

**Title: EMOTION TRANSFORMATION FEATURE:
NOVEL FEATURE FOR DECEPTION DETECTION IN VIDEOS**

Presenter: Jun-Teng Yang

Authors: Jun-Teng Yang, Guei-Ming Liu, Scott C.-H Huang

Affiliation: National Tsing Hua University, Taiwan

Session Title: IEEE ICIP 2020 (Regular Session Paper)

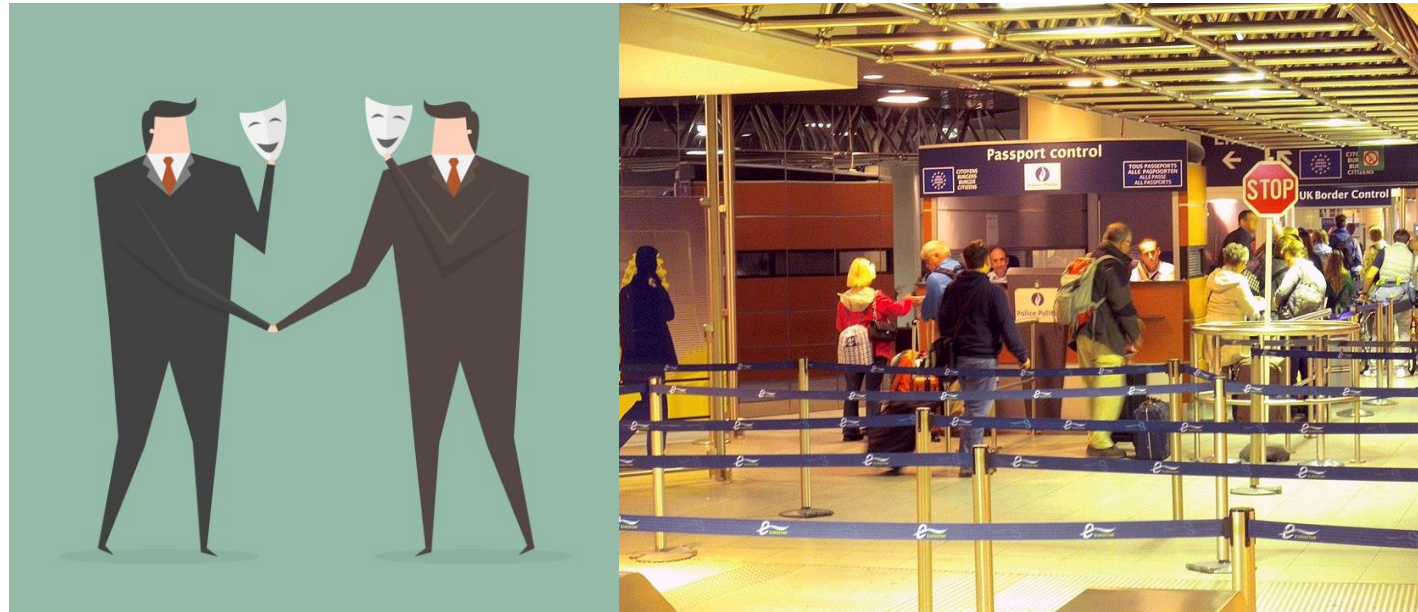


Outline

- Motivation & background
- Introduction of data set
- Our Proposed method - Emotional transformation feature
- Experiments and analysis
- Conclusions

Motivation & background

- The average accuracy of detecting lies without any special aids is 54% for normal person.
- Some deceptions may cause serious consequences: industrial spy, miscarriage of justice, job interview, etc.



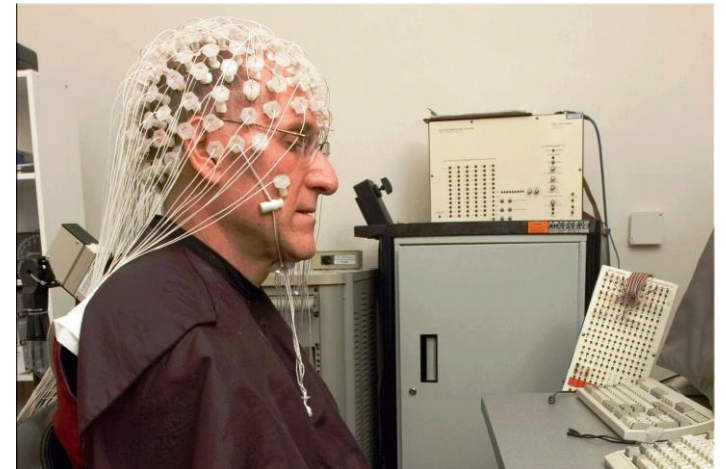
Motivation & background --- Current solution

Deception detection tools currently used:

- Polygraphs
- Event-related potential (ERP)
- functional magnetic resonance imaging (fMRI)

Disadvantages:

- The accuracies are questioned [1][2]
- Expensive and requiring experts to operate
- Limitation of usage due to skin contact



[1] N. R. Council, "The polygraph and lie detection," The National Academies Press, 2003

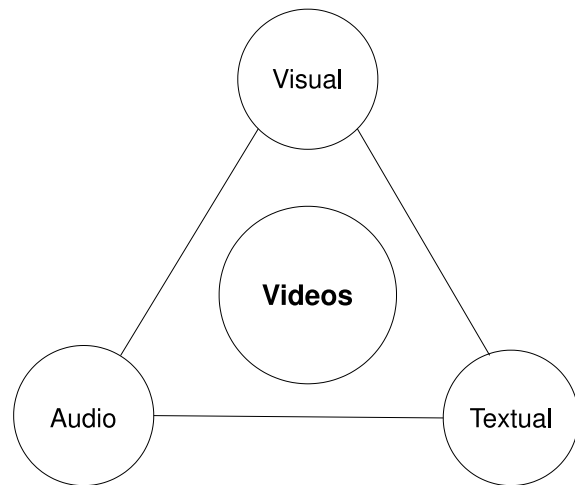
[2] M. Farah, J. Hutchinson, E. Phelps, and A. Wagner, "Functional MRI-based lie detection: Scientific and societal challenges," Nature reviews. Neuroscience, vol. 15, pp. 123–131, Feb. 2014
Pennsylvania national guard, 2 May 2013, Delegates tour NCTC, digital image, Flickr, accessed Jun.16 2020.
EconomicsUZH, 18 Mar. 2016, fMRI Scanner, digital image, Wikimedia Commons, accessed Jun.16 2020.

Antoine Lutz, 27 Jul. 2015, Barry Kerzin meditating with EEG for neuroscience research, digital image, Wikimedia Commons, accessed Jun.16 2020.

Motivation & background --- New idea

A new way to approach deception detection: Videos

- Verbal features and non-verbal features
- Lower requirement regarding the equipment, i.e., cost, operation
- Examples of practical applications, e.g., border guard (iBorderCtrl), airport (Avatar), job interview (AVIS)



Motivation & background --- Challenges

- **Computational complexity of processing videos**
 - between temporal information and complexity
- **Lack of data**
 - Requiring experts to label data
 - Not easy to collect data

Currently available resources for video deception detection:

- Dataset collected by Pérez-Rosas (Total of 121 video clips)
- Deep neural network-based methods prone to overfit

Goal : Finding features highly correlated to deception behaviors within limited data

Introduction of dataset



Basic information:

- Data are collected from real court trials.
- Including 121 video clips, 61 deceptive, 60 truthful clips
- The average length of a video clip is 28.0 seconds

Problem of data:

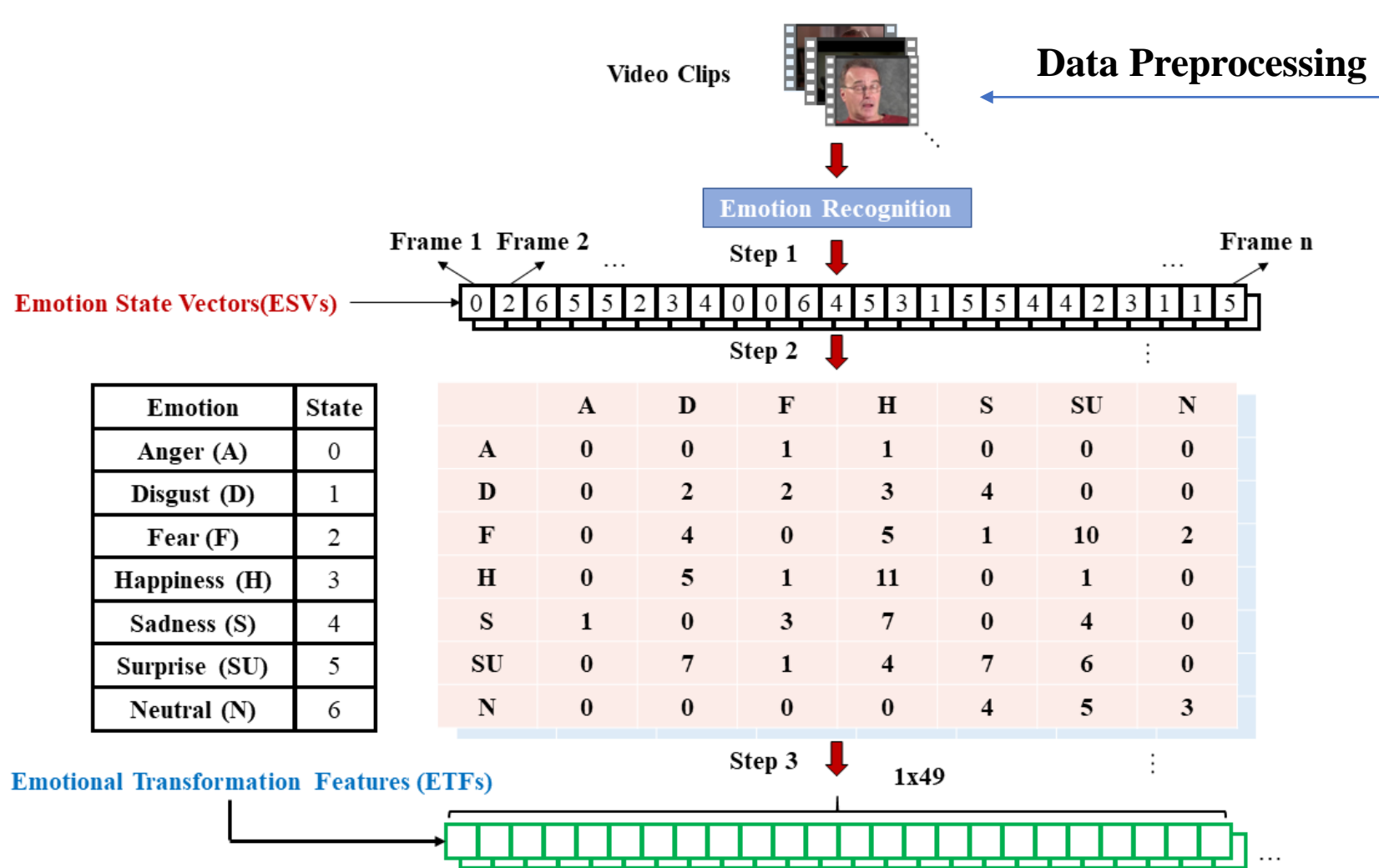
- Existence of bad data, e.g., broken scenes, the face is too small, the face is covered, transition effects

Solution:

- Manually cut and remove the broken clips



Our Proposed method - Emotional transformation feature

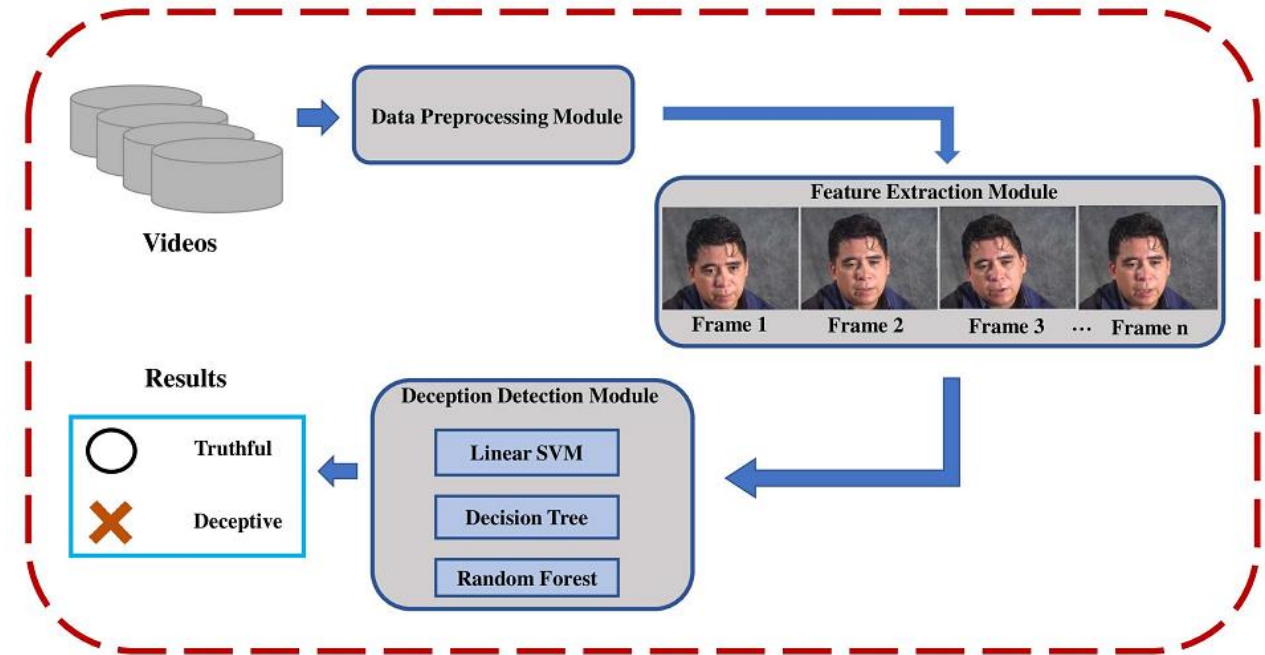


Emotion	State
Anger (A)	0
Disgust (D)	1
Fear (F)	2
Happiness (H)	3
Sadness (S)	4
Surprise (SU)	5
Neutral (N)	6

	A	D	F	H	S	SU	N
A	0	0	1	1	0	0	0
D	0	2	2	3	4	0	0
F	0	4	0	5	1	10	2
H	0	5	1	11	0	1	0
S	1	0	3	7	0	4	0
SU	0	7	1	4	7	6	0
N	0	0	0	0	4	5	3

Experiments and analysis

- Performance on several classifiers, including: decision tree (DT), random forest (RF), k-nearest neighbor (kNN), support vector machine(SVM)
- Parameter settings:
 - maximum depth of DT is 5
 - number of tree in RF is 100
 - number of neighbors is 3
 - penalty weight of SVM is 1
- 10-fold cross validation
- Comparison of accuracy with latest works



Experiments and analysis --- Parameters & settings

Papers	Features	SVM	DT	RF	kNN
Pérez-Rosas et al. [12]	Unigrams	69.49%	76.27%	67.79%	-
	Psycholinguistic	53.38%	50.00%	66.10%	-
	Syntactic Complexity	52.54%	62.71%	53.38%	-
	Facial Displays	78.81%	74.57%	67.79%	-
	Hand Gestures	59.32%	57.62%	57.62%	-
	All Features Combined	77.11%	69.49%	73.72%	-
Karimi et al. [21]	DEV-vocal	-	-	-	74.16%
	DEV-visual	-	-	-	75.00%
	DEV-Hybrid	-	-	-	84.16%
Our proposed method	ETF	65.00%	68.46%	71.15%	61.15%
	ETF + Facial Displays + Hand Gestures	87.59%	76.56%	81.66%	82.49%

Table 1. The comparison of the performance.

Conclusions

- An emotional transformation based feature is proposed.
- The performance of the proposed feature, based on only visual information, is competitive with the latest work utilizing multimodality information.
- Data collection of different occasions are suggested.

Thank You!