Facial Recognition and Emotion Detection In Environmental Installation and Social Media Applications

Pensyl William Russell, Min, Xiaoping, Song Shuli Lily
Northeastern University
r.pensyl@neu.edu
min.x@husky.neu.edu

Synonyms

Image Processing, Facial Recognition, Emotion Detection, Vision System, New Media Art Work

Introduction

Facial recognition technology is a growing area of interest, where researchers are using these new application for study in psychology, marketing and product testing and other areas. There are also applications where the use of facial image capture and analysis can be used to create new methods for control, mediation and integration of personalized information into web based, mobile apps and standalone system for media content interaction. Our work explores the application of facial recognition with emotion detection, to create experiences within these domains. For mobile media applications, personalized experiences can be layered personal communication. Our current software implementation can detect smiles, sadness, frowns, disgust confusion, and anger. [9] In a mobile media environment, content on a device can be altered, to create a fun, interactive experiences, which are responsive and intelligent. By intersecting via direct communication between peer to peer mobile apps, moods can be instantly conveyed to friends and family – when desired by the individual. This creates a more personalized social media experience. Connections can be created with varying levels of intimacy, from family members, to close friends, out to acquaintances and further to broader groups as well. This technique currently uses a pattern recognition to identify shapes within an image field using a Viola and Jones[3] Haar-like features application, OpenCV[4] and a "Feret" database[10] of facial image and a library support vector machine (LibSVM)[5][11] to classify the capture from a web camera view field and identify if a face exists. The system processes the detected faces using an Elastic Bunch Graph Matching [12] technique that is trained to determine facial expressions. These facial expressions are graphed on a sliding scale to match the distance from a target emotion graph, thus giving an approximate determination of the user's mood.

State of the Art Work

Currently, many media artists are using vision systems, sensor based systems, and other technologies to create interactive experiences and mediated arts works in public spaces. In many of these works, the images are projected onto building facades, or use embedded LED arrays on building surfaces. In Asia, it is common for newer building to use vast LED arrays on the façade of the building. These projections and LED arrays can use video playback, images changing over time or other ways to control the imagery. Our work focuses on the possible use of vision systems for the detection of facial recognition, which then can be used to control or mediate visual information on surfaces in public spaces or to allow mobile apps and web based experiences or through social media.

Overview

Considering historical examples, artists have explored the use of projected imagery, or light works as a primary medium. These works may fall into one or more genre or may be in-between different genres of art. Looking at examples of installation, or environmental art works, the work of Dan Flavin [1], is exemplary in the use of light as a singular imaging medium. Flavin's work, as he has described it, is created and experience in a strict formalist approach. Formalism focuses on the way objects are made and their purely visual aspects. Nevertheless, the works, such as Flavin's, though static light alter or inform audience spatial perception of spaces where they are installed. In our study of the used of interactive elements, can the viewers perception be altered by the shifting of color or imagery based on responses detected from the viewers themselves. Further, can we use the detection of subtle emotional cues to alter the qualities of the imagery or installation? More recently, the projection of video or animated imagery on building facades or in public spaces has become a common way to attract viewer engagement. In these types of new media art work experiences, such as the 2011 transformed façade of St. Patrick Cathedral and the New Museum in New York [2]. These altered architectural and public spaces become a "canvas" where images and media content can be viewed outside of the special circumstance of the gallery or museum. Considering possible ways to allow for audience interaction, we know that use of sensors and vision systems are being used to encourage audience participation. Can subtle emotional cues be used as well?

Facial recognition for artistic, environmental installation in public space

Detection of emotion states in a public art installation to change the environmental elements is possible. Using webcams position in specific selected locations can capture facial information, the emotion states can be detected. The detected state can be used to alter projected imagery, auditory ambiance and ambiance of lighting, intensity and color. The location of the camera need not be directly within the installation space. Indeed, the control of the qualities of the imagery, lighting, or ambiance can be collect remotely in other building location, from the internet and even by mobile apps.

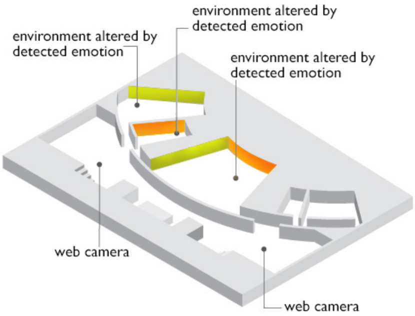

environment altered by detected emotion
environment altered by detected emotion
environment altered by detected emotion
web camera
web camera

Figure 1. A schematic view of a museum with capture location, and installation spaces.

In my work "MoodModArt" we use emotion detection to change the quality of an image based on detected moods.

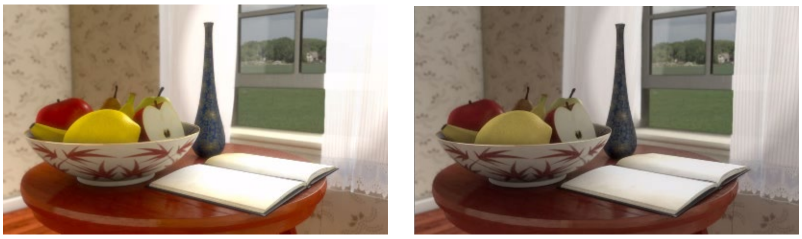


Figure 2 and 3. Images from the looped media streams in MoodModArt.

If the detected emotion of a viewer is positive, the streamed loop of video is vibrant and colorful. If the detected emotion is negative, the streamed loop of video played to a dull drab and darker view. The viewer can change quality of the image by altering their facial expression.

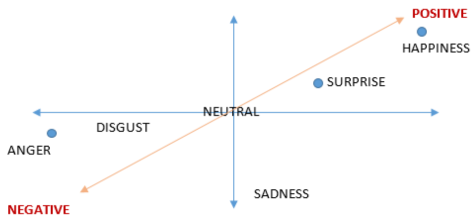
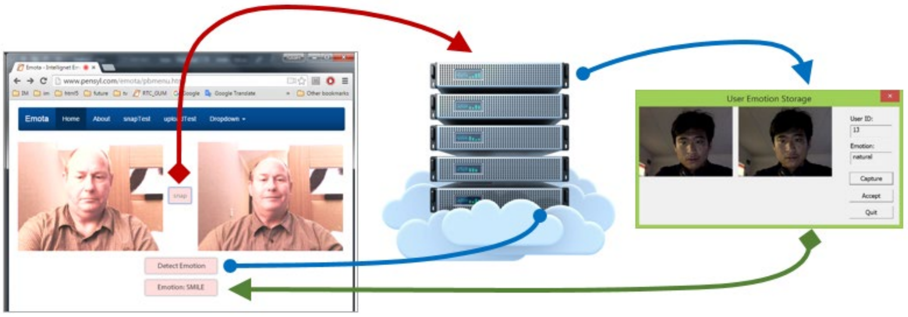


Figure 4. Graphing of emotion states on a continuum from negative to positive.

Facial recognition in mobile apps, internet webpage detection, and stand-alone kiosk systems

In mobile apps detected emotions of a viewer can be shared via social media through simple #hashtag or Facebook posts. Using HTML5/CSS3 along with Canvas, apps and webpages can be used to capture and submit an image to a back-end server application, which returns a detected emotion state. Apps and webpages submit an image to a cloud database. The server listener application listens for images arriving, tagged with user random user IDs and time stamps. The listener passes the image to a back-end server application, which returns a detected emotion state to the listener. The listener then reverts the result the webpage or app.



Emotion: SMILE

User ID: 13

Emotion: natural

Figure 5: transfer of captured images to a server application, and the return of a detected emotion.

Developmental Work in Facial recognition – Gender and Age

Our work in facial recognition began with experimentation with the detection of Gender and Age in public spaces. In our earlier project "HiPOP," we were successful in d implementing a software tool for facial recognition for use in public spaces. The focus of this work revolved around the detection of gender and age. This implementation use an image processing approach by identifying shapes within an image field using methods published byViola and Jones [3]. The technique employed a Haar-like features application [3] [5] and a “Feret” database [6] [10] of facial images. Support vector machine (LibSVM) [5] was used to classify the faces to glean attributes such as gender, age and other individual characteristics. The system segmented the captured image to recognize face rectangles; Scaling to 64x64 pixel grayscale image and equalizing the histogram to increase contrast. The OpenCV [4] library used to detect and segment faces from video images through the following methods:

1. Using a cascade of boosted classifiers working with Haar-like features.
2. Training classifiers by a database of face and non-face images.
3. Scanning input images at different scales to find regions that are likely to contain faces.
4. A SVM classifier method using data points are dealt with as a p-dimensional vector was used to detect smiles in the captured images.

Application of such a system is feasible in environments where marketing message can be targeted for individual(s) based on gender, age or other cues that can be identified. The design of the system installation allows marketing or media content to be played based on the detection of certain demographic information detected from consumers in a retail environment.

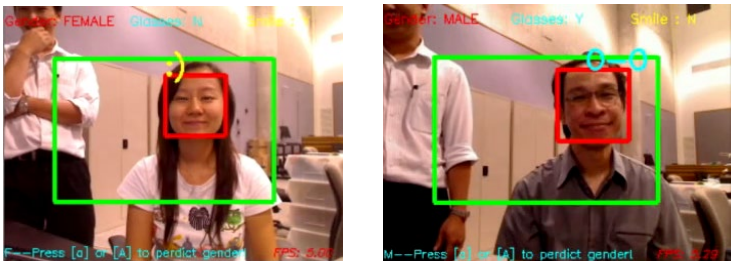
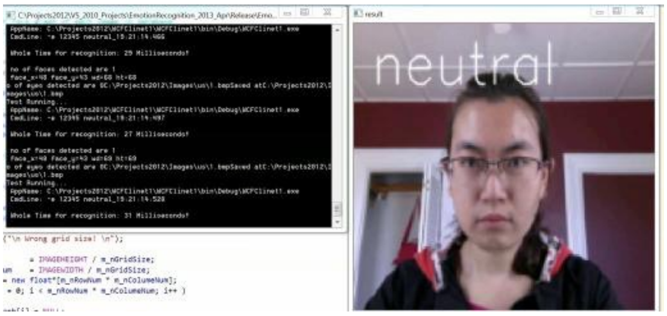

Gender: FEMALE
Glasses: N
Smile : Y
F--Press [a] or [A] to perdict gender
FPS: 5.00
Gender: MALE
Glasses: Y
Smile : N
M--Press [a] or [A] to perdict genderl

Figure 6. Detected genders invoke playback of targeted media.

Development of emotion detection - Emota v1.0

Work on emotion detection in the initial stages used a hybrid approach with a library of images, each with an Elastic Bunch Graph Match (EBGM) [7] [12] the software implementation was designed with two module to process the captured video images and give the resulting detected emotion. The “ImageNormalizer” module detected the face from an image, crops, resizes to a standard size (90 x 100 pixels), and converts these to grayscale. The normalized image is input to the EBGM program. Training for detection of emotion states in an individual was required for accuracy. The technique used a database of filtered images that are defined with an index set that are identified as one of seven emotion states. The “EmotionRecognition” module integrated with “ImageNormalizer” so that every captured frame is normalized and the detected face is stored in normalized form on the fly. “EmotionRecognition” used EBGM with the on the fly normalization program to output a detected emotion state from the captured image.



C:\Projects2012\VS_2010_Projects\EmotionRecognition_2013_Apr\Release\Emo...
result
neutral

```
 AppName: C:\Projects2012\WCFClinet1\WCFClinet1\bin\Debug\WCFClinet1.exe
 CmdLine: -e 12345 neutral_19:21:14:466

 Whole Time for recognition: 29 Milliseconds!

 no of faces detected are 1
 face_x=48 face_y=43 wd=68 ht=68
o of eyes detected are 0C:\Projects2012\Images\us\1.bmpSaved atC:\Projects2012\I
mages\us\1.bmp
Test Running...
 AppName: C:\Projects2012\WCFClinet1\WCFClinet1\bin\Debug\WCFClinet1.exe
 CmdLine: -e 12345 neutral_19:21:14:497

 Whole Time for recognition: 27 Milliseconds!

 no of faces detected are 1
 face_x=48 face_y=43 wd=69 ht=69
o of eyes detected are 0C:\Projects2012\Images\us\1.bmpSaved atC:\Projects2012\I
mages\us\1.bmp
Test Running...
 AppName: C:\Projects2012\WCFClinet1\WCFClinet1\bin\Debug\WCFClinet1.exe
 CmdLine: -e 12345 neutral_19:21:14:528

 Whole Time for recognition: 31 Milliseconds!
```

```
("\n Wrong grid size! \n");

       = IMAGEHEIGHT / m_nGridSize;
um     = IMAGEWIDTH / m_nGridSize;
= new float*[m_nRowNum * m_nColumeNum];
 = 0; i < m_nRowNum * m_nColumeNum; i++ )
```

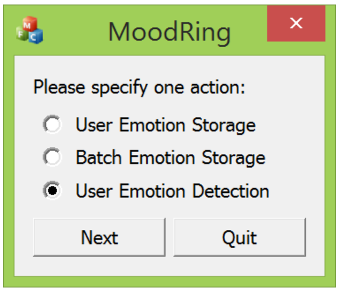
Figure 7: A screen capture of Emota v1.0 in action.

Development of emotion detection - Emota v2.0 (Mood Ring)

Additional work in the the detection of emotion has continued with the version 2.0, entitled “MoodRing.” This implementation has four modules: MoodRing MFC, MoodRing Core, Weight Trainer, and Database Compressor.

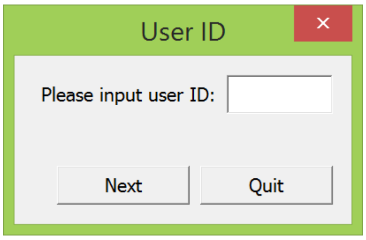
Both MoodRing MFC and MoodRing Core are implementations of the project’s core part. MoodRingMoodRing Core is the interface version which shows how to setup this project under different platforms. WeightTrainer is used to train weight of each anchor point to calculate similarity among subgraphs. Once a model is trained, elastic bunching graphs [7] [12] can be stored and compared instead of images. Database Compressor is used to compress elastic bunching graphs by comparing, searching, and combing similar graphs based on the distance among them.

MoodRing MFC is a standalone MFC version which supports emotion storage, batch training, and emotion detection. There are options for emotion storage: User Emotion Storage and Batch Emotion storage. Batch Emotion storage allows user to parse batch amount of images to xml files and add these files to data set of certain user. The batch module is designed mainly to train large amount of images in order to set up the default dataset which belongs to the default user.

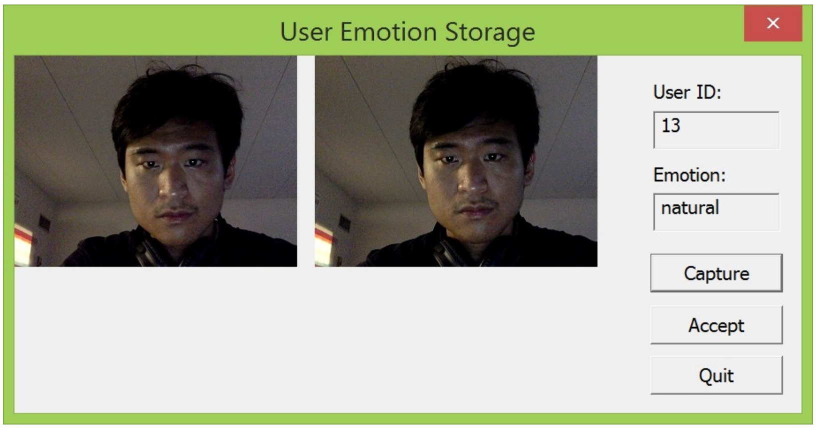


Please specify one action:

○ User Emotion Storage
○ Batch Emotion Storage
● User Emotion Detection



Please input user ID:

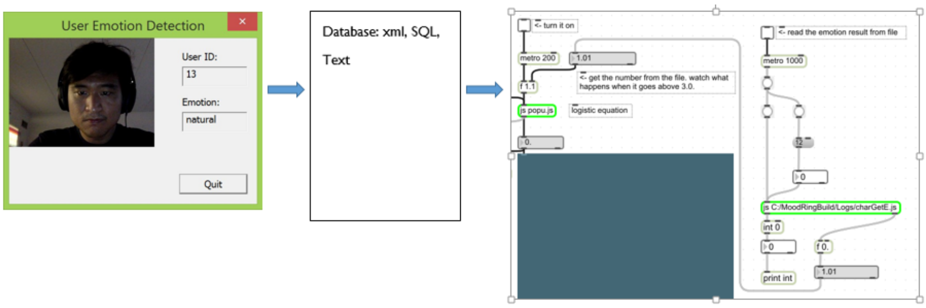

User Emotion Storage

User ID: 13

Emotion: natural

Figure 8: The interface windows for operation of MoodRing MFC.

User Emotion Storage allows user to capture, extract, and store emotions to numeric values one by one using a web camera. To use this system, the user initiates the training and capture of emotion state facial expressions by instantiating “User Emotion Storage.” The system prompts for the seven emotion state expressions, and the images are captured and accepted. Once the seven states are stored, the system is effectively trained for that user.


User Emotion Detection

User ID: 13

Emotion: natural

Database: xml, SQL, Text

Quit
<- turn it on
metro 200
1.01
<- get the number from the file. watch what happens when it goes above 3.0.
f 1.1
js popu.js
logistic equation
0.
<- read the emotion result from file
metro 1000
12
0
js C:/MoodRingBuild/Logs/charGetE.js
int 0
0
f 0.
1.01
print int

Figure 9: Detected emotion states are stored in a periodic manner to a text file that is readable by other software, such as Max/MSP.

User Emotion Detection allows real-time user emotion detection. This can be used as a control for interaction, media, and environmental elements.

Emota v2.0 Mood Ring Core and Data Processor

Image Pre-processing

First, we apply some image standardizations to get a small size gray scale image. Second, a series of image preprocessing operations are adopted, including noise removal, and Image balance.

Noise Removal[16]. For each pixel, we calculate and accumulate the difference of all its neighbor points as the weight of this pixel:

$$weight(x) = K_1 e^{-K_2 \sum_{p \subset P} |f(p) - f(x)|}$$

where P is the neighbor point set of pixel x, and *f*(x) is the pixel value of x.

Then, we traverse the image again with a weighted average filter for each pixel.

$$g(x) = \alpha \frac{\sum_{p \subset P} weight(p) \cdot f(p)}{\sum_{p \subset P} weight(p)} + (1-\alpha) \cdot f(x)$$

Image Balance. We have noticed that vague shadow will not heavily affect Haar classifier performance, and hard shadow edge can be weakened. Thus, instead of complex shadow removal algorithm, we adopt following operations to concentrate effective image information so that Haar classifiers can find target more easily:

$$f(x) = \begin{cases} \alpha K \log(x) + (1-\alpha)x & , if\ x < 127 \\ \alpha[255 - K\log(x)] + (1-\alpha)x & , if\ x \geq 127 \end{cases}$$

Face Detection

After above operations, we adopt a set of pre-trained Haar Classifiers[15] to locate only one's eyes and mouth[13] [14]. If multiple rects are found for the same part, we run a clustering method to estimate the target rect based on the Euclidean distance. Then, a set of anchor points can be delivers based on these rects. Before feature extraction, lumen compensation is adopted to detected facial part of the image such that light conditions have less effect to the feature extraction process.

Feature Extraction

Numeric presented features are extracted through convolutions with a set of pre-calculated Gabor filters called Gabor Bank.

Gabor Bank

Gabor filters are implemented to derive orientations of features in the captured image using pattern analysis, directionality distribution of the features increases accuracy of the anchor points derived in the elastic bunch graph matching. Gabor filters of all scales and orientations compose the Gabor Bank to detect edges and textures. In the Gabor filters:

$$g(x,y) = \frac{k^2}{\sigma^2} \cdot e^{-\frac{k^2(x^2+y^2)}{2\sigma^2}} \cdot (e^{ik \cdot \begin{bmatrix} x \\ y \end{bmatrix}} - e^{-\frac{\sigma^2}{2}}), where\ k = \begin{bmatrix} k_v \cos\varphi \\ k_v \sin\varphi \end{bmatrix}, k_v = 2^{-\frac{v+2}{2}} \pi$$

MIN_SIZE, STD_SIZE, MAX_SIZE: constrains the window length of Gabor filters.

COEF_KVMAX: constrains the max value of k_v.

COEF_SIGMA: contains the σ.

COEF_COMMN: simultaneously affect k and σ.

We choose 18 Gabor filters with 6 directions and 3 phases to compose a Gabor Bank.

Directions include:

$$\left| 0, \frac{1}{6}\pi, \frac{1}{3}\pi, \frac{1}{2}\pi, \frac{2}{3}\pi, \frac{5}{6}\pi \right. ;$$

phases include:

$$\left. C\pi, C\sqrt{2}\pi, C\sqrt{3}\pi \right| ,$$

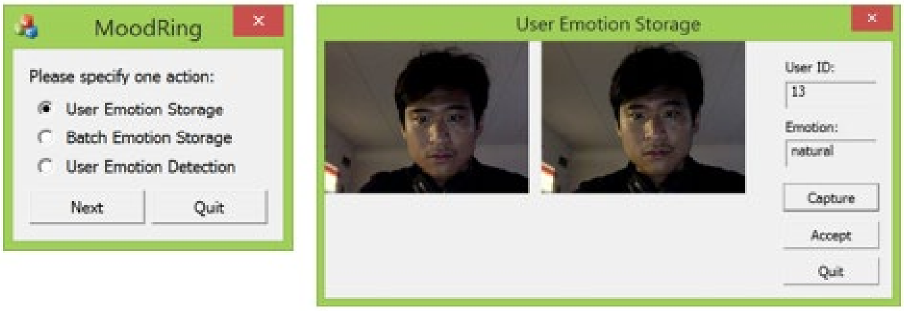
where C is a constant. Such a Gabor Bank will be initialized when the program starts, and used every time extracting features.

Elastic Bunch Graph

Operations of elastic bunching graph include graph matching, graph pruning, and adding sub-graphs from either an image or an xml file. Elastic bunching graphs applies convolutions of certain areas of images using all filters in the Gabor Bank. This results in a list of anchor information for all anchor points, where each anchor information contains a list of all convolution results corresponding to filters in the Gabor Bank. If the program is in a training mode, it will store the hierarchical results as an xml file. Otherwise, emotion detection is followed after feature extraction.

Graph pruning is the core function of Database Compressor. The pruning algorithm is basically a variety of DBSCAN [17], where the distance of sub-graphs is defined as sum of Euclidean distance of all convolution results of for all anchor points. If one cluster contains at least the minimum number of neighbor point sub-graphs, and distances of these sub-graphs are at most eps, we combine all sub-graphs in one cluster into one. Thus, very similar sub-graphs are merged to reduce storage space and comparing time. Emotion Detection

The emotion detection is a similarity comparing process. Target graph is compared with all sub-graphs in all seven emotions [9] in certain dataset. We categorize target graph the same emotion type as its most similar sub-graph. In comparison of two graphs, we can calculate a weighted average on the distance of all such convolution results of all anchors in graphs. When initialized the program, mathematical model determined by Weight Trainer will be loaded, such that weight of each anchor can be used to measuring graph similarity.


MoodRing

Please specify one action:

● User Emotion Storage

○ Batch Emotion Storage

○ User Emotion Detection

User Emotion Storage

User ID: 13

Emotion: natural

Figure 9: Training if the EmotionDetector.

There are two types of preloaded dataset used in the detection process: default graph set and user graph set. When initialized, program will load the default graph set, which only contains graph of the default user. As mentioned above, dataset for the default user is usually trained in the Batch Emotion Storage module. Since default user’s dataset of contains large amount of samples from existing database like “Feret” [6], it can be used without user graph set. However, user graph set is still a better choice because it contains fewer but more informative graphs. Based on given user ID, the program will load graphs of that user into user graph set if program can find user emotion data of this user. Otherwise, only the default graph set will be loaded.

Weight Trainer

Weight Trainer is the first step to set up the MoodRing system. Input of this module is a set of elastic bunching graphs with all seven emotions; output is a weight matrix stored as local file.

Given a set of seven graphs, G_i (i = 0, 1, 2, 3, 4, 5, or 6), and each graph G_i has sub-graphs g_{ij}, we first generate the dataset through a pair-wise comparison:

$$x = g_{ij} - g_{mn}, \quad y = \begin{cases} 0, & if\ i = m(same\ emotion) \\ 1, & if\ i \neq m(different\ emotion) \end{cases}$$

Then, because y is between 0 and 1, we apply the following logistic function on X:

$$Input\ matrix : g(X), where\ x \subset X, and\ g(x) = \frac{1}{1+e^{-t}}$$

$$Output\ matrix : Y, where\ y \subset Y$$

Now that we have transferred the dataset into this form, we adopt certain classification method, like LibSVM [11] to train the weight matrix. If size of X is small (e.g. for individual users), we will use batch training; if size of X is large (e.g. for the default user), we will use mini-batch stochastic training instead. The boundary of these algorithm is a constant value.

Future work

- Application of the detection is feasible in installation and environments, and public spaces
- Further experimentation is necessary to determine accuracy of the facial capture and emotion detection
- Further work will include continued refinement of the image processing and normalization to operate in varying lighting conditions
- Mobile app and social media application exploration will continue
- Experimentation and comparison of image library based implementations and EBGraph matching for the further development of a "universal" detector

Cross References

1. https://chinati.org/collection/danflavin.php
2. http://www.newmuseum.org/ideascity/view/flash-light-mulberry-street-installations.
3. Viola, P., & Jones, M. (2001). Robust real-time object detection. Paper presented at the Second International Workshop on Theories of Visual Modelling Learning, Computing, and Sampling
4. Bradski, G. and Kaehler, A., (2008). Learning OpenCV. OReilly.
5. Burges, C. J.C., (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121-167
6. http://www.nist.gov/huma nid/colorferet
7. Wiskott, L.; Fellous, J.-M.; Kuiger, N.; von der Malsburg, C. (1997) Face recognition by elastic bunch graph matching, Pattern Analysis and Machine Intelligence, IEEE Transactions on Machine Intelligence, Volume: 19 Issue:7 Pages 775 - 779
8. Ekman, P., (1999), "Basic Emotions", in Dalgleish, T; Power, M, Handbook of Cognition and Emotion, Sussex, UK: John Wiley & Sons,
9. Database FERET http://www.nist.gov/humanid/color
   Feret FA|FB|QR|QL|HL|HR 2. Rate of accuracy FB(dvd2) : 246/268 = 91.791%
10. Chang, C.C., Lin C.J., https://www.csie.ntu.edu.tw/~cjlin/libsvm/
11. Bolme, D., Elastic Bunch Mapping, cs.colostate.edu/~vision/publications/Bolme2003.pdf, 2003
12. Hlang, H.K.T., Robust Algorithm for Face Detection in Color Images International Journal of Modern Education and Computer Science, 2012.
13. Alpers, G.W., Happy Mouth and Sad Eyes: Scanning Facial Expressions, American Psychological Association, Emotion. Aug 11, 2011 (4):860-5. doi: 10.1037/a0022758
14. Messom, C., Barczak, A., Fast and Efficient Rotated Haar-like Features using Rotated Integral Images, Int. J. of Intelligent Systems Technologies and Applications, 2009 Vol.7, No.1, pp.40 - 57

15. Wang Mingjia, Zhang Xuguang, Han Guangliang, Wang Yanjie, "Elimination of impulse noise by auto-adapted weight filter", Optics and Precision Engineering, 2007, 15(5), pp. 779-783.
16. M. Ester, H. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise”, Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD, 96), pp.226-231, 1996.