An Application Framework for Implicit Human-Centered Tagging Using Attributed Affect

K. C. Apostolakis, P. Daras
Information Technologies Institute
Centre for Research & Technology – Hellas
Thessaloniki, Greece
kapostol@iti.gr, daras@iti.gr

Implicit Human-Centered Tagging (IHCT)

- Attempts to obtain user behavioural response for tagging purposes
  - Effectively reduces user effort in contrast to explicit (textual) annotation
- Challenges (Pantic & Vinciarelli, 2009)
  - Effort to include observed user reactions & behaviour, as well as implicit tags to the data tagging & retrieval loop
  - Develop behaviour analyzers that can attain accurate and reliable results even on audiovisual sensors built in commercial computers

Psychological Framework (Russell, 2003)

- Core Affect
  - *2-D space defined by two components*
    - ***Valence*** – Amount of pleasure experienced at any given moment
    - ***Arousal*** – Activation level in preparation for action
    - e.g. feeling *delighted*, *bored*, etc.
- Perception of Affective Quality
  - *To perceive stimuli in terms of their emotional properties*
    - e.g. *delicious* meal, *boring* lecture etc.

Psychological Framework (Russell, 2003)

Attributed Affect

- Subconscious attempt to attribute change in Core Affect to its perceived cause
- The stimulus that is identified as the cause becomes **the "Object"**
  - ***Attention*** *is shifted towards* ***the "Object"***
  - ***Behaviour*** *is directed at* ***the "Object"***
- Defines Emotional Awareness
  - *Main route to the affective quality of the stimulus*

Attributed Affect for IHCT

- Obtain user *affective response* (Core Affect)
- Obtain *the "Object"* via gaze information
  - *Identify specific stimulus depicted in the image, where the users have focused their attention on*
- Attribute *affective response* to *the "Object"*
- Image annotated with appropriate *affective tag*
- A new image containing *the "Object"* is automatically annotated with the assoc. label

Attributed Affect for IHCT

Advantages (1/4)

- Automatic annotation of large portions of the image database by looking at a single image
  - *User looks at image depicting a spider and experiences a **jittery** reaction*
  - *Spider identified & attributed as the cause → spider considered **jittery** by the user*
  - *Framework annotates all images in the collection depicting spiders with the **jittery** affective label*
    - User most likely to experience the same reaction when presented with the same stimuli

Attributed Affect for IHCT

Advantages (2/4)

- Retrieval & Recommendation readily available through annotated stimulus
  - *User looks at images of cars, trying to locate models that spark his interest*
  - *Several cars are identified as causing feelings of satisfaction – others are dismissed*
  - *Annotation of all images in the collection, depicting either dismissing or pleasing stimuli*
  - *Retrieval of images that were annotated as '**pleasing**'*

Attributed Affect for IHCT

Advantages (3/4)

- Annotation based on user personal experience
  - Users can annotate content specifically to their preferences
    - Not all spiders are considered *jittery* by all people!
  - Several culture-dependant points addressed
    - Something funny here might be considered offensive somewhere else
  - Personalized recommendation of like-valenced content
    - Horror movies scare me! → don't show me Horror movies!
    - I love Horror movies! → Show me more!

Attributed Affect for IHCT

Advantages (4/4)

- IHCT based on Attributed Affect can be applied to multitude of setups, as long as there's a means to obtain gaze and affect information
  - *Many methods for obtaining user affective response*
    - Facial Expressions, Blood Pressure, Body Temperature, EMC, etc.
  - *Many methods for obtaining gaze information*
    - Single Image, Stereo, Special apparatus (eyeglasses)
  - *Most affordable setup: Commercially available computer systems with a single low-res camera*
    - Today's Laptop computers!

The Framework

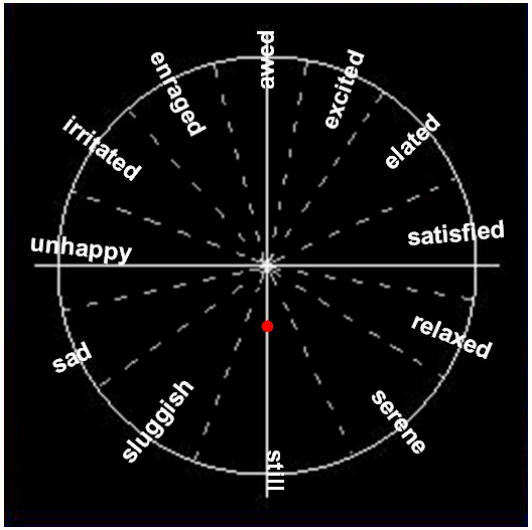
Obtaining User Affective Response (1/2)

- Affective Response obtained via Facial Expression Analysis
  - *Using Single RGB camera*
- Identifying AU activation
  - *Track key facial features corresponding to AU muscle groups*
- Active Shape Model (ASM, Cootes & Taylor 1995)
  - *Statistical model describing the shape of an object*
  - *Capable of deforming to fit to a new instance of the object*

The Framework

Obtaining User Affective Response (2/2)

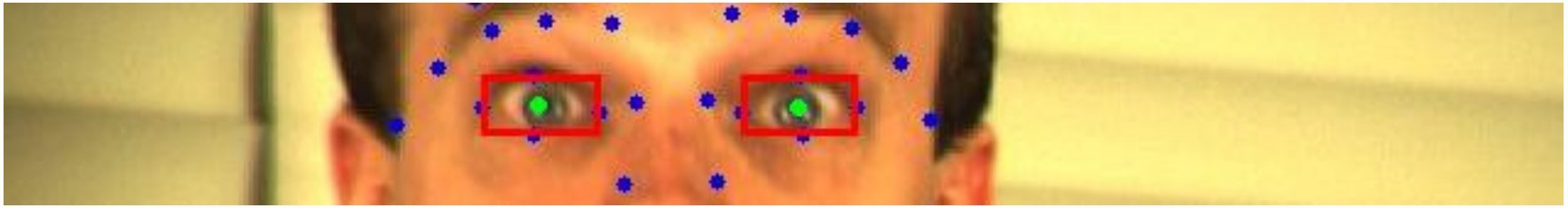
- Procedure:
  - *Take snapshot of "neutral" expression*
    - Fit ASM and save "neutral" landmark positions
    - Calculate landmark distances from the eye line
  - *For every consequent frame:*
    - Fit ASM
    - Calculate landmark distances from the eye line
    - Calculate AU intensity from the distance differences
    - Infer Valence-Arousal pair corresponding to the AUs extracted
    - Extraction of affective label via Yik et al (2011)



The Framework

Gaze Tracking (1/2)

- Single Image Gaze Tracking & Gaze Point Est.
  - *Locate the iris centre (pupil) P & eye corners E1, E2*
  - *Map current information on P, E1, E2 to 2D screen coordinates*
    - Requires calibration step
- Eye corners located via ASM (27, 29, 32, 34)
  - *Generation of Pupil Search Area (PSA)*
  - *Pupil is certain to be contained within PSA*



The Framework

Gaze Tracking (2/2)

- Locate Pupil via Automatic Adaptive Thresholding
- Gaze Point Estimation via Linear 2D Mapping
  - *Calibration*
    - Calibration points displayed on the screen to collect info
    - Users fixate their gaze on each calibration point
  - *Linear 2D Mapping*
    - Pupil centre positions P_i (x_i, y_i) stored for each calibration point K_i (α_i, β_i) during calibration
    - Minimum of 2 calibration points K_1 (α_1, β_1), K_2 (α_2, β_2)
    - Every subsequent P' (x', y') mapped to screen coordinates (α', β')

The Framework

Identifying the "Object" (1/2)

- "Object" identified via image segmentation
- Segmentation algorithms
  - *Require explicit designation of foreground – background pixel seeds*
  - *Even more difficult to unobtrusively implement using input obtained via eye tracker (Sadeghi et al, 2009)*
    - User shouldn't need to bother with explicit fg/bg designation
    - User should look at the object depicted in the image
    - The segmentation algorithm should take over the rest
  - *GrabCut Segmentation*

The Framework

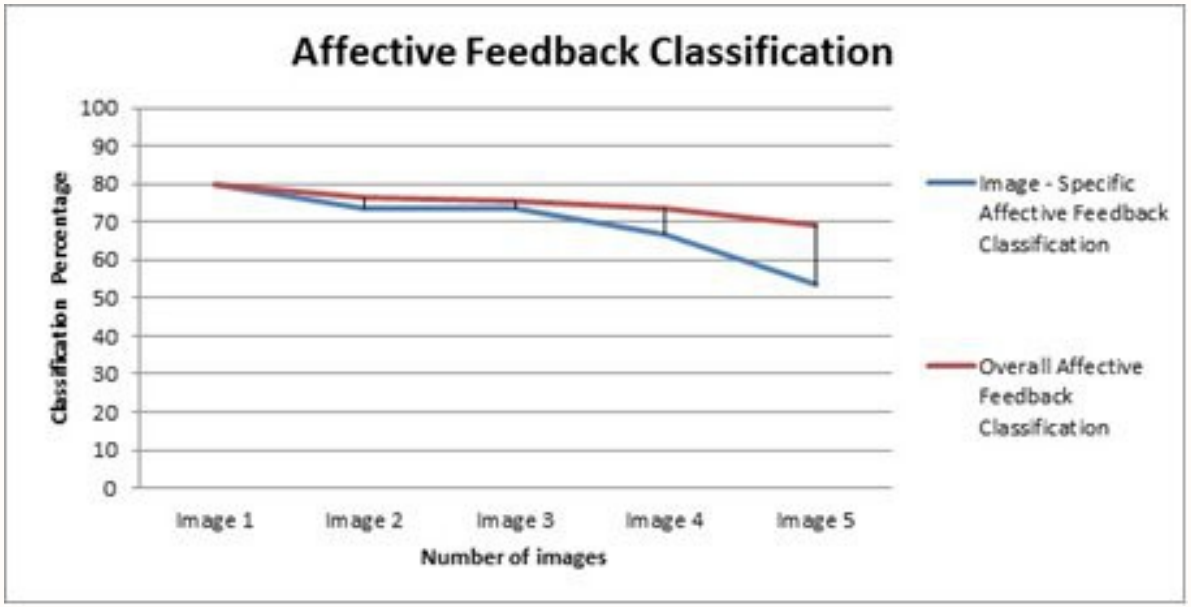
Identifying the "Object" (2/2)

- Application to Framework
  - *When the user's gaze point on screen is found to intersect one of the images displayed, a ROI is automatically generated around it*
    - Ensures the unhindered process of image annotation
- Shortcomings & Improvements
  - *Excessive or incomplete segmentations when object is non-convex or not entirely contained within ROI*
    - Solution: modify ROI width & height (mouse wheel)
    - Modified GrabCut versions (Chen et al, 2008)

Experimental Results

Topical Relevance – Affective Feedback Classification

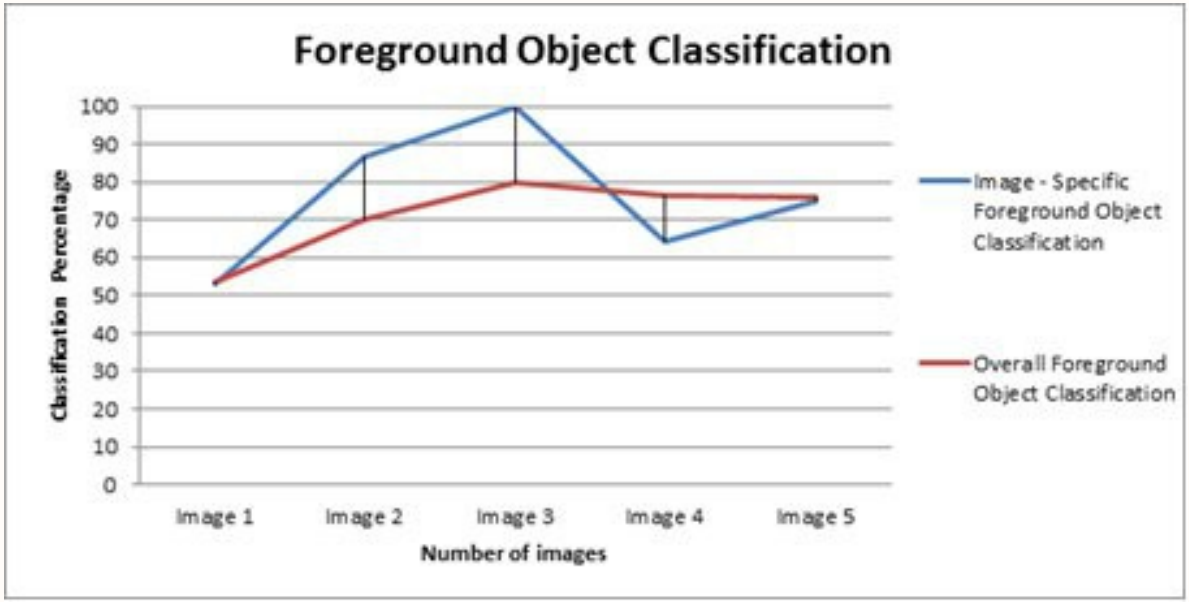
- 15 participants
- Results show the framework achieves an approximate 70% correct affective feedback classification performance



Experimental Results

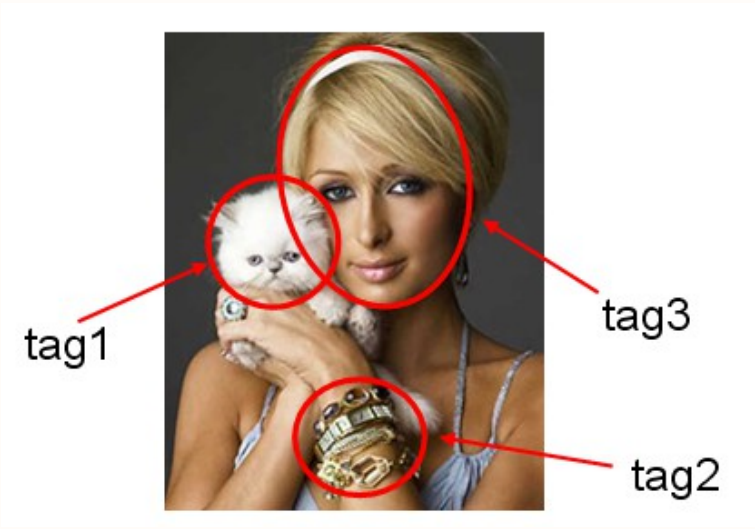
Direct Distribution of Tags to Image Categories

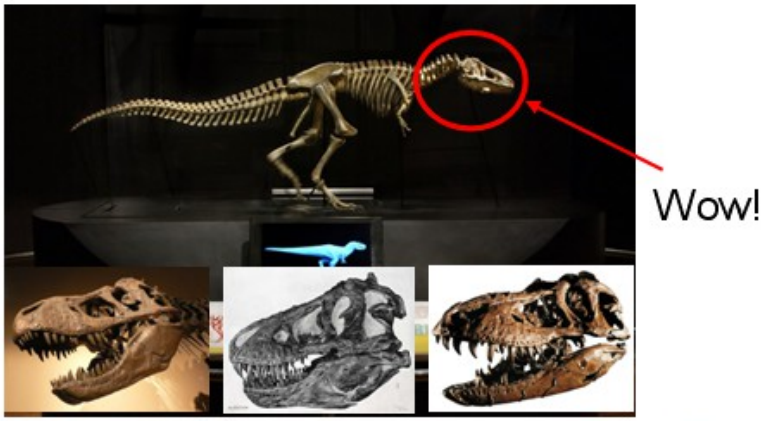
- 95% of the images undergone segmentation were classified to one of the 5 available categories (BoF)
- Overall classification performance approximately reaches 76%



Future Endeavours

- Improvements
- Framework applicability
  - *Content-based Recommender Systems*
  - *Tagging & Retrieval on Complex Image Scenes*
  - *Object recognition and display in immersive 3D environments*





Thank You!

Questions?