Facial Expression Tracking
For
Remote Conferencing Applications

An approach to tackling the Disadvantages of Remote worker Multitasking

Researcher Name: Paul Smith

Emotions and Communication

Forms of communication

- **Vocal communication** and language
  - Words, written or spoken

- **Non-verbal communication**
  - Visual communication such as gestures and facial expressions
Non-verbal communication

- Around 70-80% of a message’s emotional meaning is communicated through non-verbal clues such as facial expressions and gestures.

- Only a relatively small amount is taken from what the person actually says.

- Emotion is what gives communication life – without understanding it the message loses meaning and becomes ambiguous.

Distributed or Remote Teams

- In face-to-face work, discussion and negotiation relies strongly on non-verbal feedback.
  - Participants are continuously aware of their partner’s gestures and emotional signals.
  - This allows them to respond to peoples reactions to information sent or received.

- Continued rise of virtual teams, means that collaborative work increasingly requires tools to manage and support distributed and remote work.

- Large amounts of companies use distributed workers based in many different countries and locations around the world.
Distributed or Remote Teams cont…

- Eg. Intel (70%), IBM (40%) and Sun Micro Systems (nearly 50%) already have high percentages of virtual and distributed workers
- Remote workers can be based at home or abroad or travel frequently
- Virtual conferencing is the most convenient and economical means of communication with colleagues and clients
- It is predicted that by 2012, 40% of the USA’s workforce will qualify as distributed and the rest of the world will not be far behind

What happens in conversation when you can’t see other participants?

What if you’re not paying attention or multitasking?
Multitasking in Remote meetings

- A well known problem in this area is the effect of **multitasking** on a user’s attention, performance and awareness in a virtual meeting.

- During Virtual meetings, participants often **multiple tasks**:
  - checking emails, browsing the internet or performing other work or non-work activities.

- People multitasking during a conference call miss:
  - Certain **information points** due to lack of attention,
  - Important **non verbal information** which can provide valuable insight into participants reactions to proposed information and discussed topics.

Research Project Goals:

- Research, develop and experimentally assess **proof-of-concept AI-Based software agents** for monitoring the **presence** and emotional state of participants in virtual meetings.

**Project Requirements:**

- Tracking **important emotional states** of user and prompt other participants of user’s state.

- Must be capable of being **integrated** into a remote conferencing application.

- Designed to compliment already available **communication systems** with an aim to **reduce** the negative effect of **multitasking** of personnel during remote meetings.
Main Research Milestones:

1) Creating a prototype real-time facial expression tracking system capable of tracking user’s facial movements.

2) Devise experimental procedures for evaluating developed tracking system

3) Create a recognition system for emotion tracking software agent (E.T.A.) – effectively the agent's brain or logic engine

4) Integrate developed tracking system into a test msn style environment for proof of concept

1) real-time facial expression tracking system

The Research for this step of the project required research and development in the following areas:

- Available and affordable technology capable of providing robust motion tracking
- Previous research conducted into facial expression synthesis
- Available open source libraries or SDK’s applicable to research project.
2) Evaluation Methods

In order to assess the usability of our tracking system we:

- Developed an application to test the following elements of our system:
  - How well the tracking worked in a highly graphical application as a control method
  - Evaluate the competency of the system in capturing/recording facial expressions

3) Software Agent Emotion Recognition System (E.T.A.)

The Research for this step of the project required research and development in the following areas:

- Neural networks for pattern recognition
- Representation methods for input and training data
4) Integration of Agent into Remote meeting environment

The Research for this step of the project required research and development in the following areas:

- Existing Collaborative Environments
- Collaborative work in Virtual Spaces
- Interactions in virtual places
- Interaction limitations in virtual environments

Fundamental Emotions

- What emotions should be tracked?
- Dr Paul Ekman carried out extensive research and devised a list of basic/fundamental emotions
  - Anger
  - Fear
  - Happiness
  - Sadness
  - Surprise
  - Disgust
Modelling these Emotions

- Ekman developed a list of the key facial movements associated with these emotions.
- These 6 emotions have specific facial movements which are universal across all cultures.
- We used this work in our Emotion Tracking Agent to train our system to recognise these important emotions.

The 6 Universal Emotions

ANGER  DISGUST  HAPPINESS

FEAR  SADNESS  SURPRISE
Solving the Multitasking problem

- Our ETA is designed to remove some of these unwanted side effects of multitasking
- Keeping track of user’s non-verbal responses and emotional states
- Our ETA will prompt participants of other user’s reactions and thus:
  - improve their attention and awareness,
  - reduce the amount of information lost and
  - ultimately improve work performance

Technology Used

Our prototype tracking system consists of:

- An advanced Optical motion capture camera system known as Optitrack
- A single camera was used called an Optitrack Flex C120
- Highly reflective dots (markers) were placed on a user’s face to facilitate tracking of facial movements
Optitrack Flex C120 Camera

- **Optical Markers** are highly reflective hemispherical dots attached using spirit gum.
- Camera is specially configured to filter all light from normal surfaces.
- Markers are illuminated by 12 infrared LEDs.

Marker Setup for Facial tracking

- 9 markers placed in calculated positions on a user's face:
  - 4 on corners of each eyebrow
  - 1 on bridge of nose
  - 4 on corners of mouth and upper and lower jaw
- Designed to capture important facial movements in regions specified by Dr. Ekman's research.
Application Development

- Camera system came with a detailed SDK written in C++ to allow developers to control all its advanced features.

- Using this SDK we developed an application written in C++ to track all 9 markers and return their x and y coordinates.

- The number of markers being tracked was programmed to display on the camera’s LED number display to allow users to ensure that all markers were in view.

Markers In 2D coordinate space

- Camera view in 2d space of the facial markers.

- Origin is on top left.

- Markers are tracked from left to right, top to bottom.

- 9 markers x,y position are recorded to a text file for future use.
Tracking performance Tests

- Tests were devised to ensure robustness of the tracking system and it’s usability.
- A application was developed to test the tracking system using the 3d Game engine Torque as it’s development platform.
- Idea was to gauge how well the tracking application worked in real-time and to ensure data recorded captures enough information for our system.

Torque Game Engine

Why Torque?
- Developed by Garage Games
- Supports all modern game engine rendering features and is written in C++
- Animation system is robust and uses a biped skeletal system to allow detailed control over character animation.
- My research group has already developed a Virtual environment called VRCGroups which uses Torque as it’s development platform.
Test Setup

- First an extensive study of the game engine’s animation system and SDK was carried out

- Torque Game Engine uses a *biped skeletal system*:
  - Allows developer to build a bone structure to control models behaviour – similar to *human skeleton*

- Our test environment consisted of a *Graphical User interface* displaying a 3 Dimensional Head-model

- Head Model is rigged with a detailed *Biped bone structure* to allow full facial muscle control

Bones and Vertex Weights on Avatar Head

- **Biped Bone structure** allowed us to have full control over the movement of our Head-Models **facial Mesh**
Test Procedure

- 15 people were shown 18 different recorded emotions played from a list on the 3D head Model
- Each emotion was recorded using our face tracking application
- Participants were given a choice of the 6 emotions proposed by Dr. Ekman
- Experiment results were compared to recognition rates of tests carried out using emotions recorded on video

Test Results

- The results of this experiment showed high recognition rates for all emotions.
- Our comparison test using recorded video averaged at 80% recognition while the recorded data played back on the Head model averaged at around 81%
- This indicated that our system was sufficient for our needs and captured an adequate amount of emotional data to allow us to carry on to our next stage
Recognition System

- We chose to use a basic feed-forward neural network for our Agents recognition system.
- This allowed us to let our Agent learn from example data sets given to it as training.
- Since a neural net is a simplified model of neural processing in the brain it makes a good system for comparison with the emotion accuracy of a human – our ideal accuracy level.
- The neural network needs to be at least as accurate as the average human in order to be successful at combating the limitations caused by multitasking.

Designing the Neural network

- The basic design of our Neural network is:
  - Takes the coordinates of each marker as inputs
  - Normalises them to values between 0-1
  - Trains the network with pre-recorded emotion data
  - Return 6 outputs with a value 0-1
  - Each output represents one of the 6 fundamental emotions

![Diagram of neural network with inputs, hidden layer, and outputs](image)
Training our network

- Network is trained on groups of recorded marker coordinates which represent the facial muscle movements described in Dr. Paul Ekman’s research.

- For each group of training coordinates given to the network a value of 1 is given to the corresponding emotion node in the output layer.

- All other nodes are given a 0 value.

- A total of 240 training groups were used, 40 per emotion.

Network Inputs

- The biggest hurdles for this stage of research were:
  - Developing methods for representing our recorded coordinate data.
  - Network input schema needed to allow our network to effectively recognize similarities in facial movements between each emotion.

- We developed and tested 4 Network Input schema against each other using a cluster comparison.
  - This comparison showed how similar each group of data was to each other.

- Ideally all similar emotions should cluster together in Euclidean space to ensure that the neural network does not get confused.
Candidate Network Input Schema

- **Method 1:**
  - Converted the y coordinates of point to a fraction of the camera’s max y value 290, i.e., divided each point by 290

- **Method 2:**
  - Calculated the y distance between each point and marker 5, which is the marker on the bridge of the user’s nose and divided resulting value by 290

- **Method 3:**
  - This method calculated the max and min y value for each marker from the training data and used this to convert the distance from method 2 to a value between 0 and 1

- **Method 4:**
  - This method used the x and y coordinates of each marker to calculate specific angles between points and used the max and min angles calculated using the training data to convert each input to a value between 0 and 1

Best Input Schema

- **Method 4** proved to be best in our cluster comparison tests

- Initially this method used **7 different angles** to describe each input

- In order to determine the ideal number of angles to use, Variations of this method were developed using:
  - 9 angles
  - 11 angles
  - 13 angles
  - And 15 angles
  
  All 15 considered angles
Clustering of Normalised Training data

- Our **Angle Input method** generated the following **Dendrogram** during our cluster analysis tests on our training data.
- In the diagram each emotion is represented by one of 6 colors.
- The emotions clustered well enough together to warrant testing in our network.

![](image)

- Anger is , Disgust is , Sadness is , Happiness is , Surprise is , Fear is 

Neural Network Performance

- The next step of the project was to test our neural network’s on the following aspects:
  - Emotion recognition accuracy on data from the same recording session as our trainings data using our preferred input method.
  - Emotion recognition accuracy on newly recorded data from 3 separate recording sessions.

- Both tests used randomly chosen coordinate groups from our recorded data files and compared calculated outputs from the network with expected outputs.
Performance Test results

- The initial test using data from the same sessions as our trained data resulted in 100% accuracy as expected.

- The second test resulted in a 88% recognition accuracy and in order to evaluate this further another comparison test was developed:
  - A survey was carried out with 15 participants.
  - Each person viewed 18 video recordings of a person displaying an emotion on their face and chose which of 6 emotions they thought it represented.
  - When the training data for our network was recorded we also recorded the same session on video for this experiment.
  - Therefore this comparison should give an accurate indication to how effective our neural network really is.

Performance Test results cont...

- The survey averaged 80% accuracy over the 15 participants.

- Our result of 88% accuracy clearly indicates that our network is not only competent, but may even be more accurate than that of humans.

- The next step in the project was to implement our Agent architecture into a working collaborative virtual environment.
Integrating our ETA in a CVE

- The CVE we chose for this stage of the project was **VRCGroups** – a virtual environment already developed by our research group

- This stage required the following:
  - Integration of the Optitrack SDK into the Torque Game Engine
  - Integration of our neural network into the Torque Game Engine
  - The creation of a test environment within the VRCGroups CVE

---

Project Status

- We have completed our Integration of both the Optitrack SDK and our ETA Neural network architecture into the Torque Game Engine

- Currently we are finishing development on our test CVE within the VRCGroups environment

- The final steps will be to evaluate our ETA in Real-time
The End

- Any suggestions or comments?
- Particularly on Methodology…