Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie
AGH University of Science
and Technology

User-Generated Content (UGC)/In-The-Wild Video Content Recognition[1]

Mikołaj Leszczuk, Lucjan Janowski, Jakub Nawała, Michał Grega
qoe@agh.edu.pl

AGH University of Science and Technology

[1] Supported by Polish National Centre for Research and Development (TANGO-IV-A/0038/2019-00).

Outline

1. Introduction

Outline

Outline

Outline

Outline

Introduction

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.
2. IP video traffic generated by users is largely related to user-generated content (UGC).

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.
2. IP video traffic generated by users is largely related to user-generated content (UGC).
3. Although at the beginning of the creation of UGC, this content was often characterised by amateur acquisition conditions and unprofessional processing, the development of widely available knowledge and affordable equipment allows one to create UGC of a quality practically indistinguishable from professional content.

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.
2. IP video traffic generated by users is largely related to user-generated content (UGC).
3. Although at the beginning of the creation of UGC, this content was often characterised by amateur acquisition conditions and unprofessional processing, the development of widely available knowledge and affordable equipment allows one to create UGC of a quality practically indistinguishable from professional content.
4. Since some UGC content is indistinguishable from professional content, we are not interested in all UGC content, but only in the quality that clearly differs from professional.

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.
2. IP video traffic generated by users is largely related to user-generated content (UGC).
3. Although at the beginning of the creation of UGC, this content was often characterised by amateur acquisition conditions and unprofessional processing, the development of widely available knowledge and affordable equipment allows one to create UGC of a quality practically indistinguishable from professional content.
4. Since some UGC content is indistinguishable from professional content, we are not interested in all UGC content, but only in the quality that clearly differs from professional.
5. For this content, we use the term “in the wild” as a concept closely related to the concept of UGC, which is its special case.

Introduction

1. According to Cisco, we are facing a three-fold increase in IP traffic in five years, ranging from 2017 to 2022.
2. IP video traffic generated by users is largely related to user-generated content (UGC).
3. Although at the beginning of the creation of UGC, this content was often characterised by amateur acquisition conditions and unprofessional processing, the development of widely available knowledge and affordable equipment allows one to create UGC of a quality practically indistinguishable from professional content.
4. Since some UGC content is indistinguishable from professional content, we are not interested in all UGC content, but only in the quality that clearly differs from professional.
5. For this content, we use the term "in the wild" as a concept closely related to the concept of UGC, which is its special case.
6. In this paper, we show that it is possible to introduce the new concept of an objective "in-the-wild" video content recognition model.

Introduction

(a) Professional content with no “in-the-wild” content

(b) Professional content with “in-the-wild” content displayed in small area

(c) Professional content with “in-the-wild” content displayed in large area

(d) “In-the-wild” content with professional content mixed in large area

(e) “In-the-wild” content with professional content mixed in small area

(f) “In-the-wild” content with no professional content

AGH

www.agh.edu.pl

The Full Sets

The Full Sets

1. “In-the-wild” video quality content:

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),
   3. KoNViD-1k (2017),

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),
   3. KoNViD-1k (2017),
   4. LIVE-VQC (2018),

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),
   3. KoNViD-1k (2017),
   4. LIVE-VQC (2018),
   5. YouTube-UGC (2019) and

The Full Sets

AGH

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),
   3. KoNViD-1k (2017),
   4. LIVE-VQC (2018),
   5. YouTube-UGC (2019) and
   6. LSVQ (2021).

The Full Sets

1. "In-the-wild" video quality content:
    1. CVD-2014 (2014),
    2. LIVE-Qualcomm (2016),
    3. KoNViD-1k (2017),
    4. LIVE-VQC (2018),
    5. YouTube-UGC (2019) and
    6. LSVQ (2021).
2. Professional video quality content:

The Full Sets

1. "In-the-wild" video quality content:
   1. CVD-2014 (2014),
   2. LIVE-Qualcomm (2016),
   3. KoNViD-1k (2017),
   4. LIVE-VQC (2018),
   5. YouTube-UGC (2019) and
   6. LSVQ (2021).
2. Professional video quality content:
   1. "NTIA simulated news".

The Sub Sets

The Sub Sets

1. The number of video frames in all databases used is hundreds of thousands, but in reality the frames belonging to one shot are quite similar to each other.

The Sub Sets

1. The number of video frames in all databases used is hundreds of thousands, but in reality the frames belonging to one shot are quite similar to each other.
2. Therefore, in further analysis, we operate at the level of recognising the whole shot.

The Sub Sets

AGH

1. The number of video frames in all databases used is hundreds of thousands, but in reality the frames belonging to one shot are quite similar to each other.
2. Therefore, in further analysis, we operate at the level of recognising the whole shot.
3. For databases that are not delivered per shot, these shots are detected using PySceneDetect (a command-line application and a Python library to detect shot changes in video sequences and automatically split the video into separate clips) using its default parameters.

The Sub Sets

1. The number of video frames in all databases used is hundreds of thousands, but in reality the frames belonging to one shot are quite similar to each other.
2. Therefore, in further analysis, we operate at the level of recognising the whole shot.
3. For databases that are not delivered per shot, these shots are detected using PySceneDetect (a command-line application and a Python library to detect shot changes in video sequences and automatically split the video into separate clips) using its default parameters.
4. As a result, we get 68 shots with professional content and 2 169 shots with “in the wild” content.

Video Indicators

#	Name
1	Blockiness
2	Spatial Activity (SA)
3	Block Loss
4	Blur
5	Temporal Activity (TA)
6	Exposure
7	Contrast
8	Noise
9	Slicing
10	Flickering

Table: The list of video indicators which are used in the experiment.

Video Indicators

1. As the frames belonging to one shot are quite similar to each other, they have similar values for the video indicators.

#	Name
1	Blockiness
2	Spatial Activity (SA)
3	Block Loss
4	Blur
5	Temporal Activity (TA)
6	Exposure
7	Contrast
8	Noise
9	Slicing
10	Flickering

Table: The list of video indicators which are used in the experiment.

Video Indicators

1. As the frames belonging to one shot are quite similar to each other, they have similar values for the video indicators.
2. Consequently, the experiment operates on averaged video shots.

#	Name
1	Blockiness
2	Spatial Activity (SA)
3	Block Loss
4	Blur
5	Temporal Activity (TA)
6	Exposure
7	Contrast
8	Noise
9	Slicing
10	Flickering

Table: The list of video indicators which are used in the experiment.

Video Indicators

1. As the frames belonging to one shot are quite similar to each other, they have similar values for the video indicators.
2. Consequently, the experiment operates on averaged video shots.
3. In total, we used 10 video indicators.

#	Name
1	Blockiness
2	Spatial Activity (SA)
3	Block Loss
4	Blur
5	Temporal Activity (TA)
6	Exposure
7	Contrast
8	Noise
9	Slicing
10	Flickering

Table: The list of video indicators which are used in the experiment.

Video Indicators

1. As the frames belonging to one shot are quite similar to each other, they have similar values for the video indicators.
2. Consequently, the experiment operates on averaged video shots.
3. In total, we used 10 video indicators.
4. They come from our AGH Video Quality (VQ) team.

#	Name
1	Blockiness
2	Spatial Activity (SA)
3	Block Loss
4	Blur
5	Temporal Activity (TA)
6	Exposure
7	Contrast
8	Noise
9	Slicing
10	Flickering

Table: The list of video indicators which are used in the experiment.

Modelling

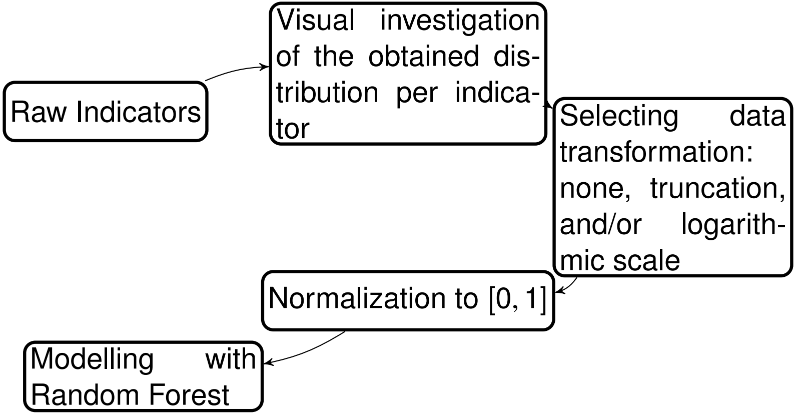


Figure: Data preparation and analysis schema.

Modelling

www.agh.edu.pl

#	Name	Truncation	Log	min	max
1	Blockiness	[0.5, 1.25]	no	0.5	1.25
2	Spatial Activity (SA)	[0, 200]	no	0	200
3	Block Loss	none	yes	0	3.33
4	Blur	[0, 22]	yes	0	1.36
5	Temporal Activity (TA)	[0, 75]	no	0	75
6	Exposure	none	no	9	222
7	Contrast	none	no	0	104
8	Noise	none	yes	0	1.79
9	Slicing	none	yes	0.16	4.22
10	Flickering	none	no	0	1

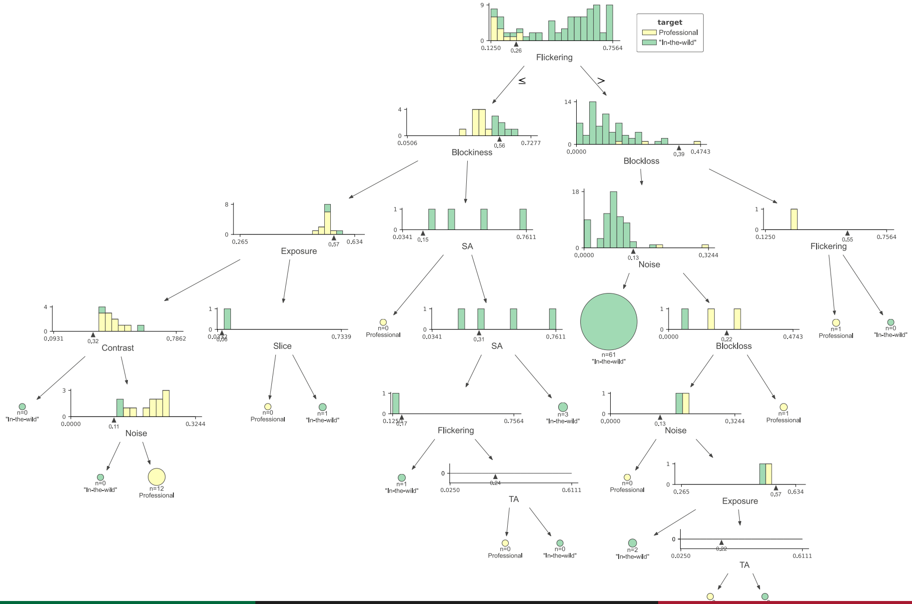
Table: Normalisation procedure for each indicator.

Results

	Accuracy	Precision	Recall	F-measure
Test set	0.956	0.976	0.971	0.973
“In-the-wild” set	0.975	1.000	0.975	0.987

Table: Decision tree results received for “in-the-wild” content recognition.

Results


target
Professional
"In-the-wild"
Flickering
≤
>
Blockiness
Blockloss
Exposure
SA
Noise
Flickering
Contrast
Slice
n=0
Professional
SA
n=61
"In-the-wild"
Blockloss
n=1
Professional
n=0
"In-the-wild"
Noise
Flickering
n=3
"In-the-wild"
Noise
n=12
Professional
TA
Exposure
n=2
"In-the-wild"
TA

Results

	Accuracy	Precision	Recall	F-measure
Test set	0.980	0.983	0.994	0.988
“In-the-wild” set	0.994	1.000	0.994	0.997

Table: Random forest results received for “in-the-wild” content recognition.

Conclusions and Further Work

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new "in-the-wild" concept of an objective video content recognition model.

Conclusions and Further Work

AGH

1. In this paper, we show that it is possible to introduce the new "in-the-wild" concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.

www.agh.edu.pl

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new "in-the-wild" concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.
3. These modelling results are obtained when the random forest learning method is used.

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new "in-the-wild" concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.
3. These modelling results are obtained when the random forest learning method is used.
4. However, it should be noted that the use of the more explainable decision tree learning method does not cause a significant decrease in prediction accuracy (measure F of 0.973).

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new “in-the-wild” concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.
3. These modelling results are obtained when the random forest learning method is used.
4. However, it should be noted that the use of the more explainable decision tree learning method does not cause a significant decrease in prediction accuracy (measure F of 0.973).
5. The results presented are work in progress.

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new "in-the-wild" concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.
3. These modelling results are obtained when the random forest learning method is used.
4. However, it should be noted that the use of the more explainable decision tree learning method does not cause a significant decrease in prediction accuracy (measure F of 0.973).
5. The results presented are work in progress.
6. While the current results are highly promising, they still require additional validation since training and test data sets are relatively limited (especially for professional content).

Conclusions and Further Work

1. In this paper, we show that it is possible to introduce the new “in-the-wild” concept of an objective video content recognition model.
2. The value of the accuracy measured for a model (the parameter of the F measure) achieved is 0.988.
3. These modelling results are obtained when the random forest learning method is used.
4. However, it should be noted that the use of the more explainable decision tree learning method does not cause a significant decrease in prediction accuracy (measure F of 0.973).
5. The results presented are work in progress.
6. While the current results are highly promising, they still require additional validation since training and test data sets are relatively limited (especially for professional content).
7. Therefore, additional selected video sequences from the collected database of 6000+ professional YouTube news clips should be used.

Publication

Mikołaj Leszczuk, Lucjan Janowski, Jakub Nawała, and Michał Grega, "User-Generated Content (UGC)/In-The-Wild Video Content Recognition", 14th Asian Conference on Intelligent Information and Database Systems, November 28-30, 2022, Ho Chi Minh City, Vietnam