Healthcare Support through Information Technology Enhancements (hSITE)

Task 2.2.1 **Sensor Information Acquisition and Feature Extraction for the Real-Time Interpretation of Patient and Workflow Information from Video Feeds**,

Task 2.2.3 **Context Aware Multimodal Information Fusion.**

Emil M. Petriu

School of Information Technology and Engineering
University of Ottawa

Annual Research Review
June 16th & 17th, 2011

Our team works on the development of Clinical Care Grade Solutions for Information Gathering from the Complex Multi-Sensor Environments (*Project 2.2*) for **Critical Care** (*Context #1*) and **Home Care** (*Context #2*) applications.

Theme 2: Context Aware Sensors Systems, Software and Applications

Project 2.2: Information Gathering from the Complex Multi-Sensor Environment

Task 2.2.1: Sensor Information Acquisition and Feature Extraction

….

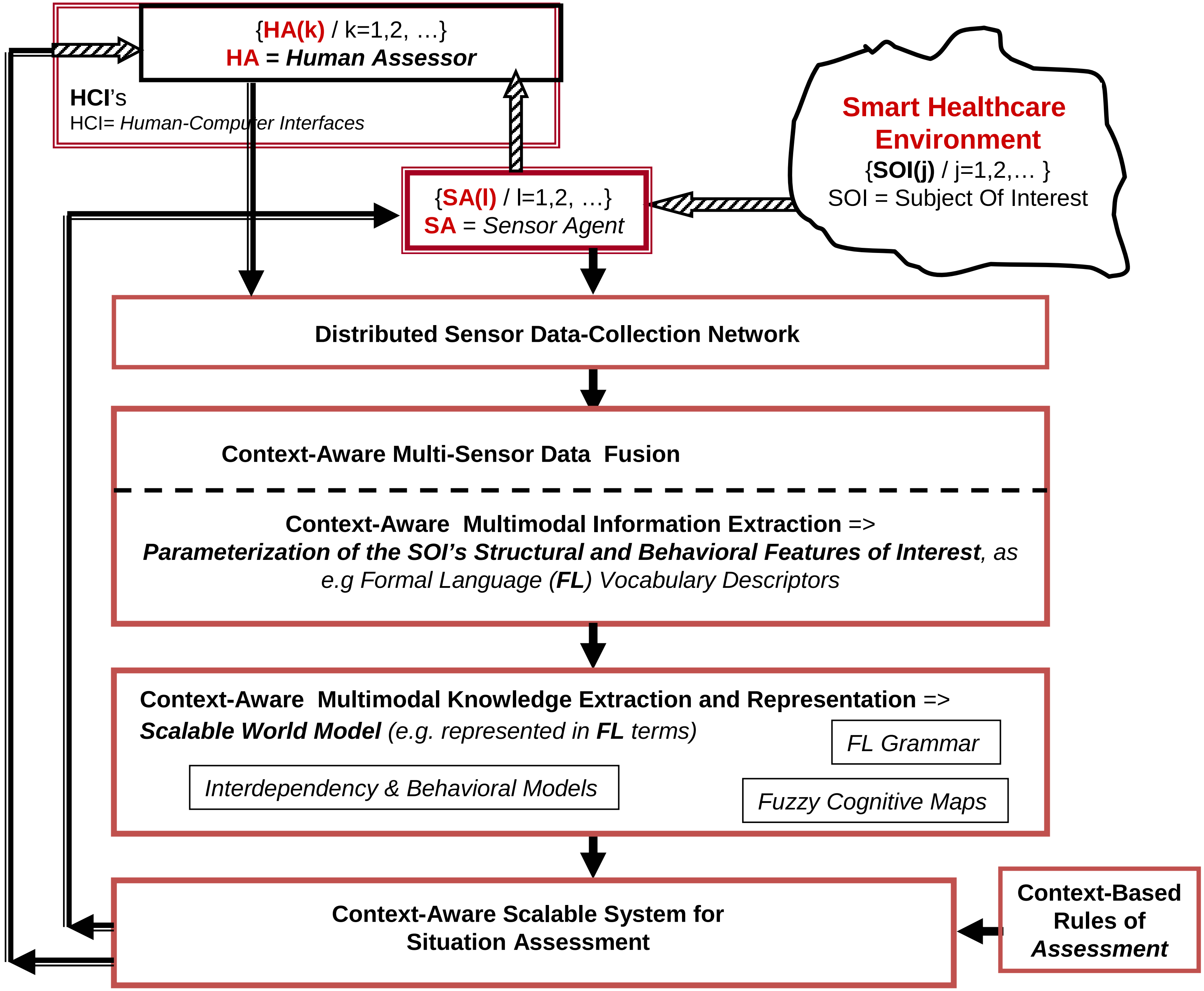
Task 2.2.3: Context Aware Multimodal Information Fusion

Research Group

Name of student / PDF	Program	Task	Start date	Graduation	Funded by:
Dr. Ana-Maria Cretu	PDF	2.2.1	Jan. 2009	Oct. 2010	hSITE & other
Md. A. Hossain	PhD	2.2.3	Jan. 2010	July 2011	hSITE & other
Dr. Qing Chen	PDF	2.2.1.	Oct. 2008	March 2011	other
Dr. Albino Cordeiro	PDF	2.2.3	Jan 2009	May 2011	other
Haifa Maamar	PhD	2.2.3	Jan. 2010	Aug. 2013	other
Alexandros Stathakis	MASc.	2.2.1	Sept. 2010	Aug. 2012	hSITE & other
Feng Shi	PhD	2.2.3	Sept 2010	Aug/ 2012	other
Eric Torunski	PhD	2.2.3	Sept. 2010	Aug. 2012	hSITE & other
Adrian Taylor	PhD	2.2.3	Sept. 2010	Aug. 2014	hSITE & other
Suzan Ureten	PhD	2.2.1	Sept. 2010	Aug. 2013	hSITE & other
Yisu Zhao	PhD	2.2.1	Jan 2010	Aug. 2012	other

Smart-healthcare environments incorporate a multitude of *time*- and *location-dependent sensor-data*, from which is possible to extract relevant information about *patient condition* (identity, location, physiological parameters), *clinical staff status* (identity, location, readiness), *specific clinical activities, medication, supplies, and equipment status* (identity, location, specs), *operating room readiness, state of the ambient environment*, etc.

Context understanding in these environments require dynamic sensor configurations and measurement capabilities similar to human perception, which pose a considerable challenge to the traditional sensor fusion methods. *Location, together with time, represents one of the basic contextual information to any context-aware system*.


{HA(k) / k=1,2, …}
HA = Human Assessor
HCI's
HCI= Human-Computer Interfaces
Smart Healthcare Environment
{SOI(j) / j=1,2,… }
SOI = Subject Of Interest
{SA(l) / l=1,2, …}
SA = Sensor Agent
Distributed Sensor Data-Collection Network
Context-Aware Multi-Sensor Data Fusion
Context-Aware Multimodal Information Extraction =>
Parameterization of the SOI's Structural and Behavioral Features of Interest, as e.g Formal Language (FL) Vocabulary Descriptors
Context-Aware Multimodal Knowledge Extraction and Representation =>
Scalable World Model (e.g. represented in FL terms)
FL Grammar
Interdependency & Behavioral Models
Fuzzy Cognitive Maps
Context-Aware Scalable System for Situation Assessment
Context-Based Rules of Assessment

The **multi-sensor fusion architecture** is based on the **mission-critical *JDL Data Fusion Model*** developed by the Joint Directors of Laboratories Data Fusion Group .

This architecture has five functional levels.

* *level 0* "Signal/Feature Assessment" and *level 1* "Entity Assessment" essentially asses the measurement data;
* *level 2* "Situation Assessment";
* *level 3* "Impact Assessment" essentially asses the information recovered from data;
* *level 4* "Performance Assessment" provides sensor management functions for process refinement.
* a supplementary knowledge-management *level 5* "User Refinement" is used delineate the human from the computer in the process refinement and allow for the adaptive decision of who can query and respectively access the information and the collected data in order to support cognitive decision support and actions.

- Human sensor information is “fuzzy quantizied” while the machine sensor information, both the symbiotic analog transducer & human, and the fully automated digital one, is “sharp & concatenated quantized”

- It is possible to reduce the uncertainty of the measurements involving humans as sensors part of multi-sensor systems, by using Fuzzy Cognitive Maps, NNs, and Associative Memories.

- Dempster-Shafer theory of evidence approach is used to incorporate human-like uncertainty management and inference mechanisms in our context-aware multi-sensor data fusion system. This approach allows us to incorporate time-variable weights representative of sensor precision which will improve the sensor fusion accuracy in dynamic environments.

- Linguistic pattern recognition techniques and semantic model representations are used to develop a semantic level situation assessment system that will allow understanding of the dynamics of a complex scene based on multimodal sensor data streams.

We **met the objectives for the following milestones**:

- **[M2.7.c /** *Task 2.2.1* **]** Evaluation and refinement of the real-time body posture recovery algorithms for video interpretation, **[Shi11].**

- **[M2.8.a /** *Task 2.2.1* **]** Development of linguistic pattern recognition algorithms for the hand gesture recognition and combination with body posture analysis, **[Dard10], [Zhao11]**

- **[M2.13.b /** *Task 2.2.3* **]** Evaluation of the fuzzy and stochastic information fusion algorithms for clinician assistance, **[Cret10], [Wide10]**,

- **[D/P 2.a]** Design of an experimental multi-sensor setup including audio sensors, cameras, temperature sensors, accelerometers, and pressure-sensitive arrays; integration of microphone array and steerable video cameras, **[Kim10], [Maam10], [Maam10a],**

We are currently working on the following milestones

- **[M2.8.b /** *Task 2.2.1* **]** Evaluation of the linguistic pattern recognition algorithms developed for the hand gesture recognition in combination with the body-posture analysis,

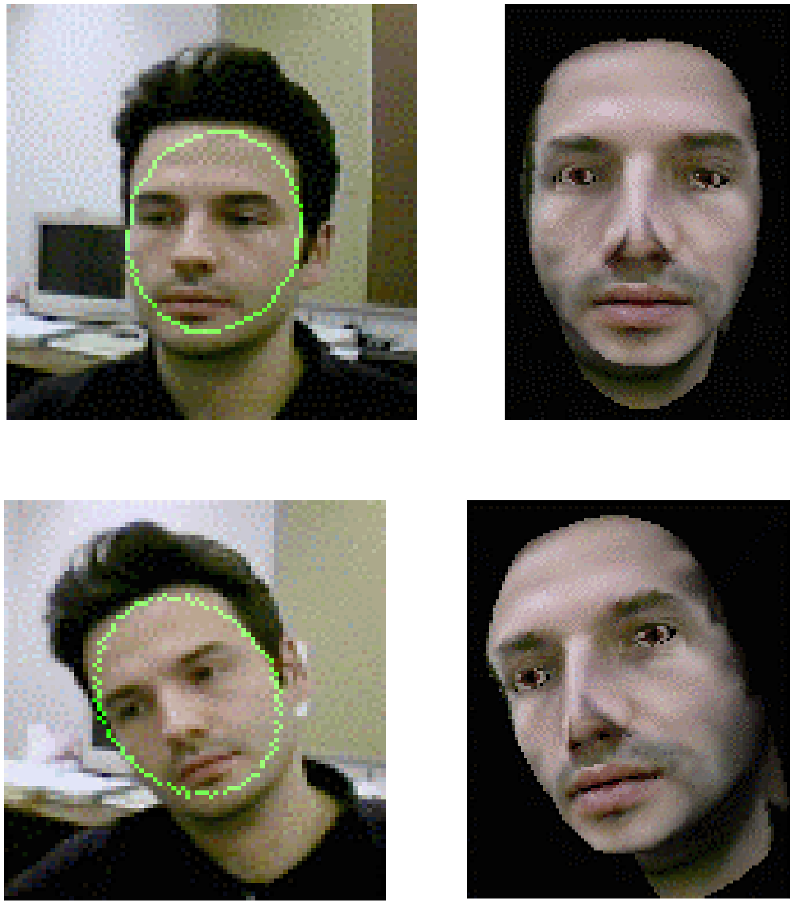
- **[M2.14.a /** *Task 2.2.3* **]** Study of context inference engine architectures based on statistic and machine learning techniques in the Context Management Framework studied by the group working on Task 2.1.2

- **[D/P 2.b]** Deployment and evaluation of the experimental of sensor setup including audio sensors, cameras, temperature sensors, accelerometers, and pressure-sensitive arrays; integration of microphone array and steerable video cameras.

- **[D/P 2.c**] Hardware/software implementation of the real-time algorithms for the body posture recovery and hand gesture recognition into the multi-sensor setup.

Real-time tracking of 2½D head parameters: position and orientation

The skin color distribution of people with different skin colors forms a compact cluster, with a regular shape in *rg* (or *HS*)-chromatic color space. => Modeling human faces as a *Mixture of Gaussian* (MOG) distributions in the 2D normalized color space.



Once the head is detected, an elliptical outline is fitted to the head contour. Every time a new image becomes available, the tracker will try to fit the ellipse model from the previous image in such a way to best approximate the position of the head in the new image. Essentially, **tracking** *consists of an update of the ellipse state to provide a best model match for the head in the new image*. The state is updated by a hypothesize-and-test procedure in which the goodness of the match is dependent upon the intensity gradients around the object's boundary and the color of the object's interior.

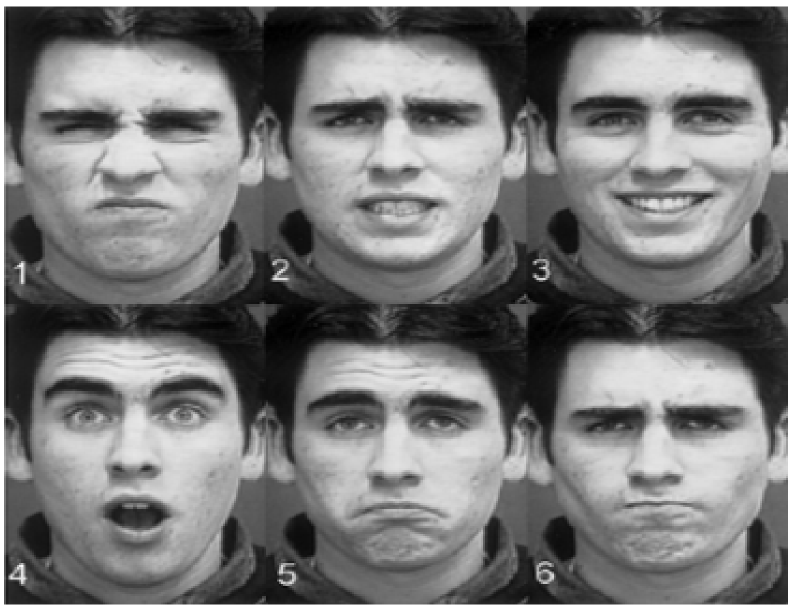
Model-Based Face Tracking and Expression Recognition

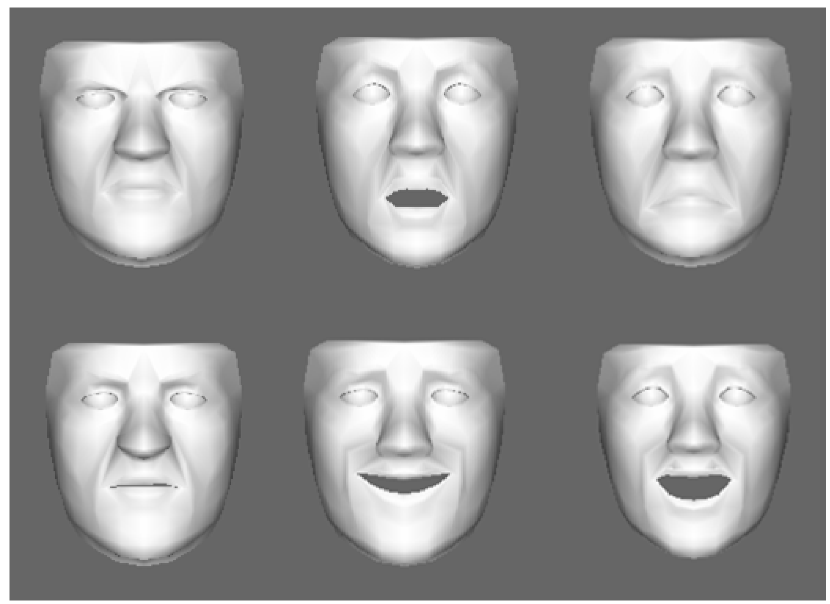
3D Face Modeling

- Modeling and animating realistic faces require knowledge of anatomy
  - Anthropometric (external) representation
    - Measurements of living subjects
    - Statistics based on age, health, etc.
  - Muscle/Skin (internal) representation
    - Over 200 facial muscles
    - Over 14,000 possible expressions

Model-Based Facial Expression Recognition

- Simulate the dynamics of real muscles
- Facial Action Coding System (FACS)
  - Facial articulation as Expression Action Units (EAUs)
  - 7 pairs of muscles + "Jaw Drop" = Expression Space



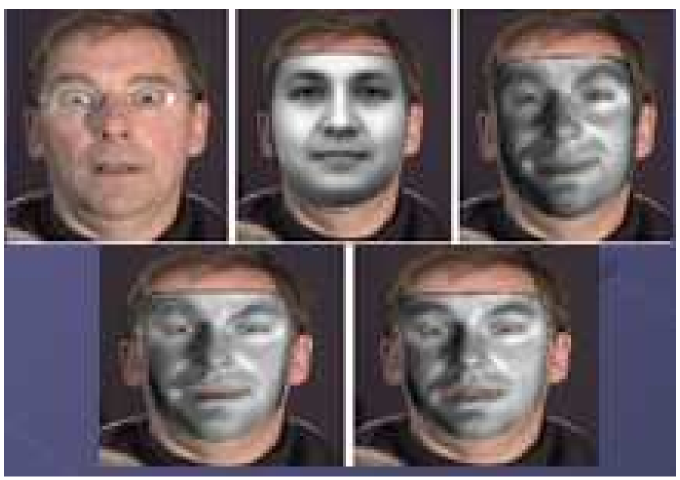


Facial Expression Recognition

- Person Dependent

AU	Signification	No.	Correct	False	Missed	Confused	Recognition Rate
0	Neutral	20	19	1	0	0	95%
1	Inner Brow Raiser	24	21	0	3	0	87.5%
2	Outer Brow Raiser	52	41	1	9	1	78.8%
4	Brow Lowerer	42	41	0	0	1	97.6%
12	Lip Corner Puller	51	48	3	0	0	94.1%
15	Lip Corner Depressor	18	17	0	1	0	94.4%
26	Jaw Drop	74	55	0	19	0	74.3%
Total		281	242	5	32	2	86.1%
False Alarm: 1.7%, Missed: 11.3%							

- Person Independent



AU	Signification	No.	Correct	False	Missed	Confused	Recognition Rate
0	Neutral	20	17	3	0	0	85.0%
1	Inner Brow Raiser	10	8	0	2	0	80.0%
2	Outer Brow Raiser	27	20	1	5	1	74.0%
4	Brow Lowerer	24	21	1	1	1	87.5%
12	Lip Corner Puller	17	13	0	4	0	76.4%
15	Lip Corner Depressor	13	10	1	2	0	76.9%
26	Jaw Drop	24	14	0	10	0	58.3%
Total		135	103	6	24	2	76.2%
False Alarm: 4.4%, Missed: 17.7%							

Publications related to our work undertaken as a direct result of the hSITE technical agenda, but funded from complementary sources

[Maam10] H.R. Maamar, A. Boukerche, E.M. Petriu, "MOSAIC - A Mobile Peer-to-Peer Networks-Based 3D Streaming Supplying Partner Protocol," *Proc. 2010 IEEE/ACM 14th Int. Symp. Distributed Simulation and Real Time Applications (DS-RT*), pp. 61-68, Oct. 2010.

[Wide10] P. Wide, E.M. Petriu, M. Siegel, "Sensing and Perception for Rehabilitation and Enhancement of Human Natural Capabilities," *Proc. ROSE 2010, IEEE Int. Workshop on Robotic and Sensor Environments*, Phoenix, AZ, USA, pp. 75-80, Oct. 2010.

[Maam10a] H.R. Maamar, A. Boukerche, E.M. Petriu, "Load balancing model for mobile peer-to-peer networks-based 3D streaming," *Proc. HAVE 2010 -IEEE Int. Symp. Haptic Audio Visual Environments and Games*, pp. 1-6, Phoenix, AZ, Oct. 2010.

[Cret10] A.-M. Cretu, E.M. Petriu, P. Payeur, F. F. Khalil, "Estimation of Deformable Object Properties from Visual Data and Robotic Hand Interaction Measurements for Virtualized Reality Applications," *Proc. HAVE 2010-IEEE Int. Symp. Haptic Audio Visual Environments and Games* pp. 168-173, Phoenix, AZ, Oct. 2010.

[Dard10] N. Dardas, Q. Chen, N.D. Georganas, E.M. Petriu, "Hand gesture recognition using Bag-of-features and multi-class Support Vector Machine," *Proc. HAVE 2010 -IEEE Int. Symp. Haptic Audio Visual Environments and Games*, pp. 1-5, Phoenix, AZ, Oct. 2010.

[Kim10] G. Kim, E. M.Petriu, "Fiducial Marker Indoor Localization with Artificial Neural Network," *Proc. 2010 IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics - AIM 2010*, pp. 961- 966, Montreal, Que. Canada, July 2010.

[Zhao11] Y. Zhao, M.D. Cordea, E.M. Petriu, T.E. Whalen, "Multimedia Based Bidirectional Affective Human-Computer Interaction," submitted to *Multimedia Image and Video Processing (*Ling Guan, Yifeng He, Sun-Yuan Kung – Editors), CRC Press, 2011

[Shi11] F. Shi, E.M. Petriu, A. Cordeiro, "Human Action Recognition from Local Part Model**,"** *Internal Report Discover Lab*., University of Ottawa, 2011

Thank You !