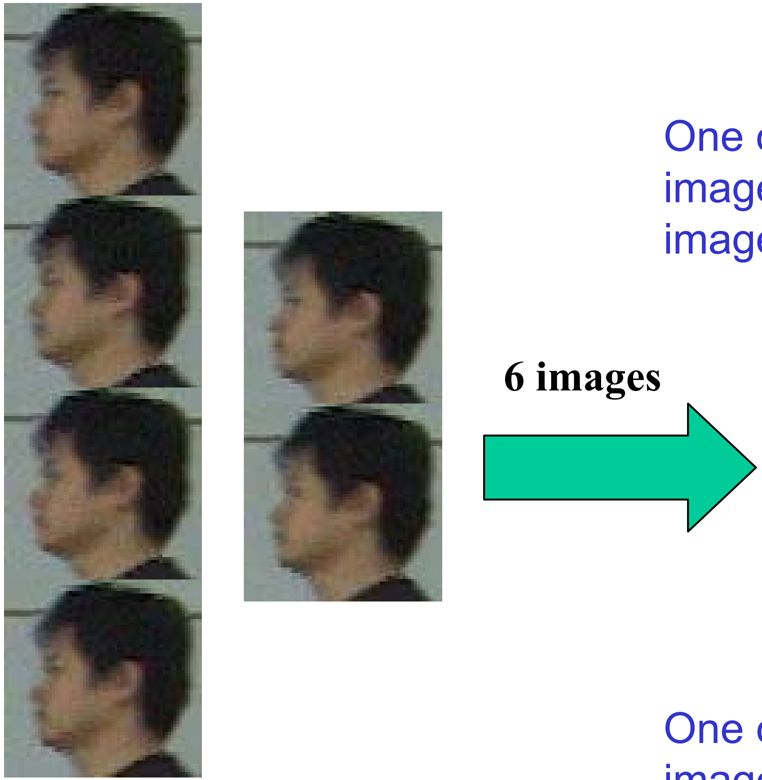
Environmental contexts: walking surface, carrying objects, recording time, etc.

Center for Research in Intelligent Systems

Comparison of Low- and high-resolution Images (AVBPA 2005)



One of the low-resolution face profile image and its edge image are resized by using bilinear interpolation. Original image size is 70x70.



One of the reconstructed high-resolution face profile image and its corresponding edge image. The resolution is 140x140.

Audio-Video Fusion (IEEE CVPR OTCBVS 2005)

"Two-person-test1"

"Two-person-test2"

"Lab"

"Behind-bush-test1"

"Behind-bush-test2"

Hierarchical Image Registration Using EC (GECCO 2005)

Original Color Images

Original Thermal Images

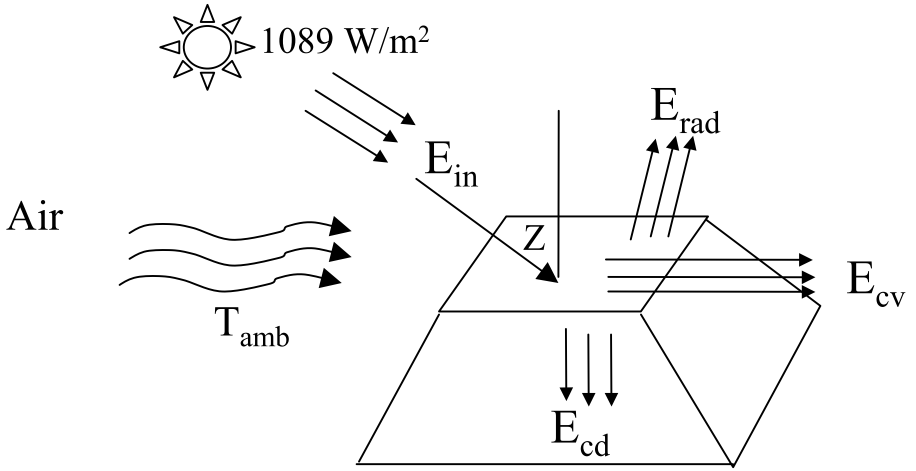
Registered Color Images

Predictive Physical Models (Thermal)

Energy Equilibrium Model

$$E_{in} = E_{out}$$

$$E_{out} = E_{rad} + E_{cv} + E_{cd}$$



Energy equilibrium model for various energies is solved for predicting surface temperature. *[Nadimi and Bhanu, MFI 2003]*

Finding the Optimal Mixture Parameters

Many search techniques exist:

- *Brute Force , Gradient-Based , Heuristics , EM*

Advantages to evolutionary-based methods:

(1) Ideal for parallel structures ,

(2) Do not require complex surface descriptions,

(3) Do not require domain specific knowledge,

(4) Do not require measures of goal distance

Updating background models using Cooperative Coevolution (CC)

Environmental Conditions

Predictive Physical Models (Visible, Thermal)
[Nadimi, and Bhanu, MFI 2003]

Collect a window of frames

Apply
Procedure CC ( )

Create and initialize 4 subpopulations. Each subpopulation consists of individuals of the form:

$<w_1, \mu_1, \sigma_1, \ldots w_k, \mu_k, \sigma_k>$

output

$\boldsymbol{<w_{IR_k}, \mu_{IR_k}, \sigma_{IR_k}, \ldots, w_{VC_k}, \mu_{VC_k}, \sigma_{VC_k}>}$

Representation of an organism

$\mathbf{IR_1}$	$\mathbf{V_1}$
IR_2	V_2
IR_3	V_3
IR_4	V_4
⋮	⋮
IR_m	V_m

$\boldsymbol{<w_{IR_k}, \mu_{IR_k}, \sigma_{IR_k}, \ldots, w_{VC_k}, \mu_{VC_k}, \sigma_{VC_k}>}$

IR = infrared,
VC = Video Channel $\in$ {R,G,B}
k = number of Gaussians

Center for Research in Intelligent Systems

Comparing Traditional GA with Cooperative Coevolution

Procedure GA( )
initialize population
loop
  evaluate individuals
  store best individual

  select mating candidates
  recombine parents and use their
  offsprings as the next generation
until stopping condition

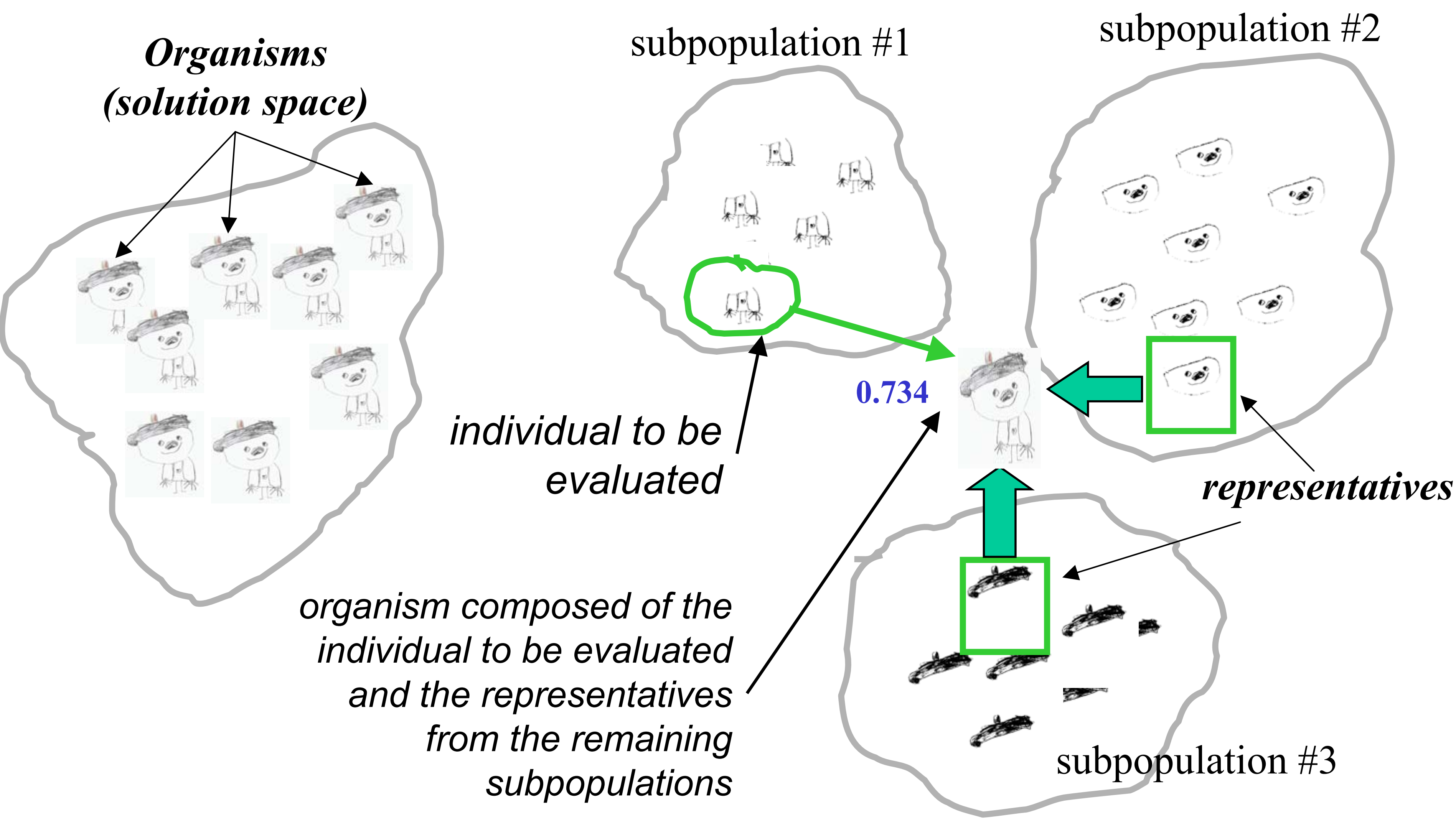
return best individual

Procedure CC( )
initialize subpopulations
loop
evaluate organisms (solutions)
store best organism
for each subpopulation
  select mating candidates
  recombine parents and use their
  offsprings as the next generation
end for
until stopping condition
return best organism

Important Features of Cooperative Coevolution

- The species cooperate to deliver the final solution.
- An individual's fitness depends on its ability to cooperate with the *context* built by the other subpopulations.
- No need for specifying explicitly the fitness function for each subtask.

Idea of Cooperative Coevolution



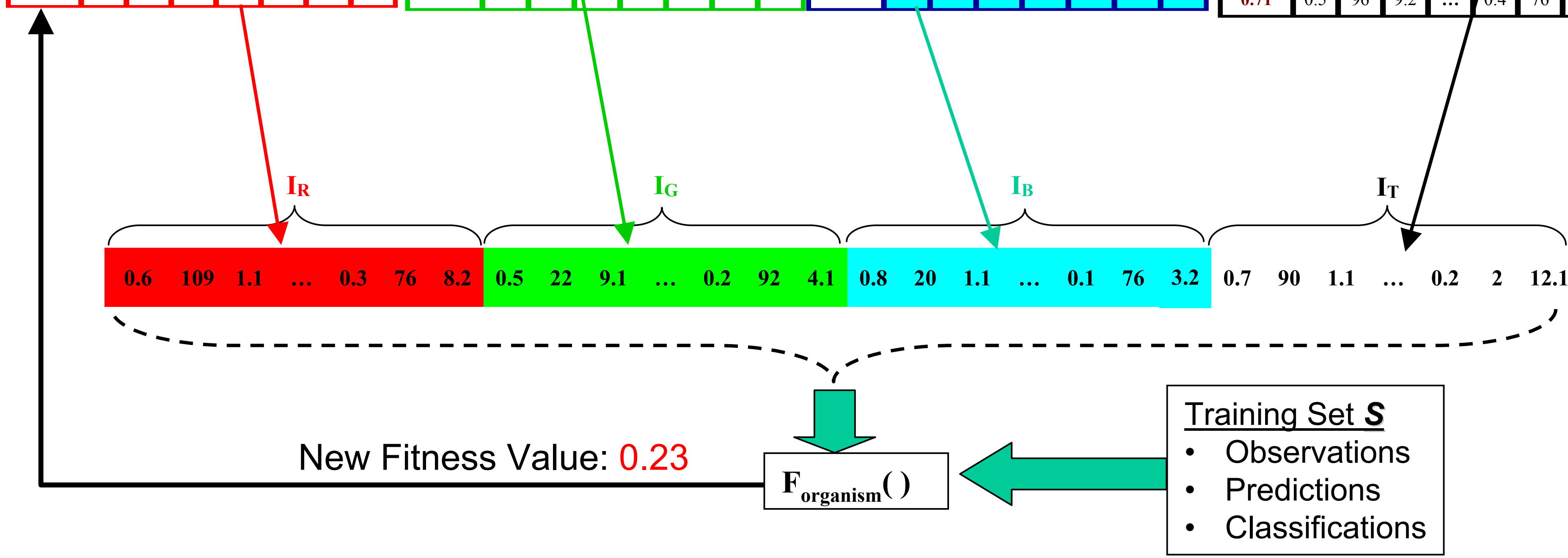
Example of Evaluating an Individual in the Red Channel

Fitness Value	W_1	μ_1	σ_1	…	W_m	μ_m	σ_m
	0.8	211	3.2	…	0.1	172	10.2
	0.7	189	2.7	…	0.2	99	4.1
	.	.	.	…	.	.	.
	.	.	.	…	.	.	.
	.	.	.	…	.	.	.
	0.6	109	1.1	…	0.3	76	8.2

Fitness Value	W_1	μ_1	σ_1	…	W_m	μ_m	σ_m
0.43	0.6	110	1.2	…	0.1	88	91
0.92	0.5	22	9.1	…	0.2	92	4.1
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
0.19	0.4	75	9.3	…	0.3	82	2.2

Fitness Value	W_1	μ_1	σ_1	…	W_m	μ_m	σ_m
0.52	0.3	21	6.2	…	0.1	46	6.9
0.32	0.7	33	2.7	…	0.2	91	4.5
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
0.83	0.8	20	1.1	…	0.1	76	3.2

Fitness Value	W_1	μ_1	σ_1	…	W_m	μ_m	σ_m
0.11	0.4	88	4.2	…	0.1	172	10.2
0.89	0.7	90	1.1	…	0.2	2	12.1
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
.	.	.	.	…	.	.	.
0.71	0.5	96	9.2	…	0.4	76	8.2



Note: Individuals with highest fitness value in their population from other channels at the previous generation are combined to form an organism (solution). The result is stored back for the individual in the red channel.

Evolutionary Adaptive Background Modeling

T = Training set which includes prediction, observation, and previous detection results for each pixel;

Note: An organism represents a solution.

---------------- Cooperative Coevolution Algorithm -----------------

For each pixel

Create and initialize 4 subpopulations (one for each channel R, G, B, Thermal)

Loop

Build 4 organisms (e.g., solution space)

Evaluate organisms using the training set T

Store the best organism

For each subpopulation

Evolve each subpopulation (Selection, Mutation, Crossover)

EndFor

Until stop Condition

Return the best organism

EndFor

Fitness Function

Let Y represents a channel, Y $\in$ {R, G, B, T}

Y_{Xobj} = Observed value of a pixel X at j^{th} frame for channel Y, j = 1 .. n ;

n = Size of the window

Y_{Xpj} = Predicted value of a pixel X by physics for the j^{th} frame for channel Y

$P(Y_X)$ = The probability distribution function for the pixel X for channel Y

G = Training Examples

$$\underset{\{1..n\}}{G_j} = \begin{cases} 1 & Background \\ 0 & Foreground \end{cases}$$

$$\mathbf{F_{organism}\ (<I_R, I_G, I_B, I_T>) = \tfrac{1}{4}\,[C_R\,F(I_R) + C_G\,F(I_G) + C_B\,F(I_B) + C_T\,F(I_T)]}$$

$$F(I_Y) = \frac{1}{n}\sum_{j=1}^{n}\left[G_J P(Y_{X\,ob_J}) + (1 - G_J)(1 - P(Y_{X\,ob_J}))\right]$$

Credibility Function

$$C = e^{-\alpha\left[\frac{1}{n}\sum_{j=1}^{n} G_j \frac{\left|T_{ob_j} - T_{P_j}\right|}{T_{ob_j} + T_{p_j}} + (1-G_j)(1-\frac{\left|T_{ob_j} - T_{P_j}\right|}{T_{ob_j} + T_{p_j}})\right]}$$

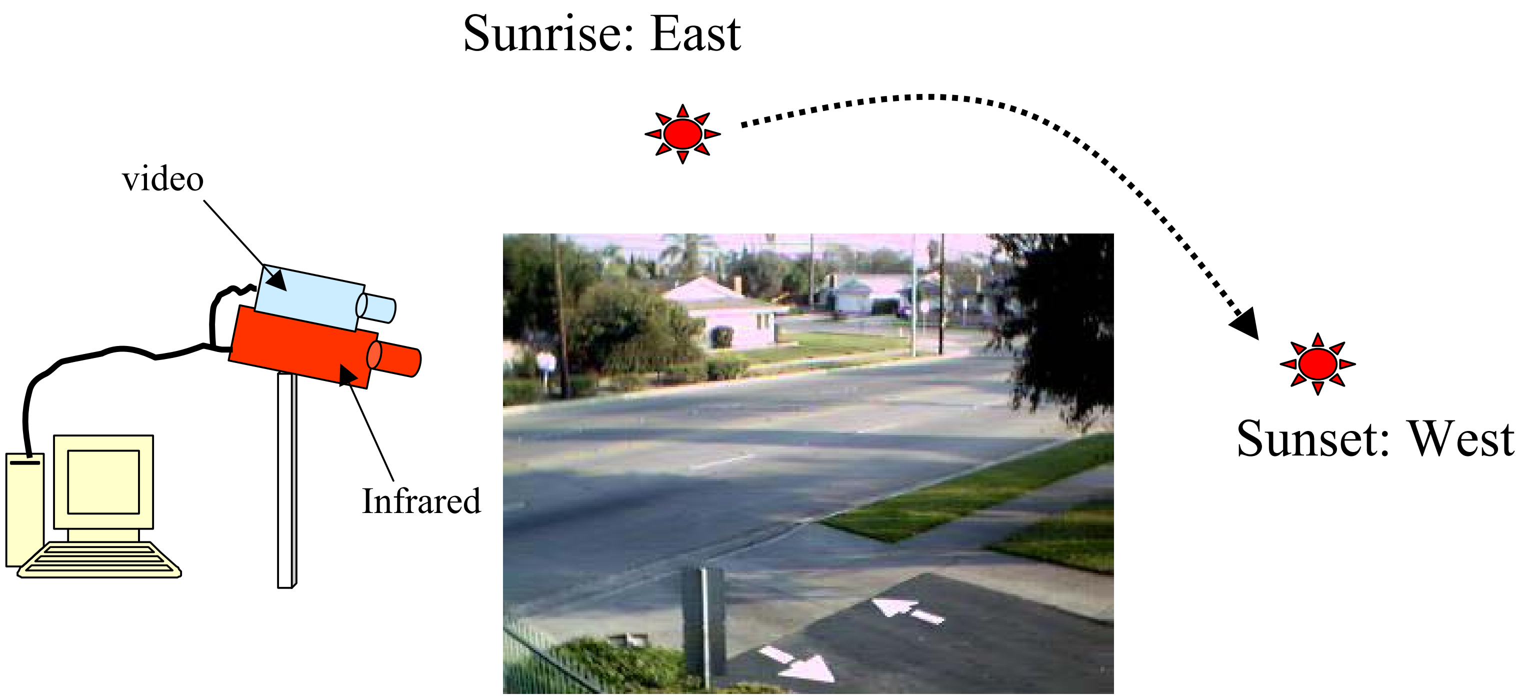
α = rate of credibility function,

$n = 20$

Classification	Difference of currently observed pixel value with the value predicted by Physics: High	Low
Background	LOW	**HIGH**
Foreground	**HIGH**	LOW

Experimental Results

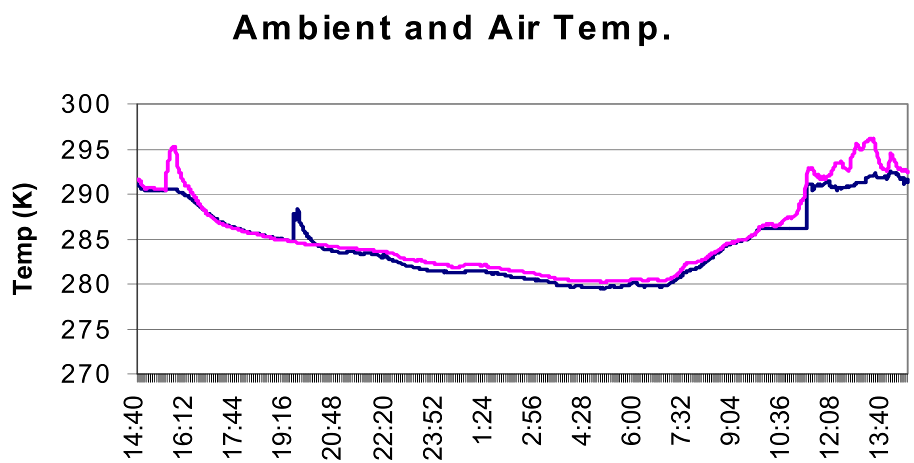
Experimental Setup

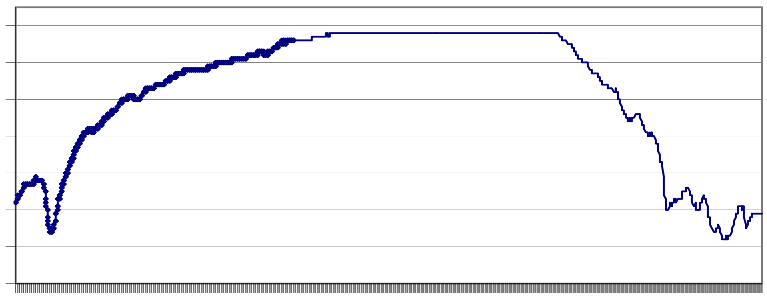


Parameters Used for the Coevolutionary Algorithm

- number of species = 4;
- population size = 60;
- crossover = single point;
- crossover rate = 0.8;
- mutation rate = 0.01;
- maximum number of generations = 60;
- training data = 20 frames;
- number of Gaussians per sensor = 3;
- $\alpha = 0.5$

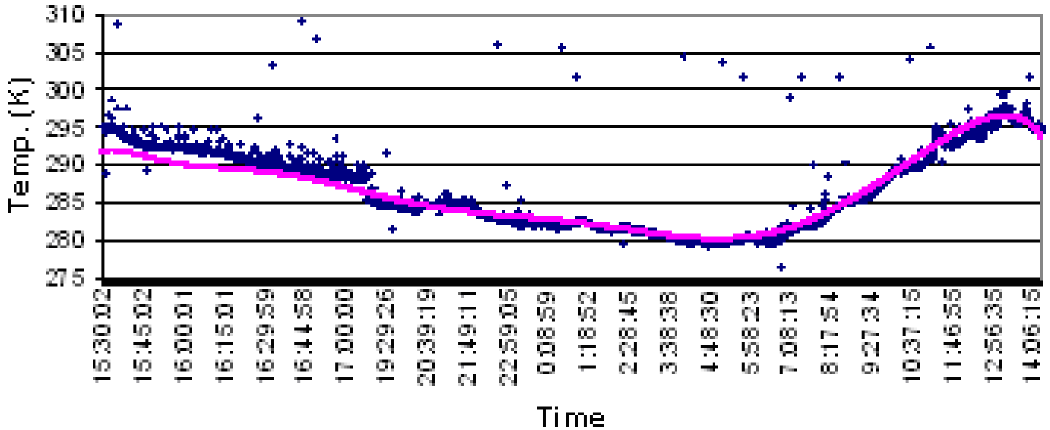
Measurements Used for Thermal Predictions



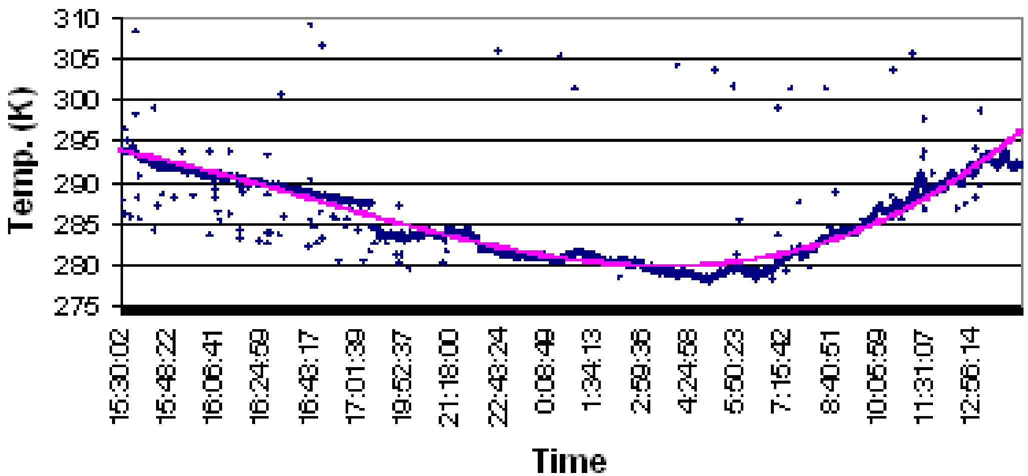


Thermal Prediction

Asphalt: Measured Vs. Predicted Temps.



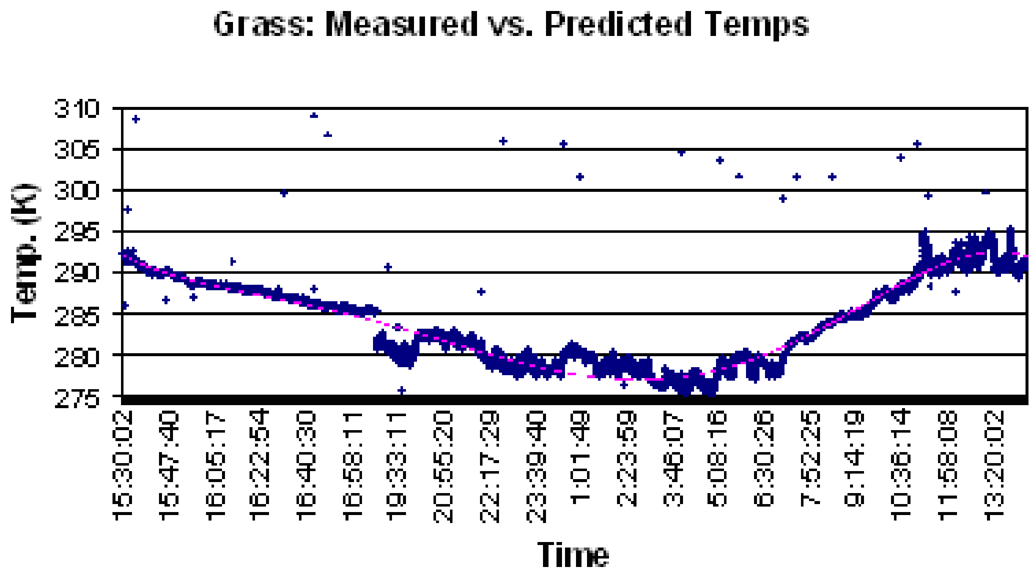
Concrete: Measured vs. Predicted Temps.

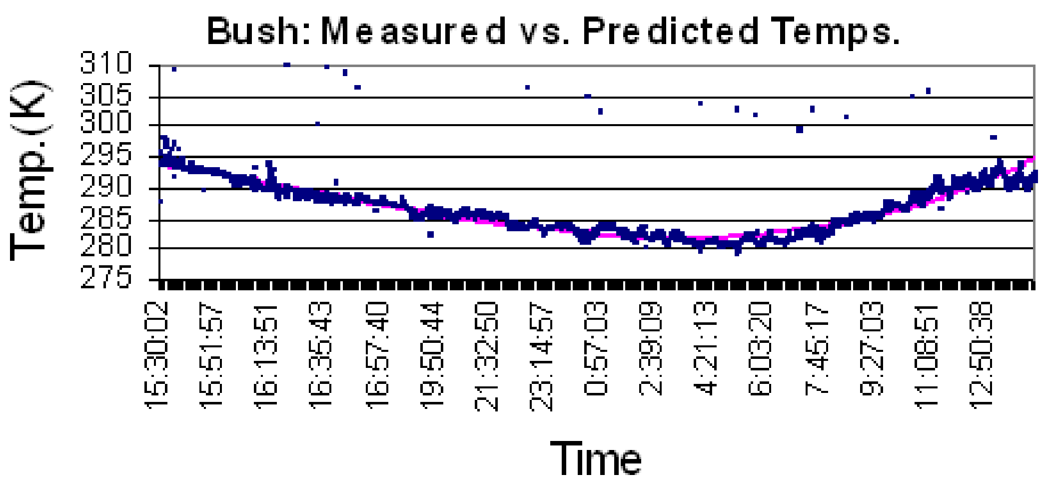


Blue = Measured

Pink = Predicted

Thermal Prediction





Blue = Measured

Pink = Predicted

Dichromatic Surface Body Color Estimation

Time	Asphalt			Concrete			Grass			Bush		
	R	G	B	R	G	B	R	G	B	R	G	B
8:30	.5727	.5726	.5867	**.5813**	**.582**	**.5687**	.6336	.726	.2672	**.5718**	**.6239**	**.5327**
9:30	.5714	.5716	.5889	**.5791**	**.5797**	**.5732**	.6343	.7189	.2844	**.5893**	**.624**	**.5132**
10:30	.5773	.5714	.5862	**.5824**	**.5824**	**.567**	.6369	.7128	.2938	**.5662**	**.6368**	**.5234**
11:30	.5669	.5676	.597	**.5737**	**.5745**	**.5838**	.632	.7193	.2883	**.5476**	**.625**	**.5563**
12:30	.5695	.5695	.5927	**.5686**	**.5749**	**.5884**	.6256	.7376	.2543	**.543**	**.637**	**.5471**
13:30	.5682	.568	.5954	**.5767**	**.5753**	**.5801**	.6249	.7364	.2591	**.5749**	**.6404**	**.5093**
14:30	.5741	.572	.5859	**.5681**	**.5752**	**.5886**	.621	.7391	.2611	**.5968**	**.6338**	**.4921**
15:30	.5635	.552	.6025	**.557**	**.5723**	**.6019**	.606	.7505	.2636	**.5639**	**.6421**	**.5193**
16:30	.5623	.5684	.6006	**.5572**	**.5802**	**.594**	.604	.7572	.2486	**.6567**	**.6369**	**.4039**
17:30	.5544	.5668	.6095	**.5566**	**.5813**	**.5935**	.6231	.738	.259	**.6321**	**.6357**	**.4431**

Center for Research in Intelligent Systems

Afternoon Results

16:58:03
Frame # 2408

16:58:34
Frame # 2422

Registered Image
(Affine transformation)

Detected (fusion)

Detected (fusion)

IR Only

Video Only

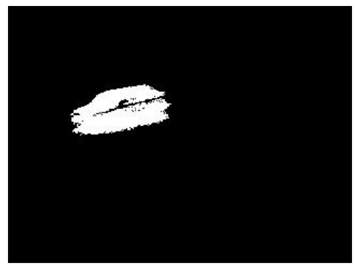
Evening Results

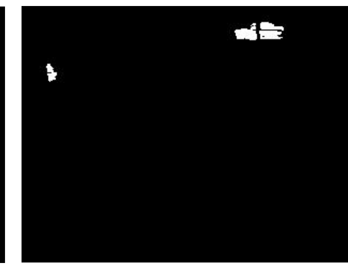
18:56:11
Frame # 2676

19:04:42
Frame # 2726

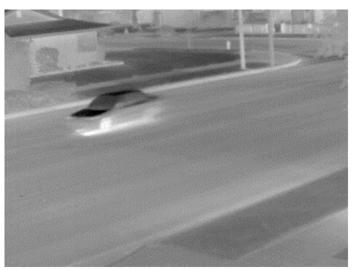
Detected (fusion) | **Detected (fusion)**

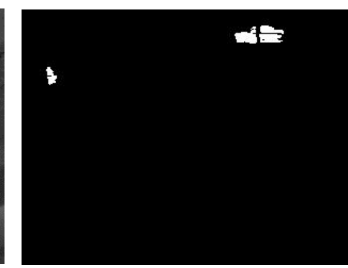
Registered Image

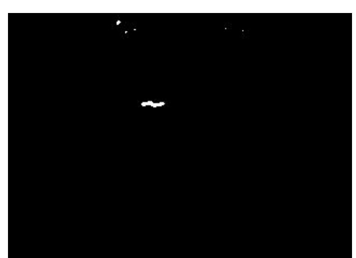


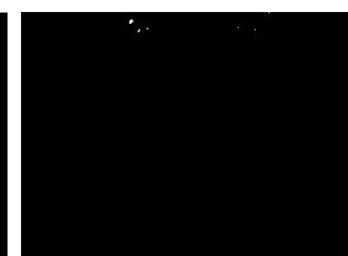
IR Only



Video Only



Morning Results

06:37:46
Frame # 6792

06:54:27
Frame # 6890

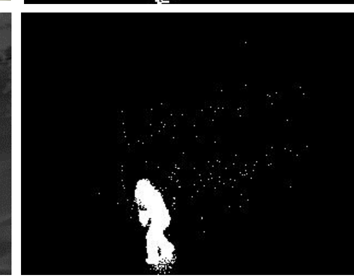
Detected (fusion)

Registered Image




IR Only



Video Only

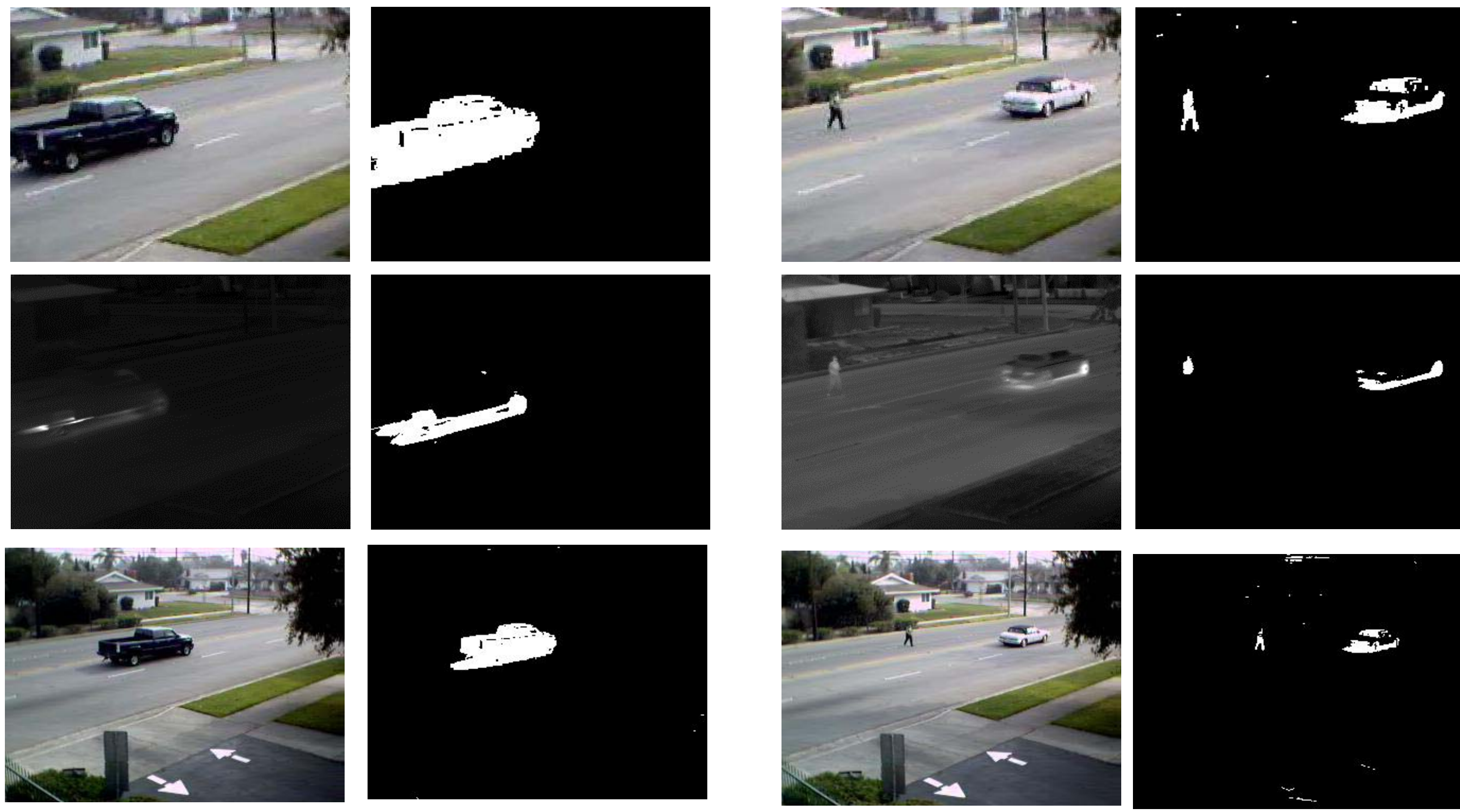



Noon-Early Afternoon Results

11:52:52
Frame # 8646

13:52:29
Frame # 9350

Detected | **Detected**



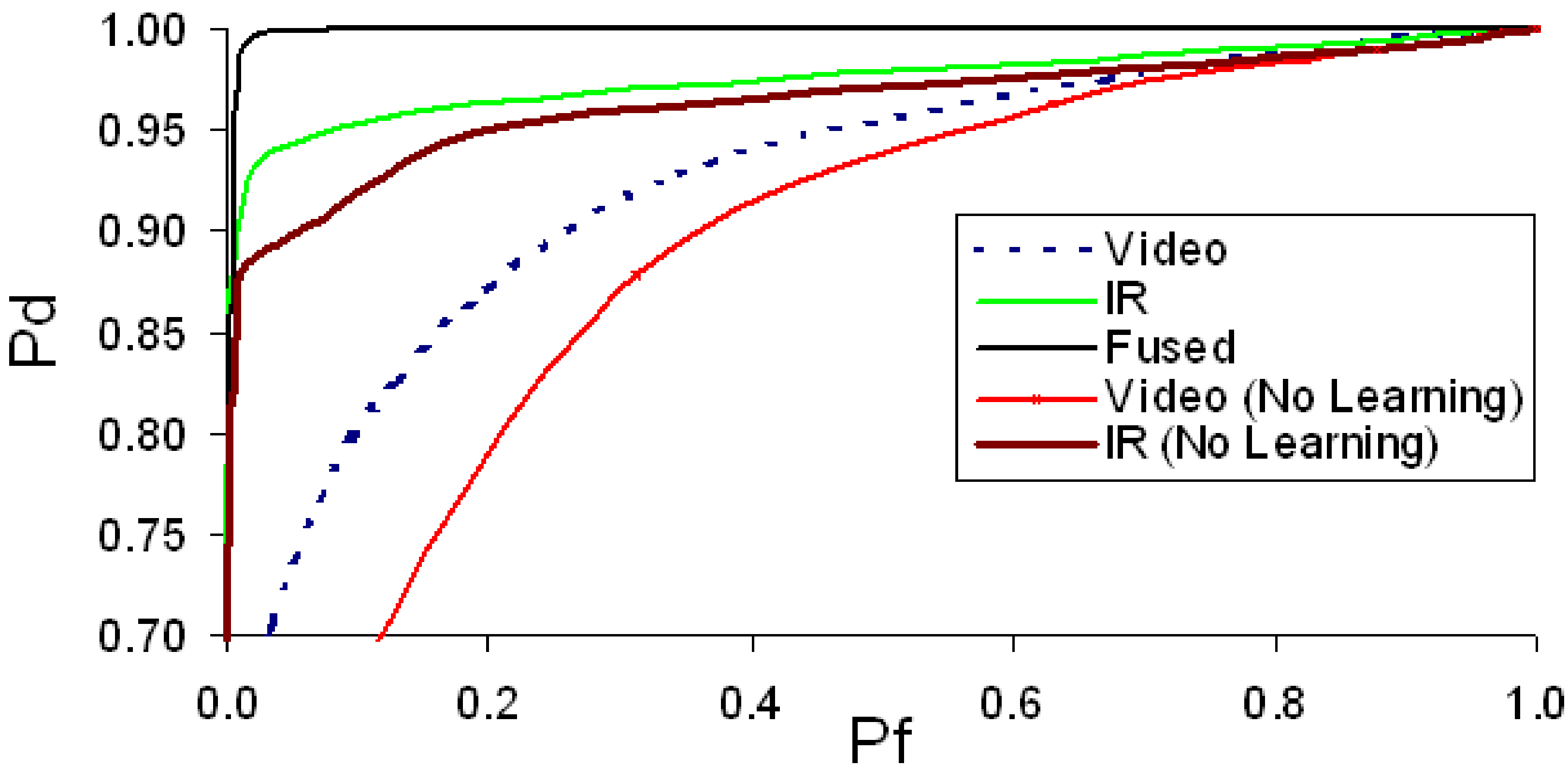
Registered Image

IR Only

Video Only

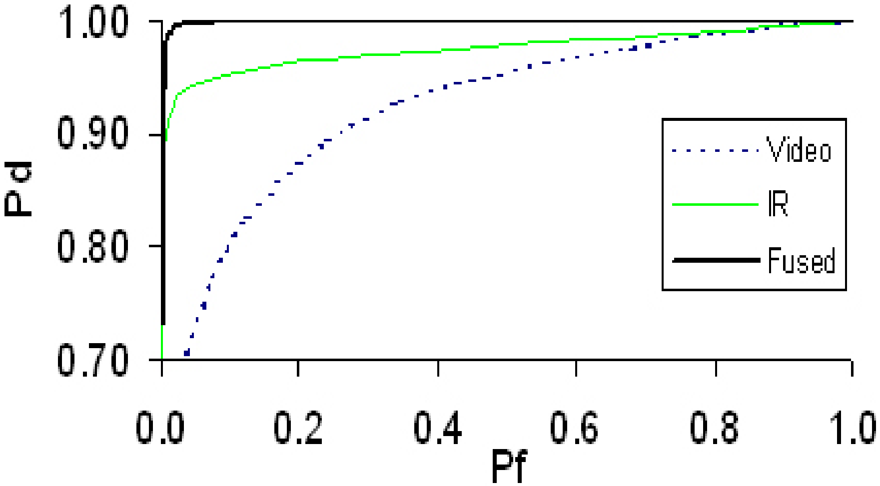
Performance Evaluation – ROC curves

Early Morning

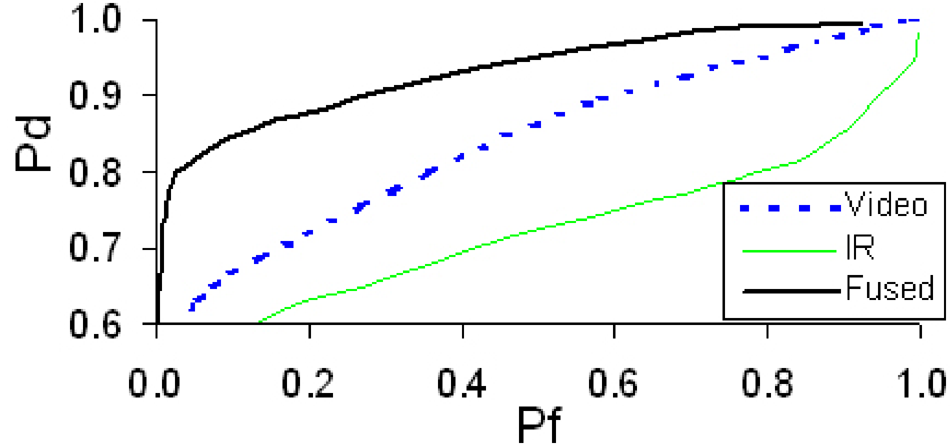


Experimental Results - ROC curves

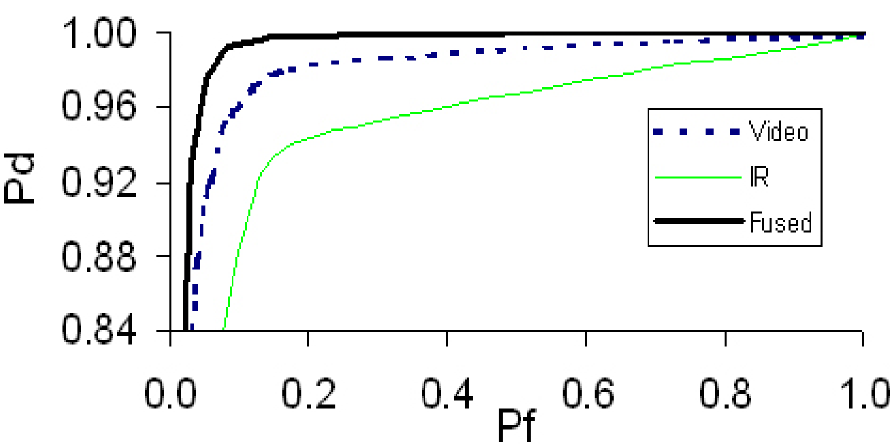
Early Morning



Noon



Afternoon



Pd = Probability of detection
Pf = Probability of false alarm

Conclusions

- **A novel sensor Fusion is introduced.**
- **Learning is achieved through an evolutionary model.**
- **Statistics and phenomenology of the sensors in the visible and longwave IR are integrated through an evolutionary computational model.**
- **Fusion model adapts to various illumination conditions and is suitable for detection under variety of environmental conditions.**
- **Full diurnal (24 hour) cycle result is presented.**

Related Publications

Bhanu, Lin and Krawiec, "Synthesis of Pattern Recognition Systems," Springer, 2005. (Monograph in Computer Science)

Nadimi and Bhanu, "Dynamic Sensor Fusion," SPIE Press, 2006