Web Video Verification using Contextual Cues

Olga Papadopoulou, **Markos Zampoglou**, Symeon Papadopoulos, Yiannis Kompatsiaris

Centre for Research and Technology Hellas (CERTH) – Information Technologies Institute (ITI), Thessaloniki Greece

2nd Multimedia Forensics and Security Workshop, MFSec2017 @ ICMR 2017 | 6 June 2017 | Bucharest, Romania

Fake videos in news reporting

- User-Generated Content (UGC): essential in news reporting
- Important to verify content and filter out fakes
- Types of "fake":
  1. **Staged videos:** actors perform scripted actions under direction, published as UGC
  2. **Out-of-context videos:** the context of the depicted events is misrepresented (e.g. the claimed video location is wrong)
  3. **Past videos:** presented as UGC from breaking events
  4. **Tampered videos:** visual or audio content has been altered through editing
  5. **Synthetic videos:** contain Computer-generated Imagery (CGI) posing as real

“Hezbollah sniper kills ISIS”

Fake videos


22.03.2016


LiveLeak
22/03/2016


BBC
TRENDING

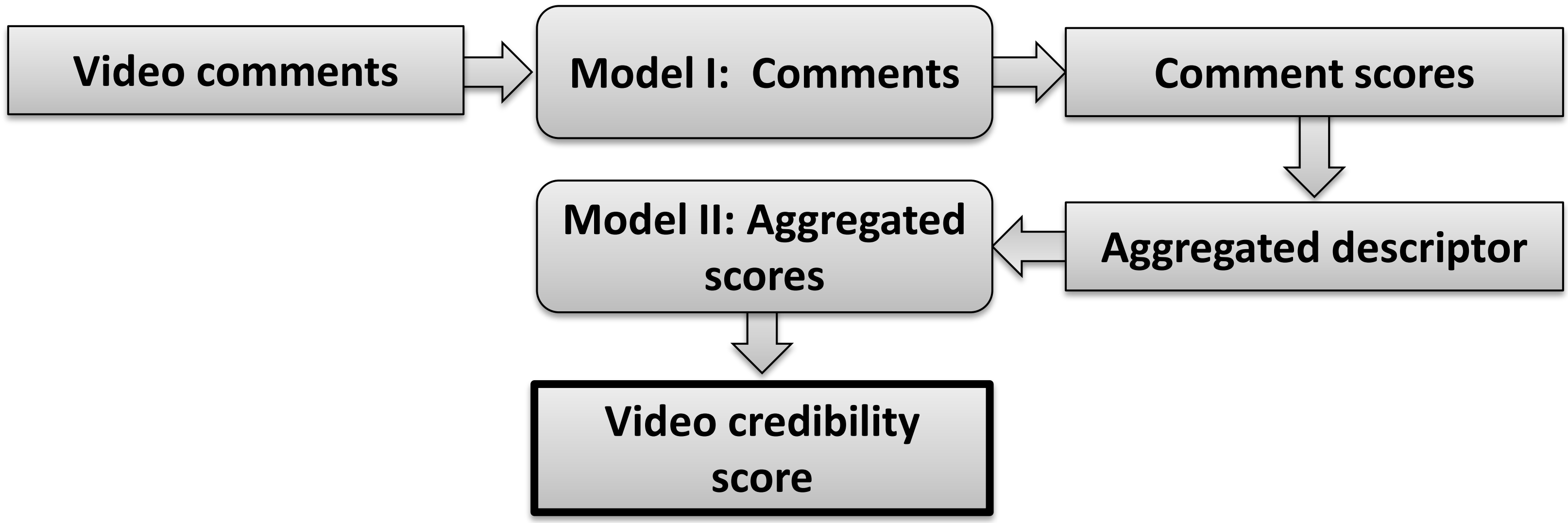

Washington
11:48 AM ET
THE POPE COMES TO AMERICA
POPE FRANCIS ARRIVES FOR U.S. BISHOPS' MEETING
LIVE
CNN
AT THIS HOUR

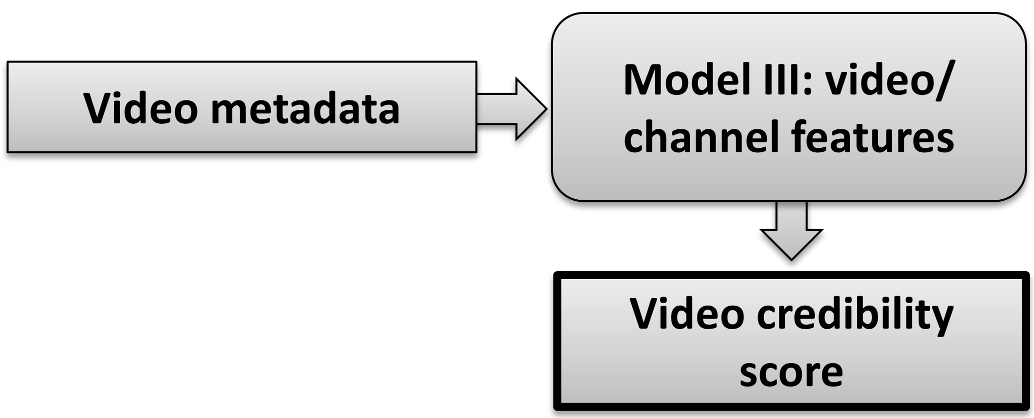
Real UGC videos

Detecting fake videos

- Tampering detection can only catch a subset of real-world fakes
- We have to exploit contextual information
  - Video metadata (video description & channel features)
  - Comment features
- Existing approaches for Tweet verification (Gupta 2013, Boididou et al, 2015 & 2016)
  - Tweet text features
  - Tweet context features (shares, retweets, etc.)
  - User profile features

Approach


Video comments
Model I: Comments
Comment scores
Aggregated descriptor
Model II: Aggregated scores
Video credibility score


Video metadata
Model III: video/ channel features
Video credibility score

Comment-based classification

#	Feature description
01	Text length
02	Number of words
03-04	Contains question/exclamation mark (Boolean)
05-06	Contains happy/sad emoticon (Boolean)
07-09	Contains 1st/2nd/3rd person pronoun (Boolean)
10	Number of uppercase characters
11-12	Number of positive/negative sentiment words
13	Number of slang words
14-15	Has ':' symbol/'please' (Boolean)
16-17	Number of question/exclamation marks
18	Readability score

- Features applicable to both tweets and comments
- Model I trained on a corpus of fake/real tweets
- Each video comment assigned a score
- Video-level aggregate descriptor formed as a 10-bin score histogram
- Model II trained on a Video Corpus & used to evaluate video credibility

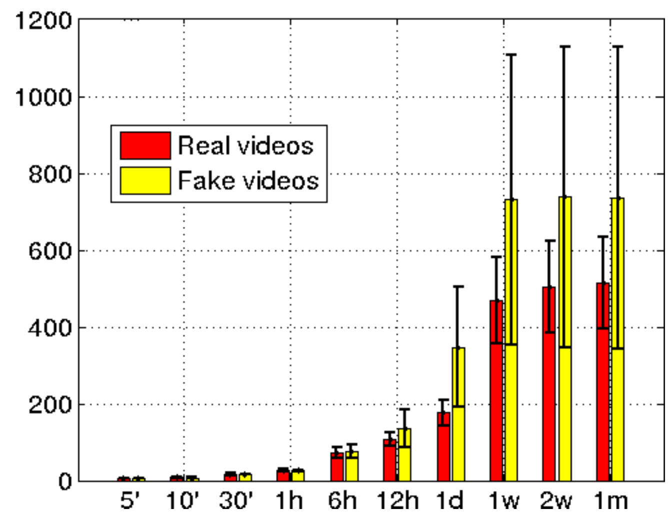
Metadata-based classification

#	Feature description
From channel description	
01	Channel view count
02	Channel comment count
03	Channel subscriber count
04	Channel video count
From video description	
05	Text length
06	Number of words
07-08	Contains question/exclamation mark (Boolean)
09-10	Contains 1st/3rd person pronoun (Boolean)
11	Number of uppercase characters
12-13	Number of positive/negative sentiment words
14	Number of slang words
15	Has ':' symbol (Boolean)
16-17	Number of question/exclamation marks

- Features extracted from channel and video description
- Model III trained on a Video Corpus
- Used to evaluate video credibility

Datasets

- Image Verification Corpus[1]
  - 17,857 tweets with images (7,229 real, 10,628 fake)
- Video Verification Corpus[2]
  - 104 YouTube videos (55 fake, 49 real)



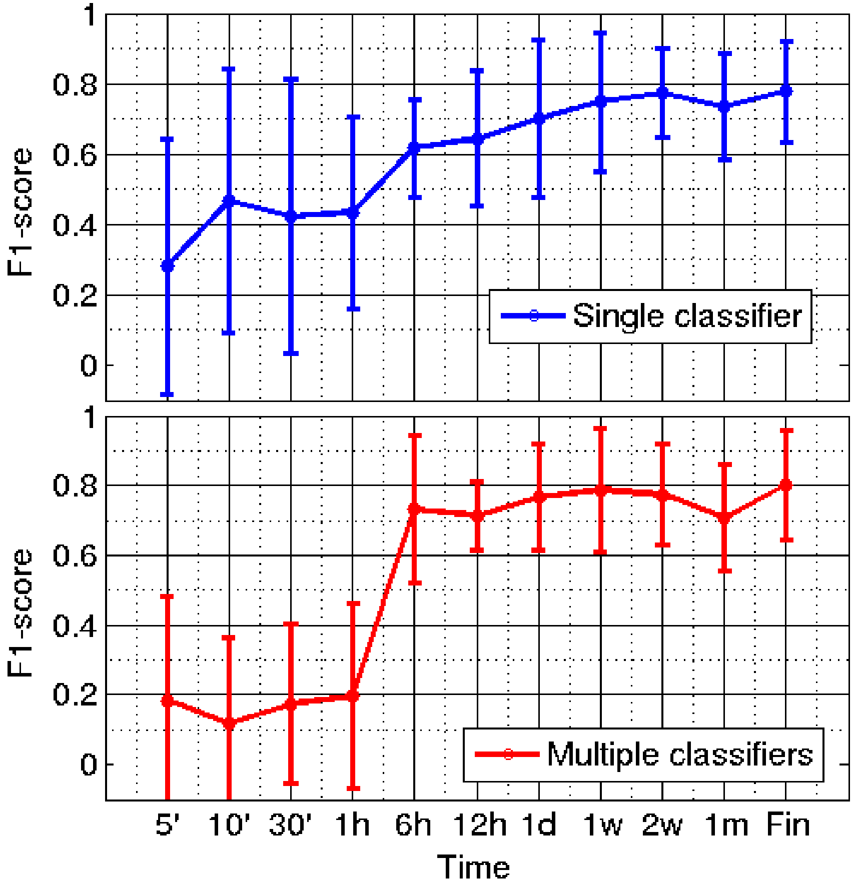
Mean number of video comments through time

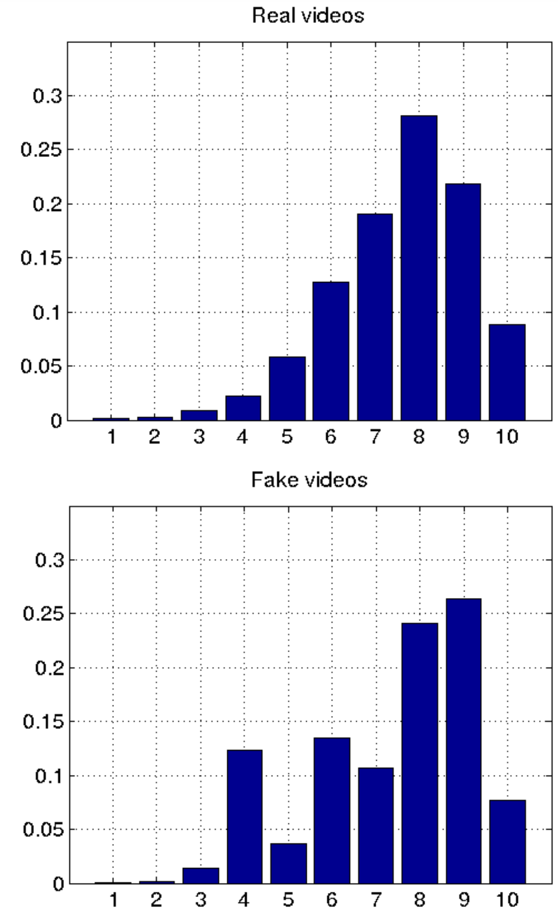
[1]https://github.com/MKLab-ITI/image-verification-corpus
[2]https://github.com/MKLab-ITI/contextual-video-verification

Experimental Results

	Precision	Recall	F1
Comments	0.88	0.74	0.79
Video metadata	0.88	0.79	0.82
Ideal fusion	1.00	0.83	0.90

Result analysis





Distribution of comment credibility estimate scores

Conclusions

- The two classifiers do not fully overlap
  - There is potential for a fusion method
- Comment-based classification requires ~6 hours after video posting to be practical
- Using different classifiers per timeframe does not yield an advantage
- The comment credibility value distributions differ significantly between real and fake videos
  - However, the actual meaning of the values is hard to interpret

Future work

- Extend the dataset
  - Ongoing work: challenging to increase the scale
- Extension to Facebook, Dailymotion, Vimeo
  - Direct correspondence between YouTube features and features in other platforms
  - Relative difficulty of finding fakes in other platforms
- Classifier fusion
  - Need for larger-scale dataset
- Explore the role of features in classification
  - Especially comment credibility scores

References

- Christina Boididou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, Stuart E. Middleton, Andreas Petlund, and Yiannis Kom-patsiaris. 2016. Verifying Multimedia Use at MediaEval 2016. In MediaEval, Vol. 1739. CEUR-WS.org.
- Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, and Yiannis Kompatsiaris. 2015. Verifying Multimedia Use at MediaEval 2015. In MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany.
- Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: characterizing and identifying fake images on Twitter dur-ing Hurricane Sandy. In 22nd International World Wide Web Conference, WWW ’13. ACM, 729–736.
- Symeon Papadopoulos, Markos Zampoglou, Ioannis Kompatsiaris, and Denis Teyssou. 2017. InVID Fake Video Corpus. (Jan. 2017). DOI:https://doi.org/10.5281/zenodo.242481
- Markos Zampoglou, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2017. A Large-Scale Evaluation of Splicing Localization Algorithms for Web Images. Multimedia Tools and Applications 76, 4 (February 2017), 4801—4834.

Thank you!

Get in touch:

markzampoglou@iti.gr

papadop@iti.gr

olgapapa@iti.gr

http://caa.iti.gr/

http://www.invid-project.eu/