# Energy Aware Recognition for Man Made Structures and other research projects at the American University of Beirut

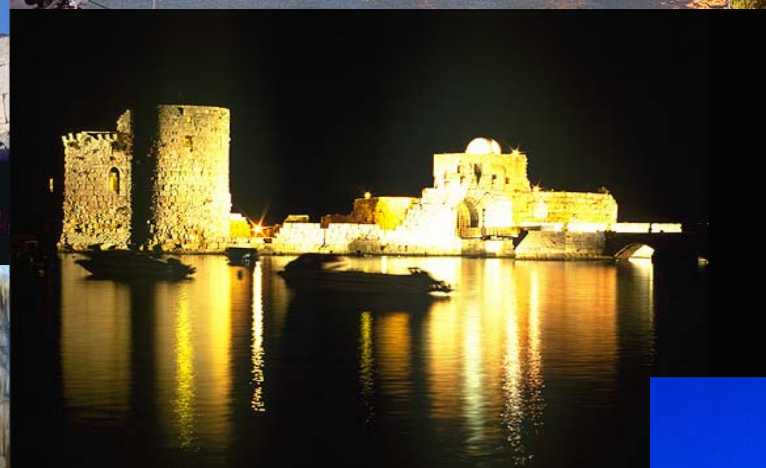
Mariette Awad
Assistant Professor
Electrical and Computer Engineering Department
American University of Beirut, Lebanon

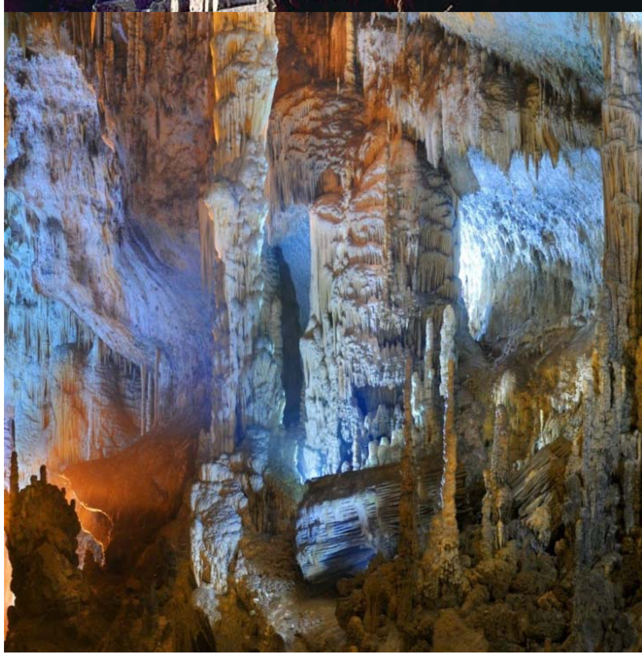
# Outline

- On Going Projects
- uSee for Man made Structures
- Biologically Inspired Deep Visual Networks

# Lebanon







# American University of Beirut

# Some AUB General Info

- AUB founded in 1866
- Since 2004 accredited by Higher Educational of the Middle States Association of Colleges and Schools in the US
- 120 programs: Bachelor, Masters and PhD degrees
- 6 faculties: Agriculture and Food Sciences, Arts and Sciences, Engineering and Architecture, Health Sciences, Medicine, Business
- Faculty of Engineering since 1944

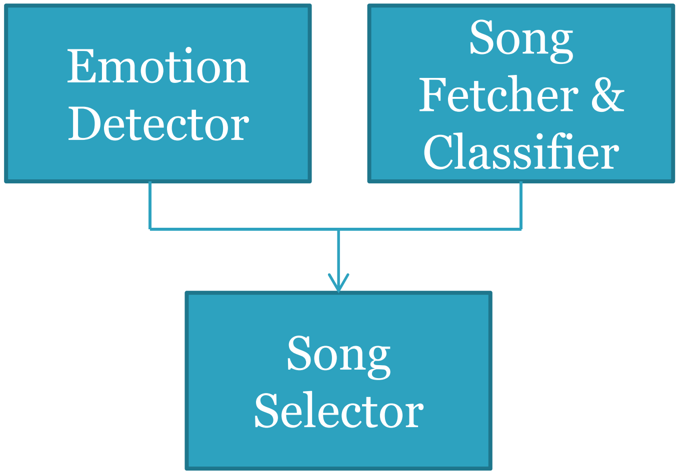
# Mood-Based Internet Radio-Tuner App


<u>Undergraduate Students</u>

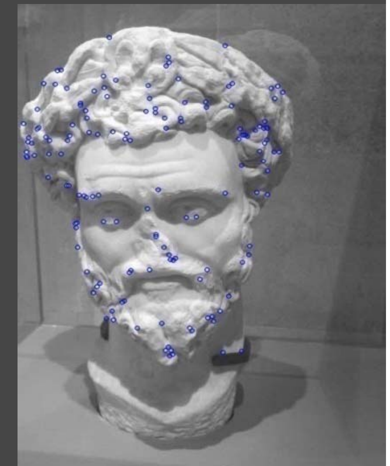
*David Matchoulian*

*Yara Rizk*

*Maya Safieddine*

# MusAR for an Augmented Museum Experience

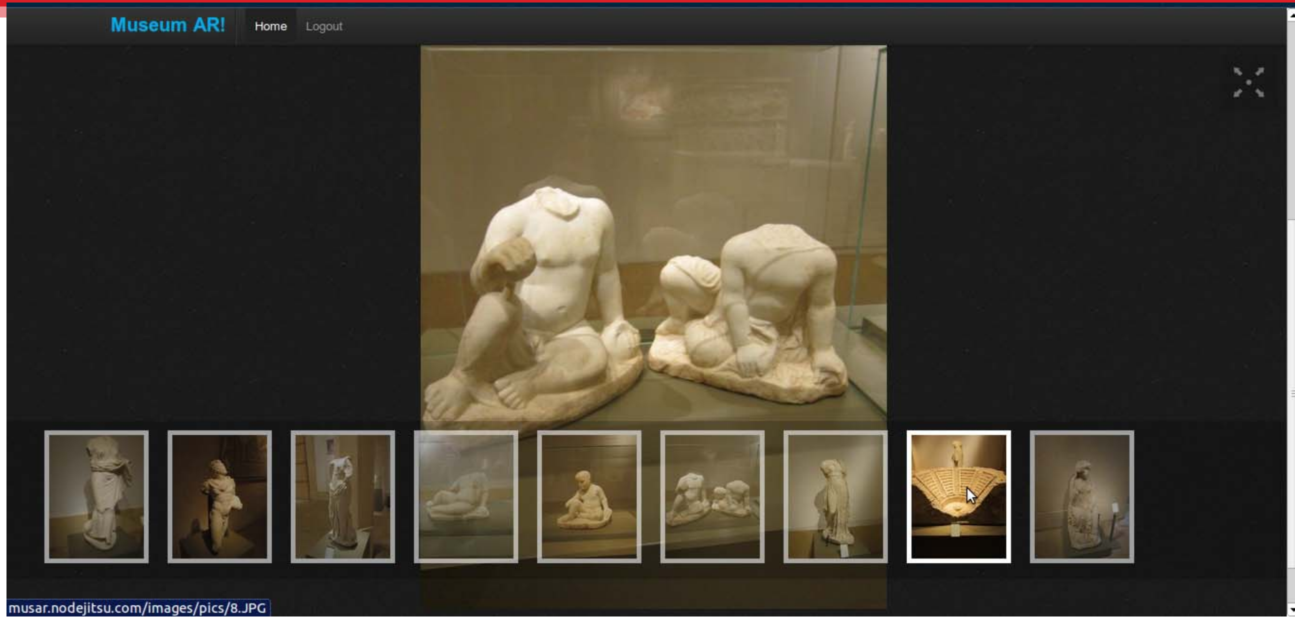


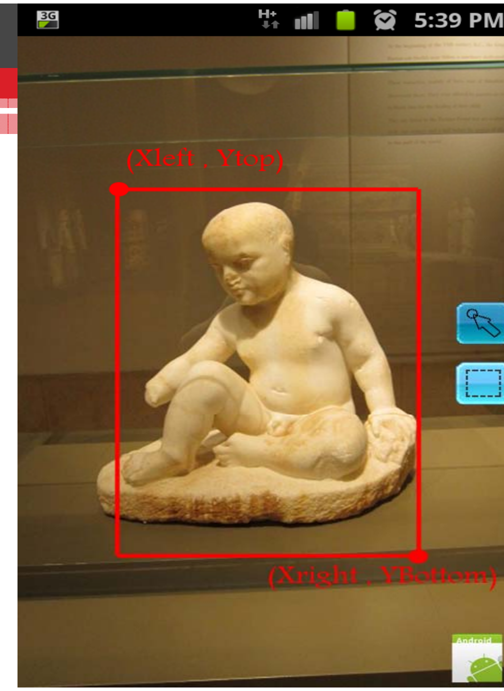

*Undergraduate Students*

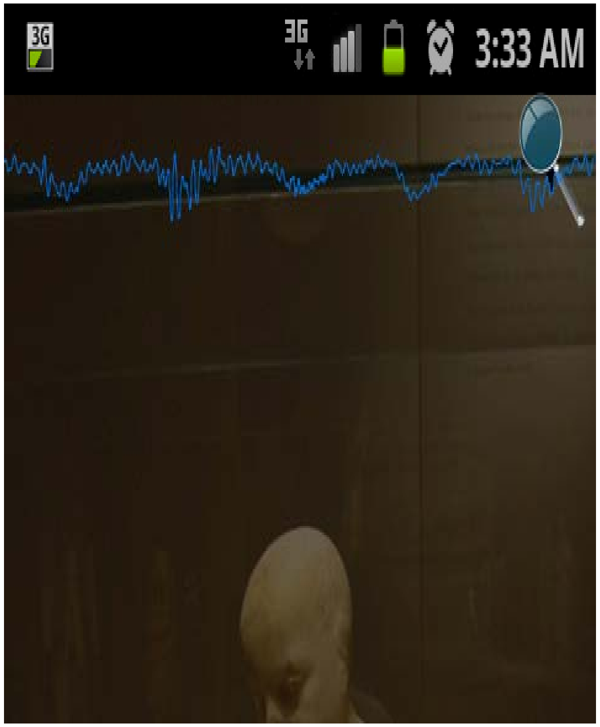
*Anis Halabi*

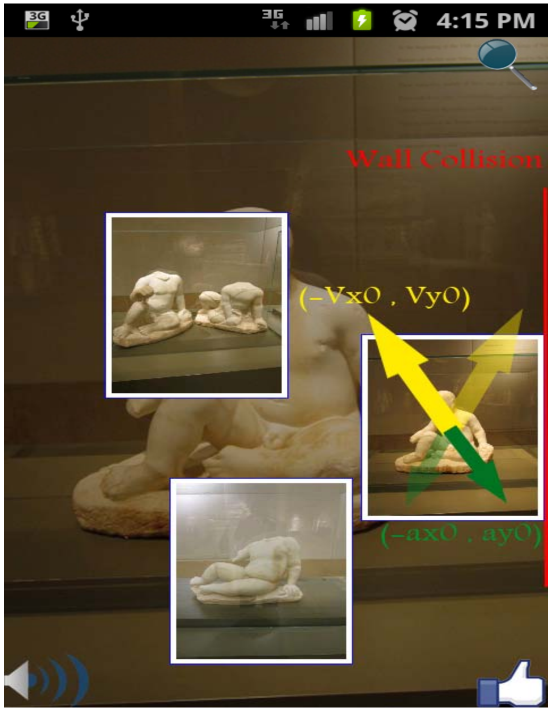
*Giorgio Saad*

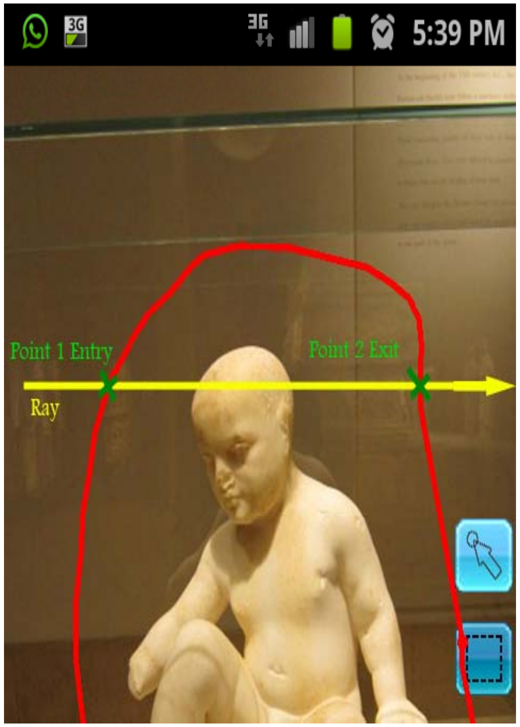
*Piotr Yordanov*


Museum AR!
Home
Logout
musar.nodejitsu.com/images/pics/8.JPG

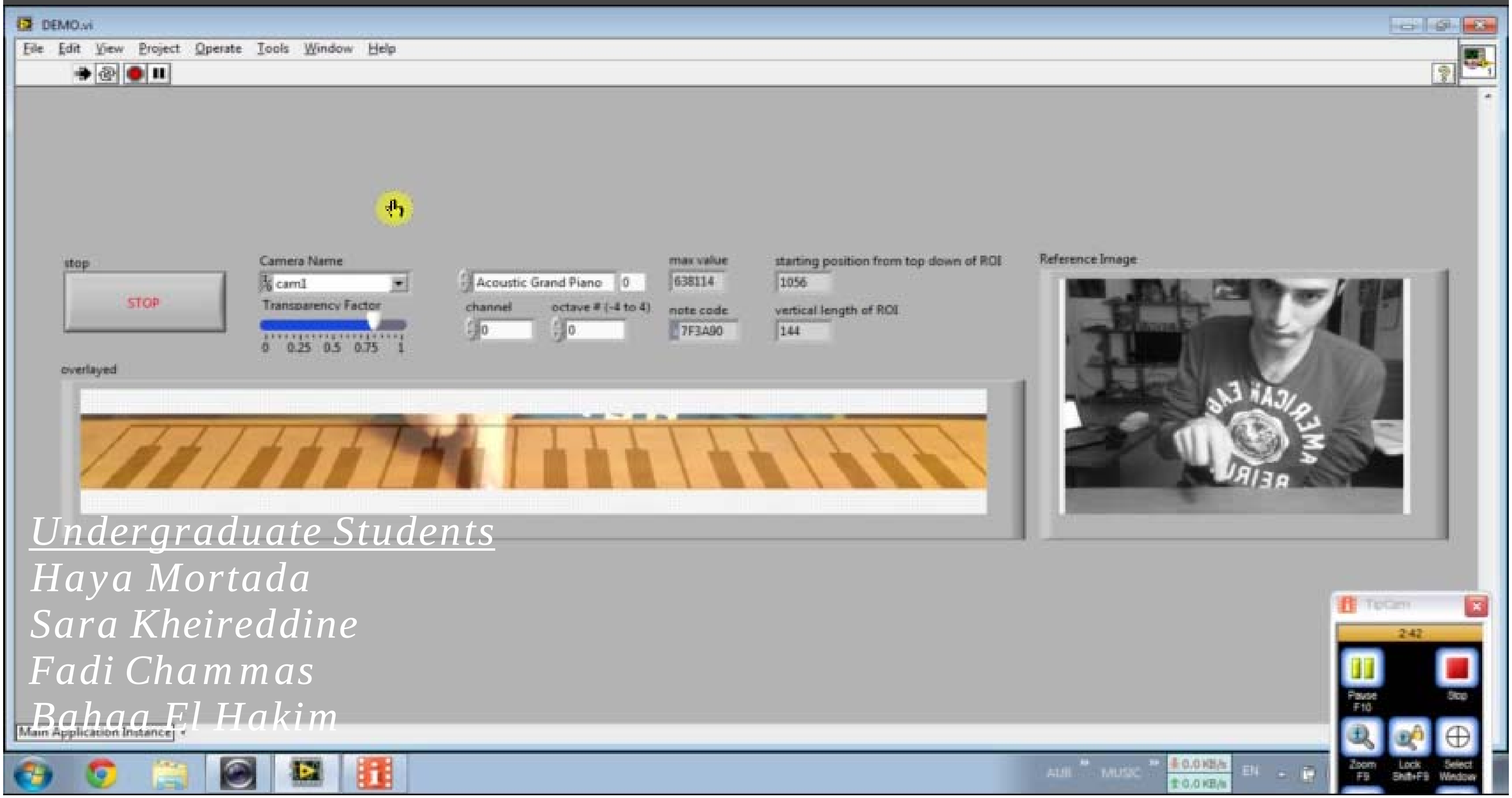

5:39 PM
(Xleft , Ytop)
(Xright , YBottom)


3:33 AM


4:15 PM
Wall Collision
(−Vx0 , Vy0)
(−ax0 , ay0)


5:39 PM
Point 1 Entry
Point 2 Exit
Ray

# Gesture Based Piano App



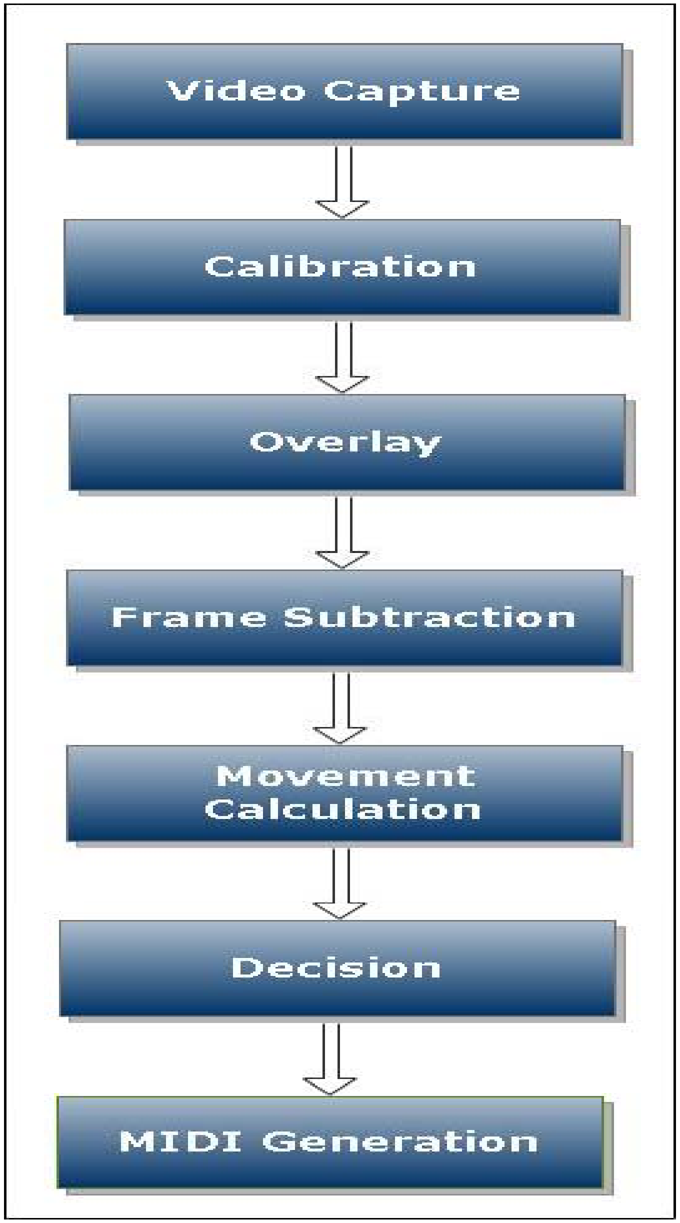

*Undergraduate Students*

*Haya Mortada*

*Sara Kheireddine*

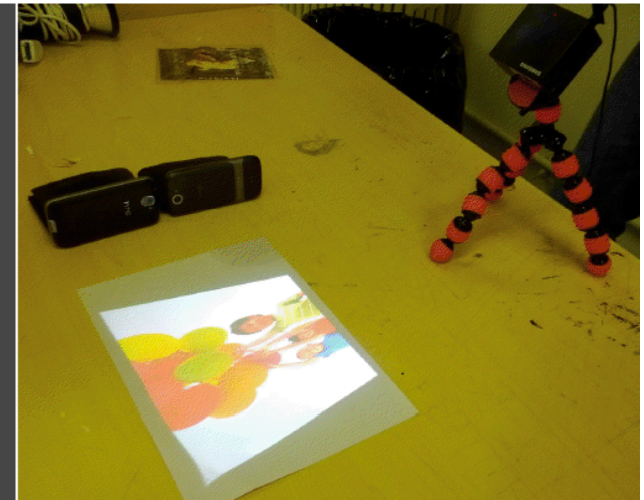
*Fadi Chammas*

*Bahaa El Hakim*



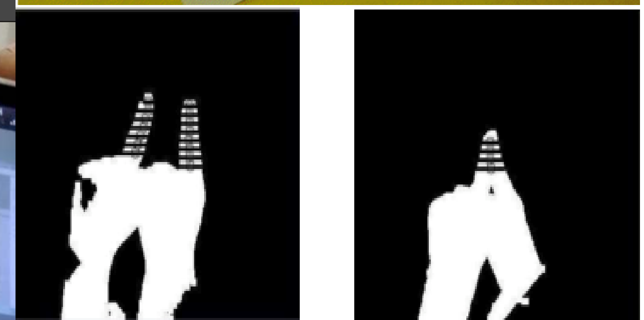

- Frame rate of 3.33 frames per second, for both options of single key and multiple key generation of notes
- Equivalent to 200 frames per minute.
- This frame rate allows for a maximum tempo of 100 beats per minute, assuming the majority of the notes played are half notes.
-Given that moderate pieces are usually played at a tempo of 108 beats per minute, and that most beginner pieces do not use shorter than half notes: playability is OK

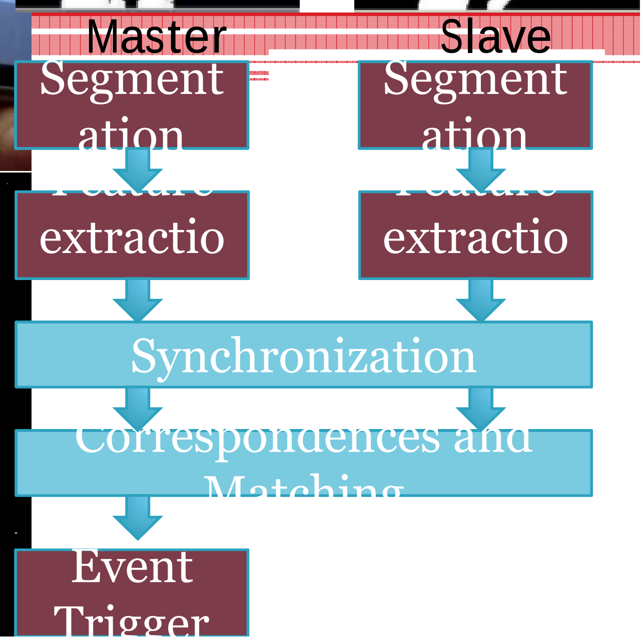
# PicoSpaces: A Mobile Projected Touch Interface for Collaborative Applications



*Prior Art*

*Proposed Solution*





*Undergraduate Stude*

*Marc Farra*

*Maya Kreidieh*

*Mohamed Mehanna*

# Outline

- On Going Projects
- uSee Project
- Biologically Inspired Deep Visual Networks

# uSee: An Energy Aware Sift based Framework for Supervised Visual Search of Man Made Structures

*Graduate Student*
*Ayman El Mobacher*

# Problem Statement

- Proliferation of digital images and videos + the ease of acquisition using smart-phones

  => opportunities for novel visual mining and search

- Energy aware computing trends and a somewhat limited processing capabilities of these handheld devices

- Required to better fit an environment where "green", "mobility", and "on-the-go" are prevailing

- **uSee:** a supervised learning framework using *SIFT* keypoints
  - exploits the physical world
  - delivers context-based services

# Prior Work

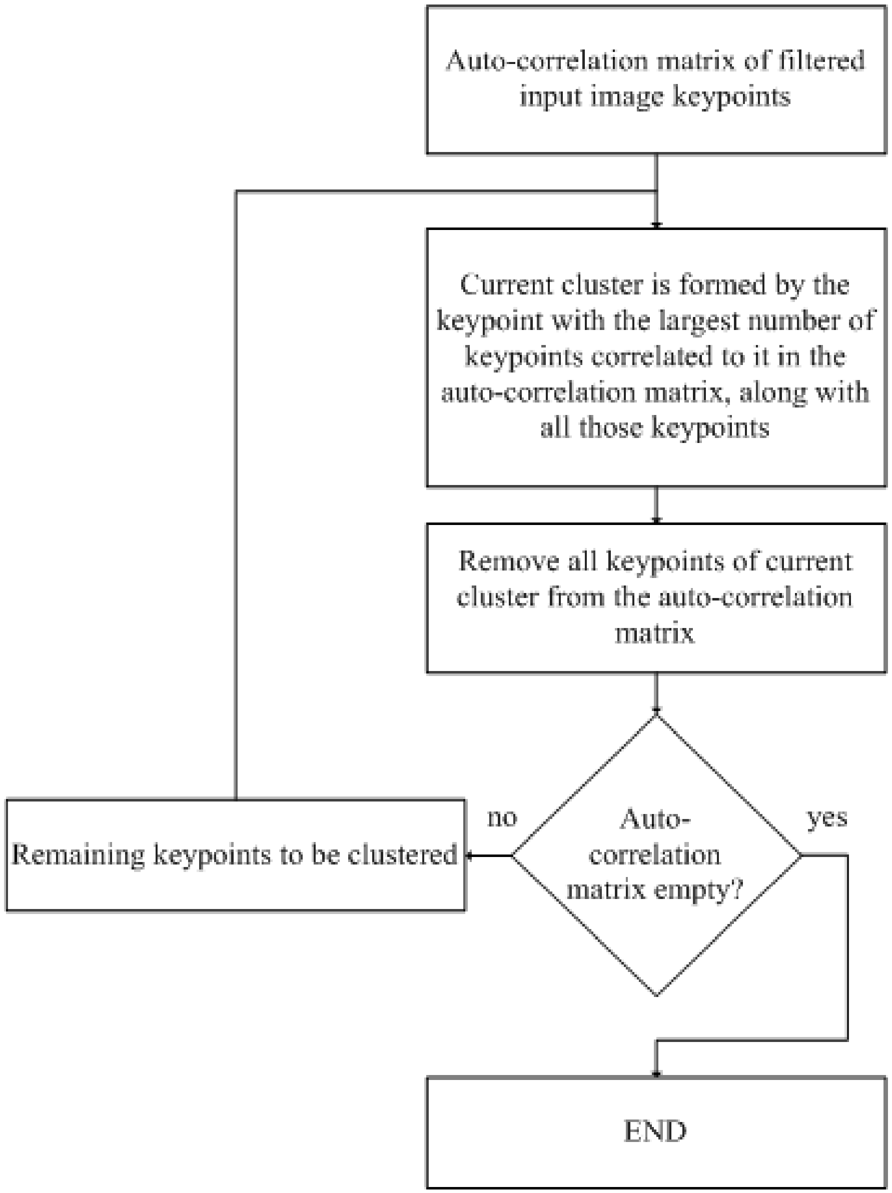
- Visual salient regions and attention model based filtration so that only keypoints within the region of interest are used in the matching process while dropping those in the background [*Zhang et al.*, *Bonaiuto et al.*]
- Consistent line clusters as mid-level features for performing content- based image retrieval (*CBIR*) and utilizing relations among and within the clusters for high level object (buildings) detection [*Shapiro et al.*]
- Causal multi-scale random fields to create a structured vs. non-structured dichotomy using image sections [*Kumar and Hebertin* ]
- Scale invariant descriptors followed by a nearest neighbor search of the database for the best match based on "hyper polyhedron with adaptive threshold" indexing [*Shao et al.*]

# Methodology

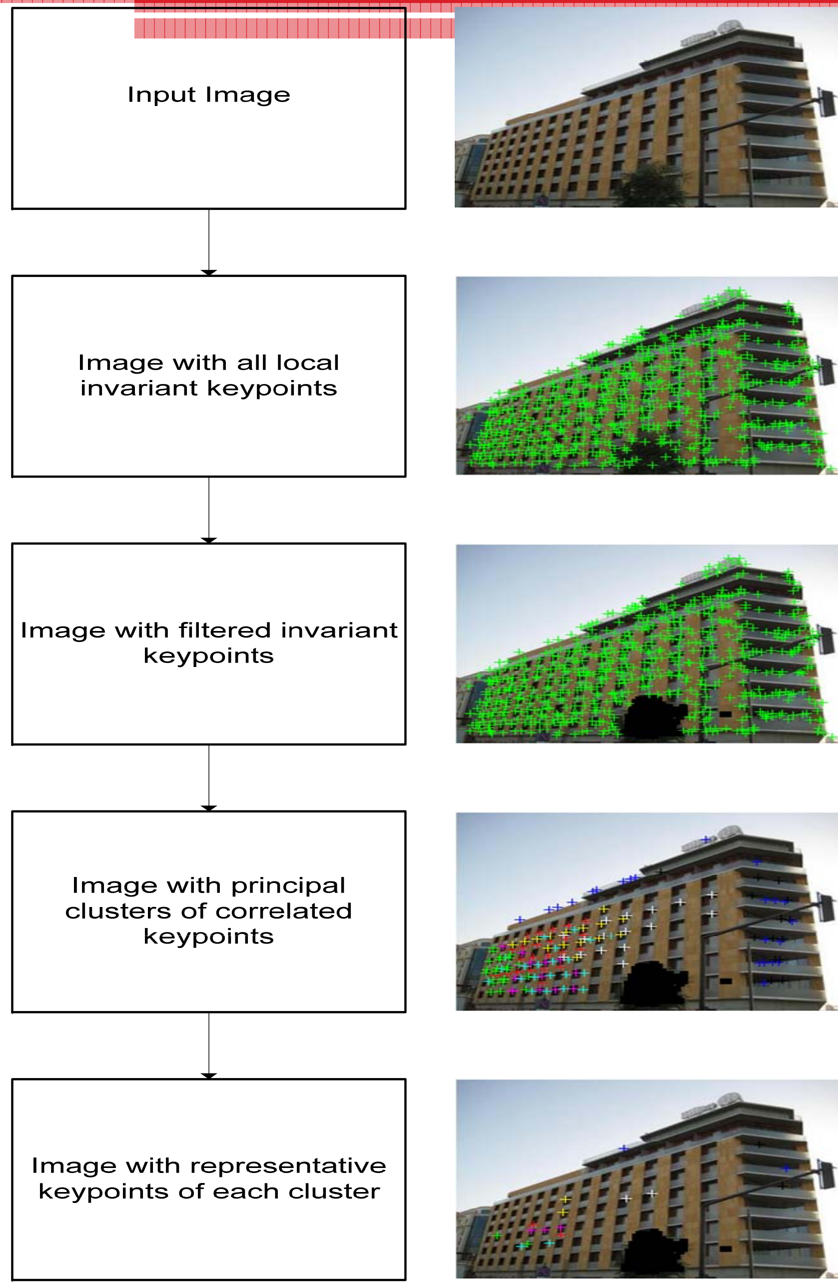

- Implemented as an on demand pull service
- Based on energy aware processing of building images
- **Pre-processing phase:**
  - via cloud ( porting it locally now)
  - highlights the areas with high variation in gradient angle using an entropy-based metric
  - image is divided into 2 clusters: low gradient angle variation vs high gradient angle variation
- **Signature Extraction:**
  - exploits the inherent symmetry and repetitive patterns in man-made structures
  - guarantees an energy aware framework for SIFT keypoints matching
  - *SIFT* keypoints are extracted -> correlated -> clustered based on threshold
  - r (%) *SIFT* keypoints are selected from clusters (r pre-defined for image)

# Preprocessing

# Methodology - Workflow

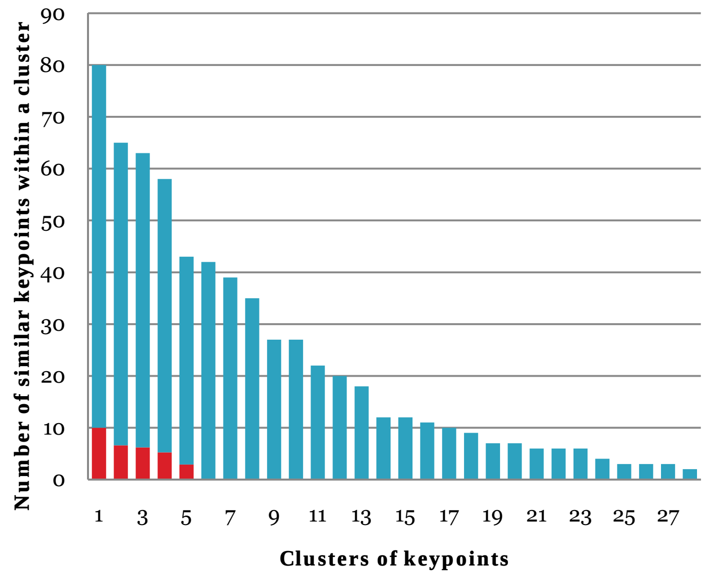



*uSee clustering workflow*




*uSee keypoints selection workflow*

# Signature Extraction




- **Identification:** when new image is acquired
  - Extract signature $C = [k_1 \quad \cdots \quad k_n]^T . [k_1 \quad \cdots \quad k_n]$
  - Compute L2 norm between the query's and all the database's signatures
  - Identification based on a maximum voting scheme

# Validation1

- ***ZuBuD***
  - 201 buildings with 5 reference views and 1 query image for each building
- Several values for r were tested for both the reference and the query images
- Average number of all *SIFT* keypoints in a given image  about 740

# Results1

- Reduction in operational complexity at runtime instead of comparing a new query image $n$ keypoints to $5^{*}n{*}d$ thus perfoming 5*n2*d comparisons, $r$ keypoints where $r<<n$, only $5^{*}r2{*}d$ comparisons are needed.

- With 50 keypoints ($r/n$ = 6.8%), we save 99.54% on computing energy without affecting accuracy results.

- Using 15.5% of SIFT keypoints exceeded all prior results achieved, to the best of our knowledge, on ZuBuD: reached 99.1% accuracy in building recognition.

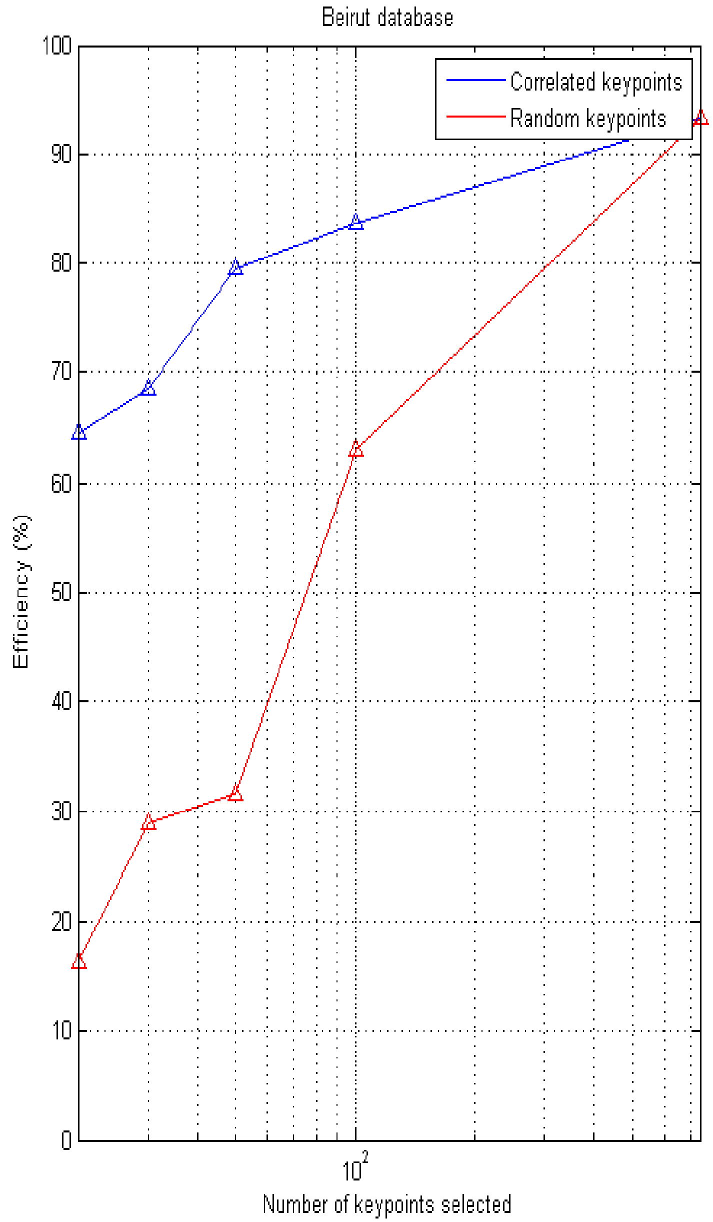
| Method | # of keypoints in reference image | # of keypoints in query image | r/n | Recognition rate |
|---|---|---|---|---|
| [8] | All | All | - | 94.80% |
| [24] | All | All | - | 90.4% (Correct classification)<br>94.8% (Correct match in top 5) |
| [26] | All | All | - | 90.4% (Correct classification)<br>96.5% (Correct match in top 5) |
| [27] | 335 | 335 | 45.3% | 96.50% |
| uSee | 20 | 20 | 2.7% | 91.30% |
| | 30 | 30 | 4.1% | 94.80% |
| | 40 | 40 | 5.4% | 95.70% |
| | 50 | 50 | 6.8% | 96.50% |
| | 100 | 75 | 10.1% | 98.30% |
| | 100 | 115 | 15.5% | 99.10% |

# Validation 2

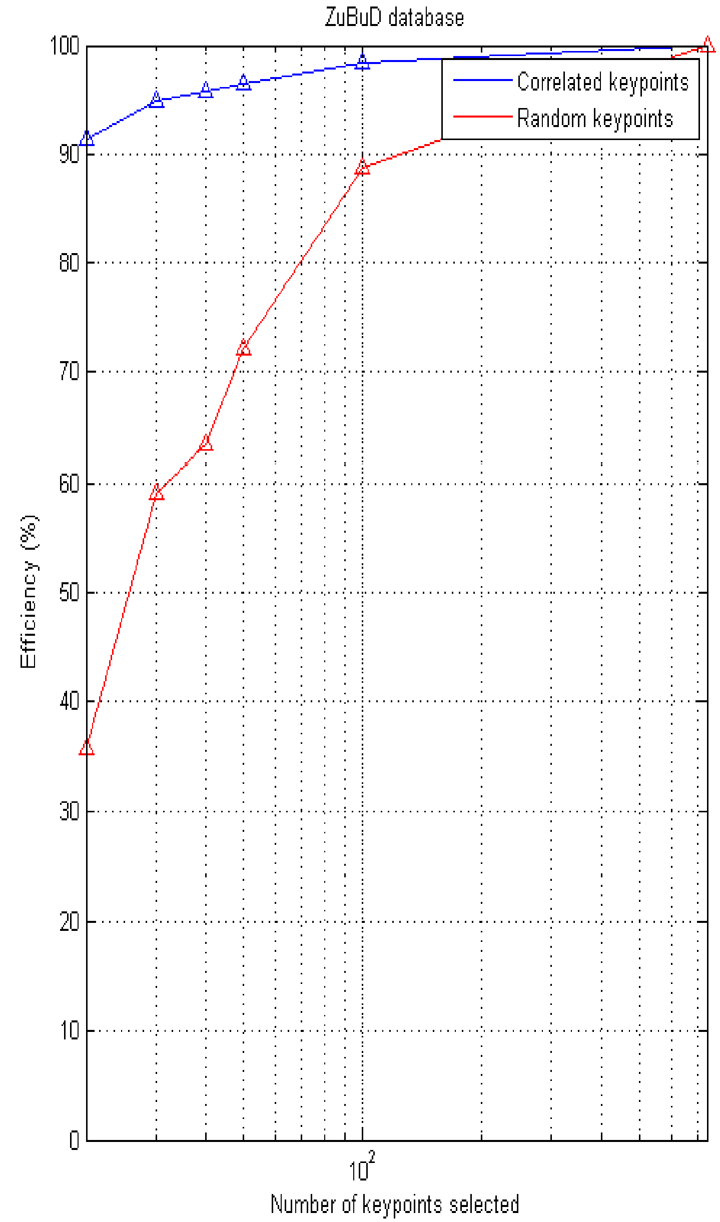
- Further tests conducted on home grown database of buildings from the city of Beirut (**Beirut Database**)
  - 5 reference images taken at the same time of day
  - 1 query image at different times and weather conditions
  - total of 38 buildings



The test images in Beirut db are different from their corresponding reference in illumination, camera angle, and scale which are major image processing challenges not present in the *ZuBuD* db



Beirut database
Correlated keypoints
Random keypoints
Efficiency (%)
Number of keypoints selected

ZuBuD database
Correlated keypoints
Random keypoints
Efficiency (%)
Number of keypoints selected

# Outline

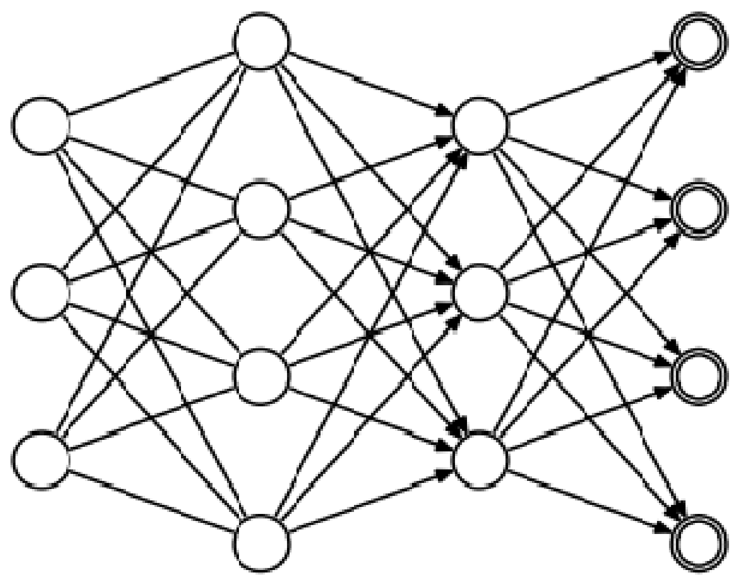
- On Going Projects
- uSee Project
- Biologically Inspired Deep Visual Networks

# Biologically Inspired Deep Networks for Visual Identification

Graduate Student
L'emir Salim Chehab

# Deep Belief Networks

- Deep belief networks: probabilistic generative models composed of multiple layers of stochastic variables (Boltzmann Machines)
- First two layers an undirected bipartite graph (bidirectional connection). The rest of the connections are feedforward unidirectional



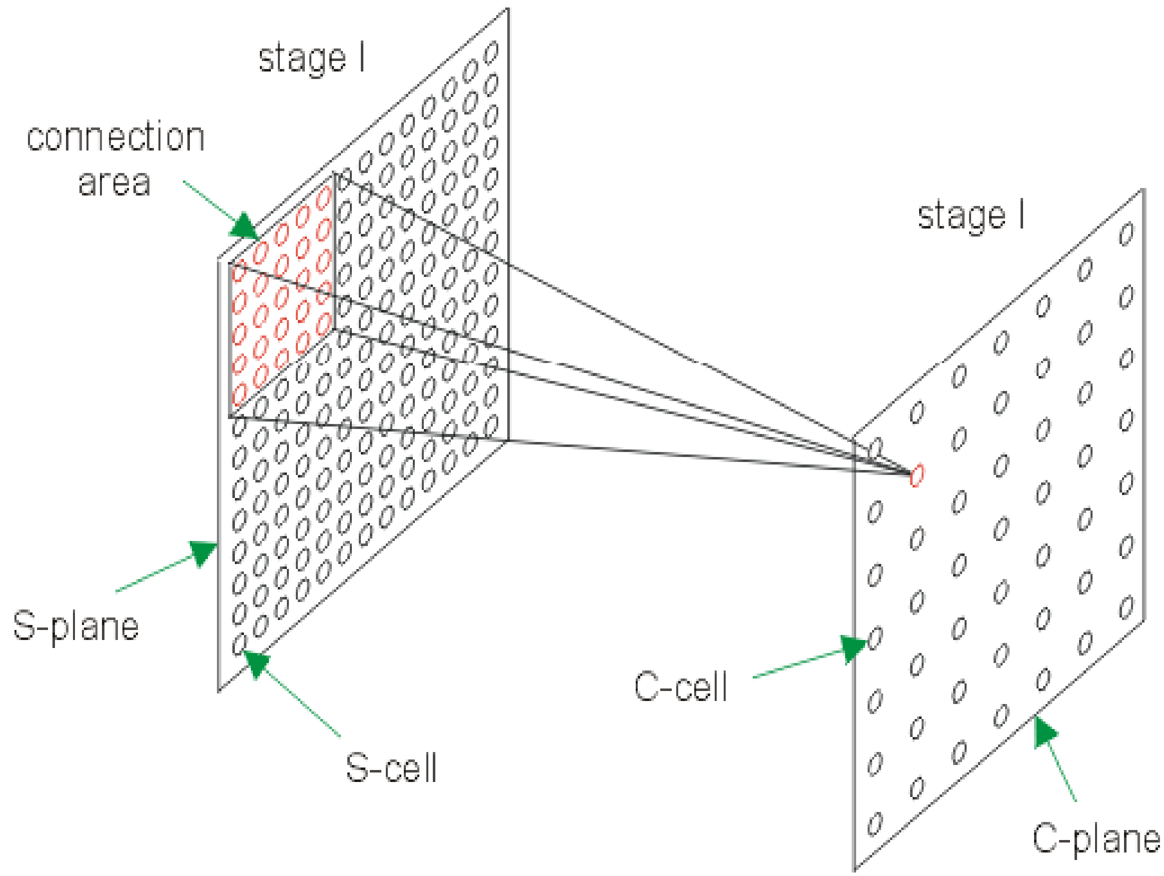
Deep Belief Network

# Fukushima Work

- Four stages of alternating S-cells and C-cells in addition to inhibitory surround from S-cells to C-cells and a contrast extracting layer

- S-cells equivalent to simple cells in primary visual cortex and responsible for feature extracting

- C-sells allows for positional errors

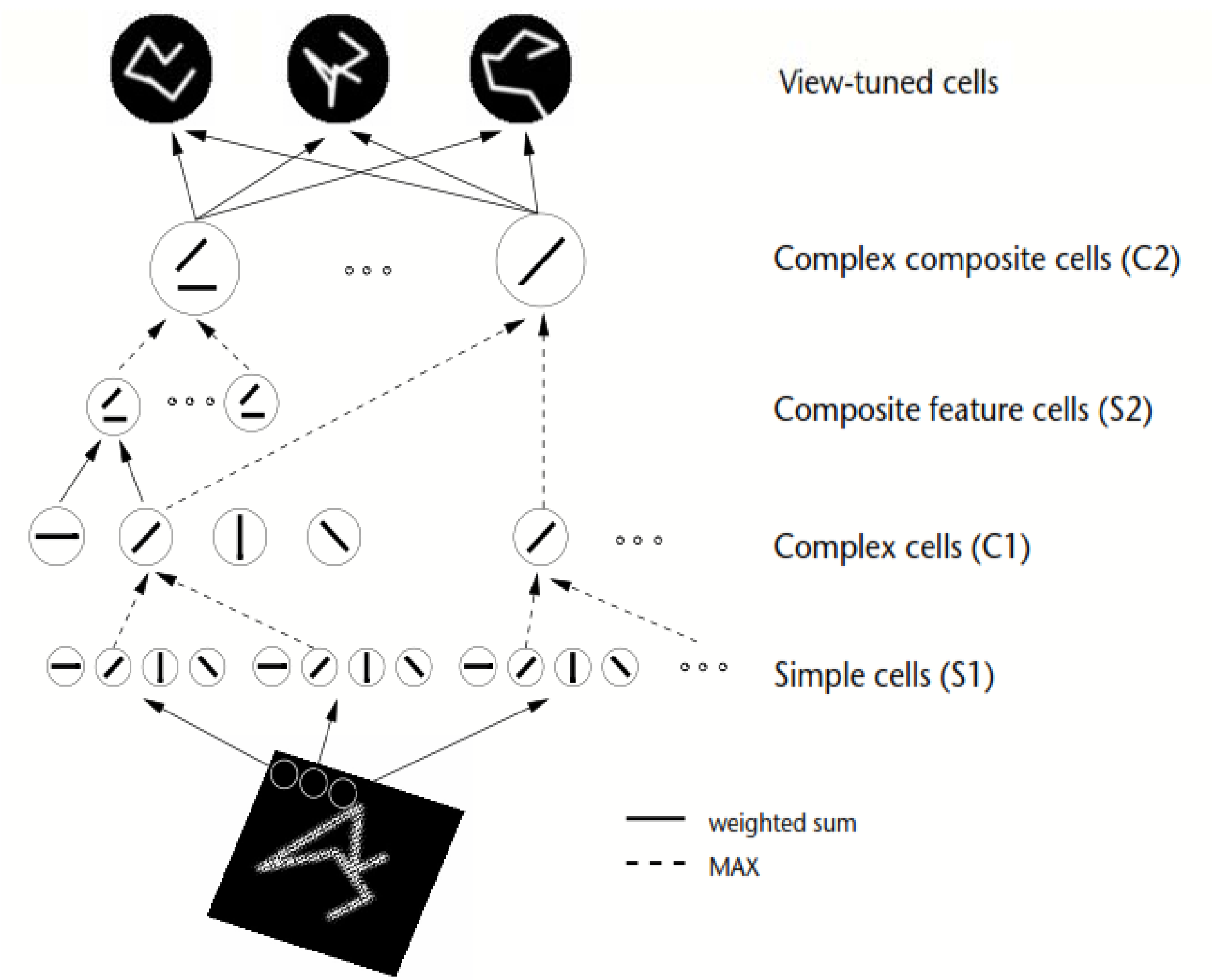



Connection of S-cells and C-cells

# Riesenhuber Prior Work

Hierarchical feedforward architecture composed of 2 stages and 2 operations:

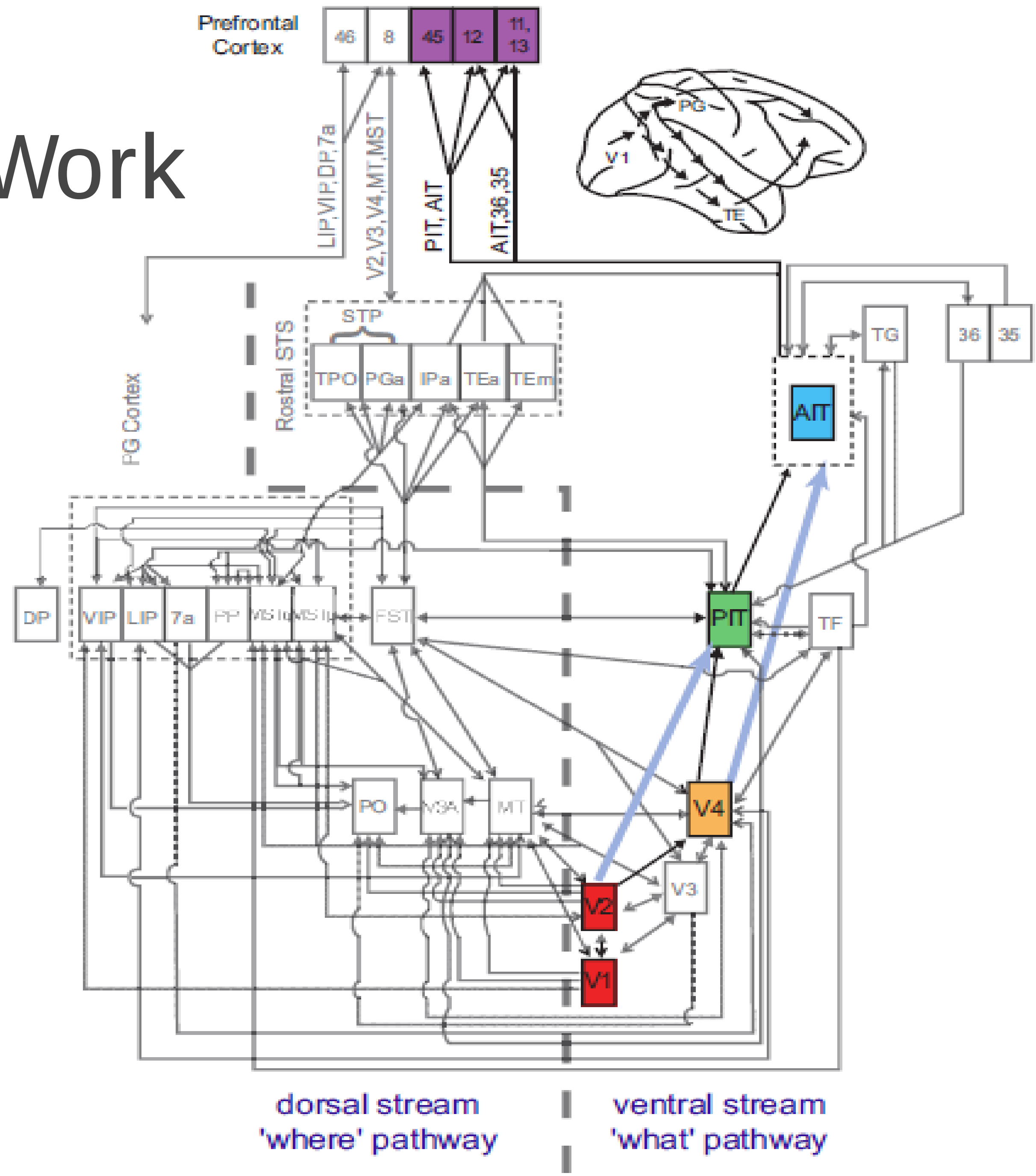
- weighted linear summation
- nonlinear maximum operation




Riesenhuber et al. simple feedforward network

# Poggio Prior Work

- Battery of Gabor filters  applied to obtain $S_1$ stage's response
- $S_1$ proceeds to $C_1$ and layers alternate between S-cells and C-cells
- A total of $10^8 – 10^9$ neurons used
- Accounts mainly for ventral stream part of the visual cortex




Poggio et al. feedforward model

# Methodology

- Includes the “Photoreceptor layer” before the first layer

- 8 main layers: every two consecutive layers represent a stage of one of the four visual regions

V1 - V2 Layers:

- Consecutive S and C-cells in each layer
- Bidirectional connections similar to a deep belief network
- Supervised training

V4 - IT Layers:

- Unidirectional feedforward connections leading to the output
- Unsupervised learning

# Proposed Model

- Network structure: Input – photoreceptor layer – 1000 – 1000 – 500 – 500 – 200 – 200 – 100 – 100 – output
- The cross-entropy error as a cost function
- Training using Hinton's algorithm

# Results

- *MIT-CBCL Street Scene Database:* 3547 images. 9 object categories = [cars, pedestrians, bicycles, buildings, trees, skies, roads, sidewalks, stores].
- Data set is split into 70% training and 30% testing.
- Results are based on the average of 15 runs.
- 90% correct classification rate versus an 88% achieved in Poggio's model on the same data set.

# Thanks for Your Time

**Contact info**
**mariette.awad@aub.edu.lb**