

BEX3012 Project Report Detecting Facial Expressions in Professional Tennis Matches

Stephanie Kobakian

28 March 2017

Contents

1	Introduction	2
2	Materials	3
2.1	Images	3
2.2	Software APIs	3
3	Procedure	5
3.1	Process Images	5
3.2	Process API results	5
4	Analysis	8
5	Validation	8

1 Introduction

Facial Recognition is beginning to be explored in sports environments, this presents quality issues.
helpful for applications

2 Materials

2.1 Images

The image set contained 6406 still shot 800x450px size video frames, taken every 3 seconds from 5 minute segments sampled from 105 Australian Open 2016 match broadcast videos. The broadcast videos are a combination of multiple camera feeds, these angles vary depending on the court the match was played on.



Figure 1: Example Image: 2016_CT6_R01_JMillman_AUS_vs_DSchwartzman_ARG_MS159_clip.0020.png

2.2 Software APIs

Three emotion recognition APIS were considered. They were chosen for their accessibility and online reviews of their performance, and these are well recognised options for Facial recognition currently available.

Table 1: This details the capabilities we considered important in recognising emotions in images of faces.

Attribute	Google	Microsoft	Skybiometry
Call Limits	10 per second	20 per minute	100 per hour
Emotion Output	Categorical Likelihood	Numeric Proportions, (0.0-1.0)	Confidence Value, (0-100)
Number of Emotions	4	8	7
Cost and Access	Account and Payment	Account	Account and Payment
API Access	REST	REST	REST

As can be seen above, a noticeable difference between the three is the amount of times the API can be called within a given time frame. Skybiometry had the largest imposition on Call Limits as it only allowed 100 API calls to be processed per hour. Microsoft also had a limit imposed, but this allowed for much more to be processed with the possibility of 1200 images to be processed within one hour, after accounting for the wait time between each group of 20. Google Vision's API call limit had a minimal effect for our purposes.

The number of emotions shown in the results were not congruent. These differences made it necessary to pre process the result to provide the most comparable results for analysis. Microsoft provided the most emotion options, contempt being the addition that was not provided by Skybiometry. Google provided the least, only giving the options for four different emotions.

The outputs from the APIs were very different. Google provides likelihoods of an emotion occurring on a particular face in categorical options ranging from Very Unlikely to Very Likely. Microsoft provide Proportions, ranging between 0 and 1, whereas Skybiometry results in a Confidence of the emotion occurring in the specified face in values from 0 to 100.

3 Procedure

3.1 Process Images

The original image set was subset to create images of faces specifically suited for emotion recognition purposes. A crop was performed to create a new image that focused on the area marked as a face by Google¹. A border to frame the face was also included for visual appeal. These 1319 new images, of varying sizes, were hosted on Google Drive to allow for URL access from the API to the individual images.

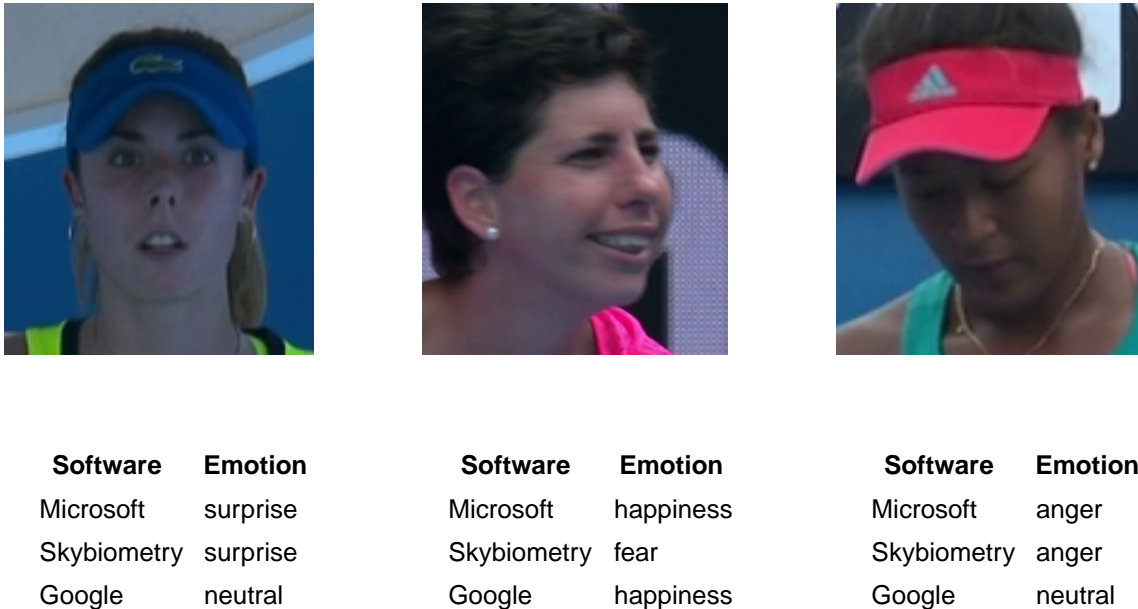


Figure 2: Three Faces in the image set, after extracting faces from the full broadcast video stills, and the dominant emotion each software recognised for them.

3.2 Process API results

The results extracted from each of the APIs differed, re configuring of these results was needed to create comparable sets for analysis.

Microsoft provided a POST request object in a JSON format. This was easily formatted to create a data set of the eight numeric emotion values provided for each face. In the event that emotion values were not returned for a certain face, NAs were presented in the data set.

Table 2: The Microsoft API provided eight numerical values, one for each emotion.

¹based on Google's bounding box coordinates

anger	contempt	disgust	fear	happiness	neutral	sadness	surprise
0.001504	0.000631	0.002371	0.003267	3e-05	0.716105	0.004738	0.271353

Skybiometry as provided a POST request object in a JSON format. However, unlike Microsoft, the results included a True or False Value for each of the seven emotions and a confidence level ‘confidence value as a percentage from 0 to 100’² representing the confidence of the emotion being present on a player’s face.

Table 3: The Skybiometry API provided seven numerical Confidence values, one for each emotion.

neutral	anger	disgust	fear	happiness	sadness	surprise
0	21	41	0	0	5	76

The process undertaken to receive results Google produces was the same as the previous APIs. However rather than numeric values, it returned categorical likelihoods for the four emotions it considered.

Table 4: The Google Vision API provided four likelihood possibilities, one for each emotion.

joy	sorrow	anger	surprise
VERY_UNLIKELY	UNLIKELY	VERY_UNLIKELY	UNLIKELY

These differing outputs needed to be arranged in a comparable manner. Noticing that the sums of the individual emotions for each face in the Microsoft results was approximately one provided the basis for transformation for the other APIs. Dividing the individual Skybiometry emotion confidences by the sum of the confidences would give the proportional likelihood of each emotion being the dominant emotion in a face.

This calculation was simple to perform and resulted in numeric values comparable to Microsoft’s values.

Skybiometry Transformed:

Table 5: The Skybiometry API provided 7 numerical Confidence values, one for each emotion.

neutral	anger	disgust	fear	happiness	sadness	surprise
0	0.146853	0.286713	0	0	0.034965	0.531469

Google’s results then required a more complex transformation for the output to also be comparable. As there were five possible categories assigned to each emotion the numeric values were chosen to

²<https://skybiometry.com/documentation/#document-21>

sit in the middle of five 20 point ranges. This allowed a numerical representation of each possible likelihood value and would allow for numeric comparisons to be made. The values were assigned according to the table below:

Table 6:

Likelihood	Value
VERY_UNLIKELY	10
UNLIKELY	30
POSSIBLE	50
LIKELY	70
VERY_LIKELY	90

As both Microsoft and Skybiometry included a neutral category we considered how to incorporate a neutral category for the Google results.

Our approach considered that when a face was “very unlikely” to be any of the four emotion categories it could be deemed “neutral”. However when one emotion was stronger than other it could still be possible that the face only had a small lean toward this emotion and it may not be the obvious choice. Therefore the Confidence value for neutral was calculated according to the formula below:

$$neutralConf_i = 100 - \max(joyConf_i, sorrowConf_i, angerConf_i, surpriseConf_i)$$

The process to derive the numeric values for Google followed the process undertaken for Skybiometry values. Where the individual emotion confidences were divided by the sum of the emotion confidence values for each face. This resulted in congruent emotion data.

Table 7: The Google Vision API provided four likelihoods, one for each emotion.

joy	sorrow	anger	surprise	neutral
0.0666667	0.2	0.0666667	0.2	0.4666667

4 Analysis

Each API returned a different amount of non null results. Microsoft produced emotion results for 732 faces. This was more than Skybiometry, as it produced only 634 results for emotions on faces. Google provided the most, with 1242 emotion results³.

5 Validation

We considered a subset of images where the emotions on the faces were manually considered.

³As the initial set was faces identified by Google there is the possibility of bias toward being able to identify emotions in faces it was able to find originally.