BABEŞ BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

EMOTION RECOGNITION

– ITSG report –

Team members
Bartha Melania Beata, SDI, 254
Pirvu Alexandru, SDI, 254

2019

Abstract

One can evaluate the satisfaction level of an adult easily, by questionnaires, forms, live interviews after the reviewed activity or even self analysis techniques. This cannot be applied to preschoolers as they have a limited means or capacity for submitting forms or objectively answering questions. This project addresses the issue of automatic satisfaction level detection by the means of face feature analysis techniques. Machine learning has proven itself to be capable of automated complicated task. The proposed approach uses AI and machine learning to do real time analysis and emotion prediction to give comprehensive information about the satisfaction level relative to the performed task. Two support vector classifier models based on different type of data are created and compared. Test results show high integration potential right away, with minimum requirements and few teacher briefing information.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 What? Why? How?

We want an objective measurement of the emotions that children experience during the interaction. This requires the development of an application that allows the identification of the emotional states of a preschooler during the course of an activity. We need to associate tasks that children do and the frequency of an emotion. We want to detect emotions through facial expressions. For this association we need artificial intelligence algorithms.

We then use support vector machines to classify the facial expressions and emotions. Support vector machines have been proven useful in a number of pattern recognition tasks including face and facial action recognition.

1.2 Paper structure and original contribution(s)

The research presented in this paper advances the theory, design, and implementation of several particular models.

The main contribution of this report is to present an intelligent algorithm for solving the problem of . . ..

The second contribution of this report consists of building an intuitive, easy-to-use and user friendly software application. Our aim is to build an algorithm that will help . . ..

The third contribution of this thesis consists of . . ..

The present work contains *xyz* bibliographical references and is structured in five chapters as follows.

The first chapter/section is a short introduction in . . ..

The second chapter/section describes ....

The chapter/section 4 details ....

Chapter 2

Scientific Problem

2.1 Problem definition

It can be hard to extract objective, relevant information regarding a specific task from a fully grown, schooled adult, which is capable of self evaluation, introspection and has a bigger experience to compare certain feelings to. With children, a lot of the helping factors are not present. They can loose focus and interest very fast, can be unpredictable and easily influenced. As in every field in the present day, machine learning and AI can be integrated and prove itself useful. The main goal is to automatically extract facial features as the children are performing a certain task, compute the emotion at small intervals of time and analyze those prediction in order to give an overall impression of the satisfaction level while performing the task. It can immediately be seen that an intelligent algorithm will make possible a fast feedback mechanism with just a frontal camera. Depending of the used algorithms and the needed precision and performance, there can be limitations in terms of performance or ambiance. The variations are not significant and in average conditions, the performance of the system is satisfactory. The workload of the project can be split in few different tasks as follows:

- train and validate a model to predict an emotion based on facial features
- extract real time facial features using the camera in from of the monitor
- predict emotions and analyze the result in order to give relevant information to the teacher

2.2 Choosing the right tool

For emotion recognition there is necessary to use different supervised machine learning algorithms in which a large set of annotated data is fed into the algorithms for the system to learn and predict the

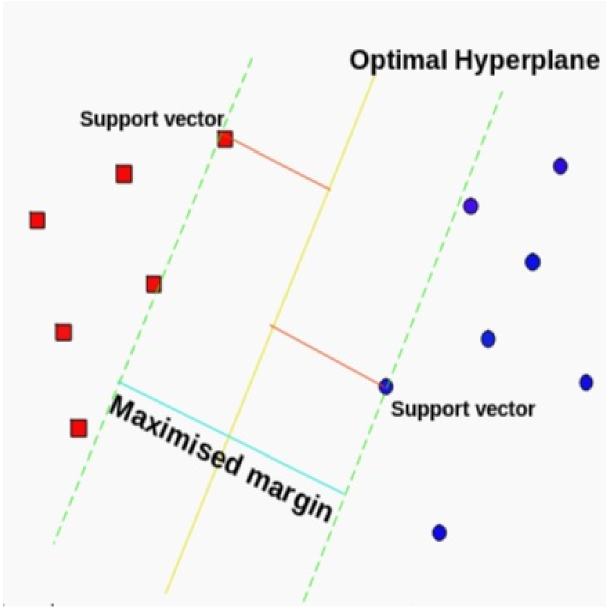


Figure 2.1: Optimal Hyperplane using the SVM algorithm

appropriate emotion types. Machine learning algorithms generally provide more reasonable classification accuracy compared to other approaches, but one of the challenges in achieving good results in the classification process, is the need to have a sufficiently large training set.

The intelligent algorithm we used is SVM (Support Vector Machine). SVM is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

In SVM, it is easy to have a linear hyper-plane between these two classes. SVM has a technique called the kernel trick. These are functions which takes low dimensional input space and transform it to a higher dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then find out the process to separate the data based on the labels or outputs youâve defined. [4]

Advantages:

- SVM works relatively well when there is clear margin of separation between classes
- SVM is more effective in high dimensional spaces
- SVM is effective in cases where number of dimensions is greater than the number of samples
- SVM is relatively memory efficient

Disadvantages:

- SVM algorithm is not suitable for large data sets
- SVM does not perform very well, when the data set has more noise i.e. target classes are overlapping
- In cases where number of features for each data point exceeds the number of training data sample , the SVM will under perform
- As the support vector classifier works by putting data points, above and below the classifying hyper plane there is no probabilistic explanation for the classification

Chapter 3

State of art/Related work

Automatically detecting facial expressions has become an increasingly important research area.

In 2000, the Cohn-Kanade database was released for the purpose of promoting research into automatically detecting individual facial expressions. [2] They recorded facial behavior of 210 adults. Participants were 18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups. For the CK+ distribution, they have augmented the dataset further to include 593 sequences from 123 subjects. They identified 7 basic emotion categories: Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise. They uses support vector machines to classify the facial expressions and emotions.

Their results were considerable and the hit rates for each emotion were : Angry - 75.00%, Disgust - 94.74%, Fear - 65.22%, Happy - 100%, Sadness - 68.00%, Surprised - 77.09%, Neutral - 100%. [2]

Tarnowski et. al in their article presented the results of recognition of seven emotional states. Coefficients describing elements of facial expressions, registered for six men aged 26-50, were used. Each subject participated in two sessions. A participant mimicked all seven examined emotional states. As a result, 42 5-second sessions were registered for each user. The entire database contained a total of 252 facial expressions. [6] They used nearest neighbor classifier (3-NN) and two-layer neural network classifier (MLP) with 7 neurons in the hidden layer. The input of the network were six AU, and the output was one of the seven emotional states.

They tested two ways to recognize emotions: a) subject-dependent - for each user separately and b) subject-independent - for all users together. In both cases, for 3-NN classifier, data were randomly divided on the teaching part (70%) and the testing part (30%) and for MLP into three groups: teaching (70%), testing (15%) and validation (15%). In subject-independent approach, the classifier accuracies (CA) for 3-NN and MLP algorithms were respectively 95.5% and 75.9%. For user-independent classification the highest classification accuracy (73%) was achieved for MLP neural network. [6]

Chapter 4

Proposed approach

The project was build upon three conceptual technical parts, solved using either existing tools or by implementing simple AI algorithms:

1. Previous experiments arose the need to have facial feature extraction tools which were made public and open source by their contributors. One of the most comprehensive and well documented tool is Open face. It has multiple utility functions to extract various information from various sources. It can process batch files, recorded videos or real time video using the camera device connected to the computer. For the described tasks addressed by this project, there will be two uses of OpenFace:

   - for validation and testing purposes: extracting AUs from each frame of every test video; this can be done just once and the results be used after each validation step
   - end result integration: start the camera device and make computations on the real time frames in order to deliver the action units to the next processing phase as early in time as possible

2. Scikit-learn is a software machine learning library for Python. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN. We used the given SVM implementation, which has a simple interface to configure and use the classifier.

3. This module accomplished data collection, training the classifier and appliying the result on the tast data and finally on the running end software. Collecting the data from the different datasets described bellow proved to be a difficult task. Datasets were very different, most notably the

given labels did not fully coincide, which made cross testing and validation hard and necessitated the exclusion of almost 30% of the data in order to make the comparison possible.

AU	Description
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser
7	Lid Tightener
9	Nose Wrinkler
10	Upper Lip Raiser
11	Nasolabial Deepener
12	Lip Corner Puller
13	Cheek Puffer
14	Dimpler
15	Lip Corner Depressor
16	Lower Lip Depressor
17	Chin Raiser
18	Lip Puckerer
20	Lip stretcher
22	Lip Funneler
23	Lip Tightener
24	Lip Pressor
25	Lips part**
26	Jaw Drop
27	Mouth Stretch
28	Lip Suck
41	Lid droop**
42	Slit
43	Eyes Closed
44	Squint
45	Blink
46	Wink
51	Head turn left
52	Head turn right
53	Head up
54	Head down
55	Head tilt left
56	Head tilt right
57	Head forward
58	Head back
61	Eyes turn left
62	Eyes turn right
63	Eyes up
64	Eyes down

Table 4.1: The action units and their codes

4.1 Dataset

4.1.1 Cohn-Kanade Dataset (CK+)

We used for training the the Extended Cohn-Kanade Dataset (CK+). Facial behavior of 210 adults was recorded using two hardware synchronized Panasonic AG-7500 cameras. Participants were 18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups. Image sequences for frontal views and 30-degree views were digitized into either 640x490 or 640x480 pixel arrays with 8- bit gray-scale or 24-bit color values. Full details of this database are given in. For the CK+ distribution, they have augmented the dataset further to include 593 sequences from 123 subjects The image sequence vary in duration (i.e. 10 to 60 frames) and incorporate the onset (which is also the neutral frame) to peak formation of the facial expressions. In this Phase there are 4 zipped up files. They relate to:

1) The Images - there are 593 sequences across 123 subjects which are FACS coded at the peak frame. All sequences are from the neutral face to the peak expression. 2) The Landmarks - All sequences are AAM tracked with 68points landmarks for each image. 3) The FACS coded files - for each sequence (593) there is only 1 FACS file, which is the last frame (the peak frame). Each line of the file corresponds to a specific AU and then the intensity. An example is given below. 4) The Emotion coded files - ONLY 327 of the 593 sequences have emotion sequences. This is because these are the only ones the fit the prototypic definition. Like the FACS files, there is only 1 Emotion file for each sequence which is the last frame (the peak frame). There should be only one entry and the number will range from 0-7 (i.e. 0=neutral, 1=anger, 2=contempt, 3=disgust, 4=fear, 5=happy, 6=sadness, 7=surprise).[2]

4.1.2 The Child Affective Facial Expression (CAFE) set

CAFE database used for training. Participants: One hundred undergraduate students (half male, half female) from the Rutgers University-Newark campus participated (M = 21.2 years). The sample was 17% African American, 27% Asian, 30% White, and 17% Latino (the remaining 9% choseâOtherâ or did not indicate their race/ethnicity). [1] The CAFE is a collection of photographs taken of 2 to 8 year-old children (M = 5.3 years; R = 2.7 â 8.7 years) posing for six emotional facial expressions based on Ekman and Friesenâs(1976) basic emotional expressionsâsadness, happiness, surprise, anger, disgust, and fearâplus a neutral face. In total, we had 154child-models (90 F, 64 M) pose each of these seven expressions. There was substantial variability across the faces, with a mean of 66% accuracy

Emotion Labels	Number of Videos	Number of Children
Curiosity	385	51
Uncertainty	344	53
Excitement	355	49
Happiness	604	60
Surprise	298	49
Disgust	137	35
Fear	50	20
Frustration	131	31

Figure 4.1: Number of videos containing each emotion and number of people who have expressed that emo- tion in EmoReact.[3]

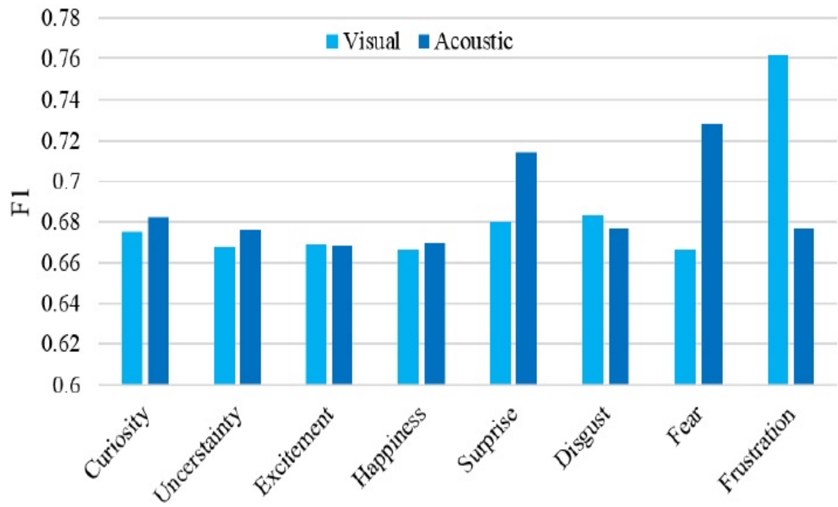


Figure 4.2: Comparison between visual and acoustic models in predicting emotions.[3]

across the 1192 photographs of the set, and a range of 0â98% correct.

4.1.3 EmoReact - video dataset

EmoReact database used for testing. YouTube has become a significant source of video data where hundreds of hours of new videos are uploaded every minute . They have selected React channel from YouTube as the source from which we downloaded videos of children who are reacting to different subjects. These videos contain children between the ages of four to fourteen years old, from different races and both genders. They have downloaded videos of children reacting to 37 subjects that include food, technology, YouTube videos and gaming devices. Dataset of 63 children from which 32 are female and 31 are male, total of 1254 video clips.[3] Due to the nature of the data, labels were given differently than the other two datasets which used images rather than videos. Each video was assigned a list of emotions present int he video. Also, different labels were used from which only half can be exactly mapped to previous datasets labels.

4.2 Training

Facial Action Coding System (FACS) is a system to taxonomize human facial movements by their appearance on the face. Movements of individual facial muscles are encoded by FACS from slight different instant changes in facial appearance. Using FACS it is possible to code nearly any anatomically possible facial expression, deconstructing it into the specific Action Units (AU) that produced the expression. It is a common standard to objectively describe facial expressions. OpenFace is able to recognize a subset of AUs, specifically: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45. [5]

We trained two models for the Cohn Kanade dataset and CAFE. Due to different data structures and labels, the data gathering was split:

- CK: the action units were already extracted for each final image from the data set; the already given AUs were used and attributed as training data to the given label
- CAFE: for each emotion label, we extracted with OpenFace the AUs in order to build the training data

The resulting models were saved for later used in validation, testing and execution.

4.3 Testing and Results

The third dataset(EmoReact) was used as testing data for both models. Given the differences in labels, the not matching labels were treated as not applicable(N/A). For each video, the tested model predicted emotions for each frame at a given interval and the results were gathered in a top emotion list. The first two emotions in the list were compared to the given list of labels from the testing dataset. If any of the top two emotions is present in the label list, we count it as a succes, otherwise as a failure. In can be seen that the cafe model achieves an almost 50% success rate, while the CK model, which was trained using "perfect" emotion expressions, achieves a significantly worse result.

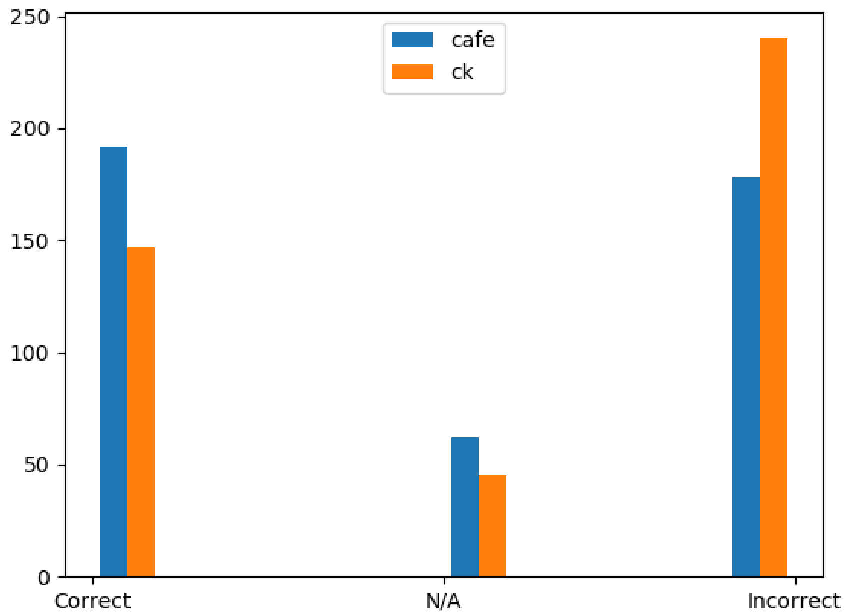


Figure 4.3: Side to side comparison of the two models

Chapter 5

Application (numerical validation)

Explain the experimental methodology and the numerical results obtained with your approach and the state of art approache(s).

Try to perform a comparison of several approaches.

Statistical validation of the results.

5.1 Methodology

- What are criteria you are using to evaluate your method?
- What specific hypotheses does your experiment test? Describe the experimental methodology that you used.
- What are the dependent and independent variables?
- What is the training/test data that was used, and why is it realistic or interesting? Exactly what performance data did you collect and how are you presenting and analyzing it? Comparisons to competing methods that address the same problem are particularly useful.

5.2 Data

Describe the used data.

5.3 Results

Present the quantitative results of your experiments. Graphical data presentation such as graphs and histograms are frequently better than tables. What are the basic differences revealed in the data. Are they statistically significant?

5.4 Discussion

- Is your hypothesis supported?
- What conclusions do the results support about the strengths and weaknesses of your method compared to other methods?
- How can the results be explained in terms of the underlying properties of the algorithm and/or the data.

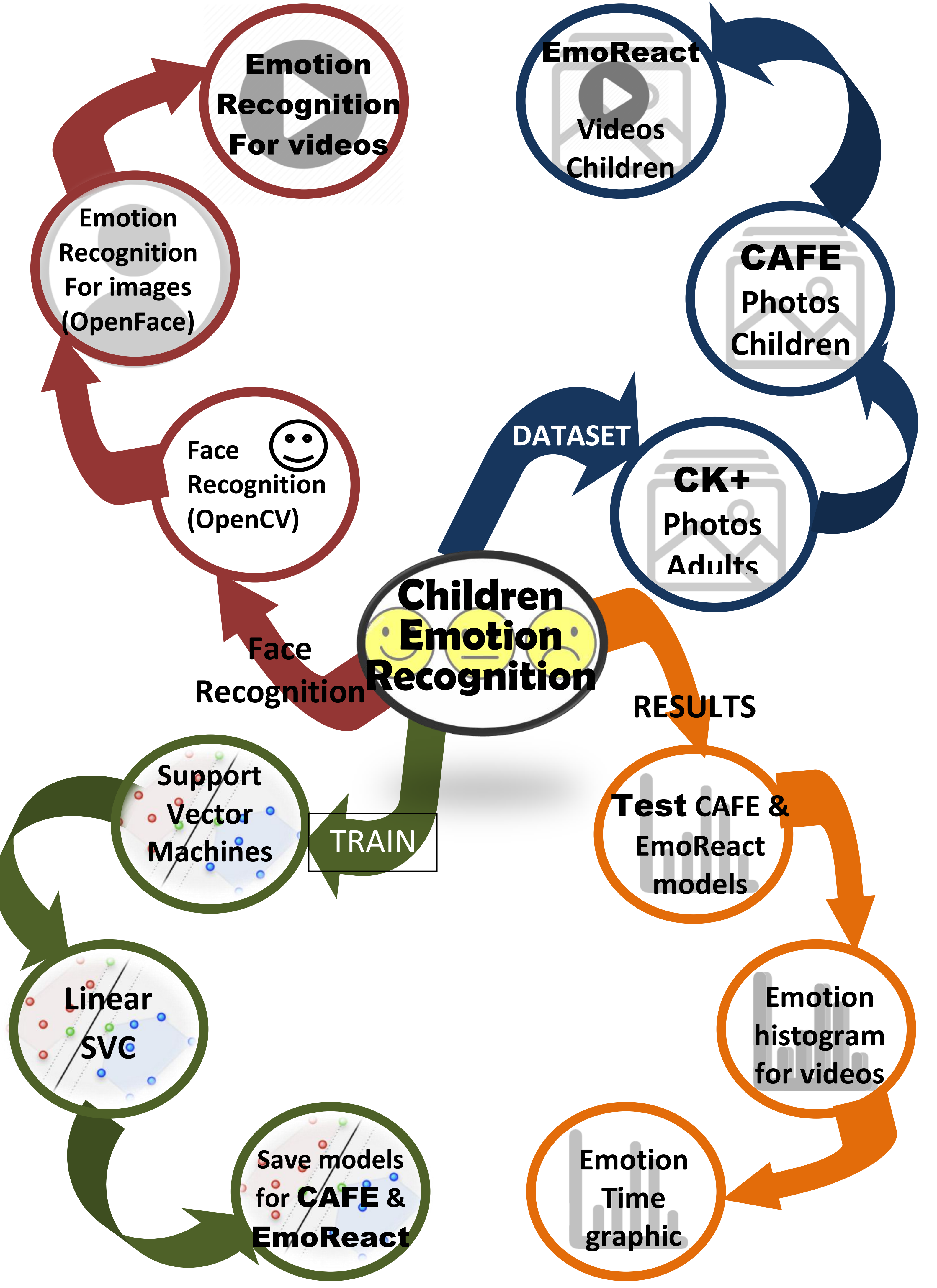
Chapter 6

Conclusion and future work

Try to emphasise the strengths and the weaknesses of your approach. What are the major shortcomings of your current method? For each shortcoming, propose additions or enhancements that would help overcome it.

Briefly summarize the important results and conclusions presented in the paper.

- What are the most important points illustrated by your work?
- How will your results improve future research and applications in the area?


Children Emotion Recognition
DATASET
CK+ Photos Adults
CAFE Photos Children
EmoReact Videos Children
Face Recognition
Face Recognition (OpenCV)
Emotion Recognition For images (OpenFace)
Emotion Recognition For videos
TRAIN
Support Vector Machines
Linear SVC
Save models for CAFE & EmoReact
RESULTS
Test CAFE & EmoReact models
Emotion histogram for videos
Emotion Time graphic

Bibliography

[1] Vanessa LoBue and Cat Thrasher. In *The Child Affective Facial Expression (CAFE) set*. Databrary, 2014.

[2] Patrick Lucey, Jeffrey Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. In *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*, pages 94 – 101, 07 2010.

[3] Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E. Hughes, and Louis-Philippe Morency. Emoreact: A multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, pages 137–144, New York, NY, USA, 2016. ACM.

[4] Sunil Ray. Understanding support vector machine algorithm from examples (along with code). https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/?fbclid=IwAR1F7gUZcC2iqLleuii30qWvCuDwEGqHAwlIVLseLHnHdYeGNumHljVD9gY, 2017. Accessed: 2019-11-28.

[5] Yao Chong Lim Tadas Baltrusaitis, Amir Zadeh and Louis-Philippe Morency. In *OpenFace 2.0: Facial Behavior Analysis Toolkit*, 2018.

[6] Pawel Tarnowski, Marcin Kolodziej, Andrzej Majkowski, and Remigiusz J. Rak. In *Emotion recognition using facial expressions*, pages 1175 – 11184, 06 2017.