Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

AGH University of Science
and Technology

Objective Video Quality Assessment Method for Face Recognition Tasks

Mikołaj Leszczuk[1], Lucjan Janowski[1], Jakub Nawała[1], Atanas Boev[2]

qoe@agh.edu.pl

[1]AGH University of Science and Technology

[2]Huawei Technologies Dusseldorf GmbH

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Introduction

Introduction

1. Nowadays, there are many metrics for overall Quality of Experience (QoE), both those with Full Reference (FR), such as the peak signal–to–noise ratio (PSNR) or structural similarity (SSIM), and those with No Reference (NR), such as Video Quality Indicators (VQI), which are successfully used in video processing systems to evaluate videos whose quality is degraded by different processing scenarios.

Introduction

1. Nowadays, there are many metrics for overall Quality of Experience (QoE), both those with Full Reference (FR), such as the peak signal–to–noise ratio (PSNR) or structural similarity (SSIM), and those with No Reference (NR), such as Video Quality Indicators (VQI), which are successfully used in video processing systems to evaluate videos whose quality is degraded by different processing scenarios.
2. However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).

Introduction

AGH

1. Nowadays, there are many metrics for overall Quality of Experience (QoE), both those with Full Reference (FR), such as the peak signal–to–noise ratio (PSNR) or structural similarity (SSIM), and those with No Reference (NR), such as Video Quality Indicators (VQI), which are successfully used in video processing systems to evaluate videos whose quality is degraded by different processing scenarios.
2. However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).
3. Therefore, correctly estimating the performance of the video processing pipeline in both manual and Computer Vision (CV) recognition tasks is still a major research challenge.

Introduction

AGH

1. Nowadays, there are many metrics for overall Quality of Experience (QoE), both those with Full Reference (FR), such as the peak signal–to–noise ratio (PSNR) or structural similarity (SSIM), and those with No Reference (NR), such as Video Quality Indicators (VQI), which are successfully used in video processing systems to evaluate videos whose quality is degraded by different processing scenarios.
2. However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).
3. Therefore, correctly estimating the performance of the video processing pipeline in both manual and Computer Vision (CV) recognition tasks is still a major research challenge.
4. In response to this need, we show in this paper that it is possible to develop the new concept of an objective model for evaluating video quality for face recognition tasks.

Contributions

Contributions

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;

Contributions

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;
2. Set a series of degradation scenarios based on the model of the digital camera and how the luminous flux reflected from the scene will eventually become a digital image;

Contributions

AGH

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;
2. Set a series of degradation scenarios based on the model of the digital camera and how the luminous flux reflected from the scene will eventually become a digital image;
3. Evaluate the resulting degraded images using a face recognition CV library – based on the state-of-the-art Deep Learning dlib software library as well as VQI – eleven (11) of which are from our AGH Video Quality (VQ) team and another eight (8) from external labs;

Contributions

AGH

www.agh.edu.pl

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;
2. Set a series of degradation scenarios based on the model of the digital camera and how the luminous flux reflected from the scene will eventually become a digital image;
3. Evaluate the resulting degraded images using a face recognition CV library – based on the state-of-the-art Deep Learning dlib software library as well as VQI – eleven (11) of which are from our AGH Video Quality (VQ) team and another eight (8) from external labs;
4. Develop a new concept for an objective model to evaluate video quality for face recognition tasks;

Contributions

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;
2. Set a series of degradation scenarios based on the model of the digital camera and how the luminous flux reflected from the scene will eventually become a digital image;
3. Evaluate the resulting degraded images using a face recognition CV library – based on the state-of-the-art Deep Learning dlib software library as well as VQI – eleven (11) of which are from our AGH Video Quality (VQ) team and another eight (8) from external labs;
4. Develop a new concept for an objective model to evaluate video quality for face recognition tasks;
5. Train, test, and validate the model;

Contributions

1. Collect a representative set of image sequences – based on the subset of the ”Labelled Faces in the Wild (LFW)” database;
2. Set a series of degradation scenarios based on the model of the digital camera and how the luminous flux reflected from the scene will eventually become a digital image;
3. Evaluate the resulting degraded images using a face recognition CV library – based on the state-of-the-art Deep Learning dlib software library as well as VQI – eleven (11) of which are from our AGH Video Quality (VQ) team and another eight (8) from external labs;
4. Develop a new concept for an objective model to evaluate video quality for face recognition tasks;
5. Train, test, and validate the model;
6. Show that it is possible to achieve a measure of model accuracy, expressed as the value of the F-measure parameter, of 0.87.

The Face Recognition Set

Figure: Example image № 0002 of Aaron Sorkin from LFW database

The Face Recognition Set

1. The source of the full data set for Face Recognition is the Labelled Faces in the Wild (LFW) database.

Figure: Example image № 0002 of Aaron Sorkin from LFW database

The Face Recognition Set

1. The source of the full data set for Face Recognition is the Labelled Faces in the Wild (LFW) database.
2. LFW is a public benchmark for face verification, also known as pair matching.

Figure: Example image № 0002 of Aaron Sorkin from LFW database

The Face Recognition Set

1. The source of the full data set for Face Recognition is the Labelled Faces in the Wild (LFW) database.
2. LFW is a public benchmark for face verification, also known as pair matching.
3. The data set contains 13,233 images of the faces of 5,749 different people collected on the Web.

Figure: Example image № 0002 of Aaron Sorkin from LFW database

The Face Recognition Set

1. The source of the full data set for Face Recognition is the Labelled Faces in the Wild (LFW) database.
2. LFW is a public benchmark for face verification, also known as pair matching.
3. The data set contains 13,233 images of the faces of 5,749 different people collected on the Web.
4. Each face has a resolution of 250×250 and is labelled with the name of the person pictured.

Figure: Example image № 0002 of Aaron Sorkin from LFW database

The Face Recognition Subset

Figure: The montage of selected SRC frames (from LFW database) for face recognition

The Face Recognition Subset

1. The whole set is subsampled, resulting in 120 images divided into a training set, a test set, and a validation set, in a ratio of 80 vs 20 vs 20, respectively.

Figure: The montage of selected SRC frames (from LFW database) for face recognition

Preparation of Hypothetical Reference Circuits (HRC)

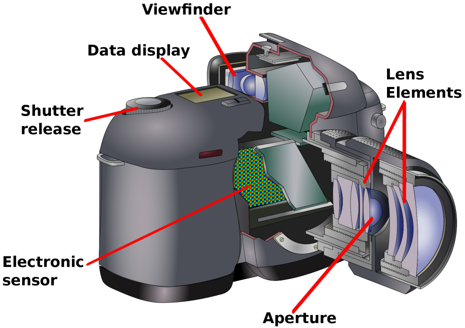


Figure: Diagram of a single-lens reflex camera with basic labels. Based on Reflex camera labels.svg. The author of the original base image is Jean François WITZ. By Astrocog – Own work, CC BY-SA 3.0

Preparation of Hypothetical Reference Circuits (HRC)

1. The HRC set is based on the digital camera model and how the luminous flux reflected from the scene eventually becomes a digital image.

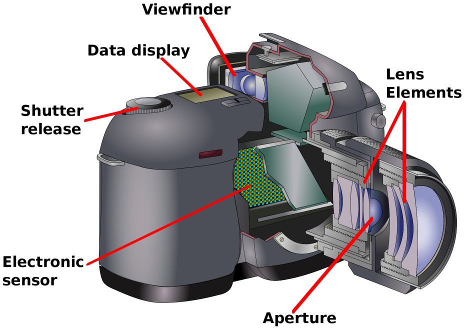
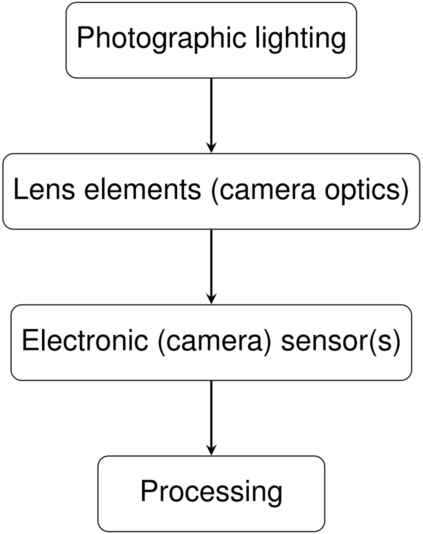


Figure: Diagram of a single-lens reflex camera with basic labels. Based on Reflex camera labels.svg. The author of the original base image is Jean François WITZ. By Astrocog – Own work, CC BY-SA 3.0

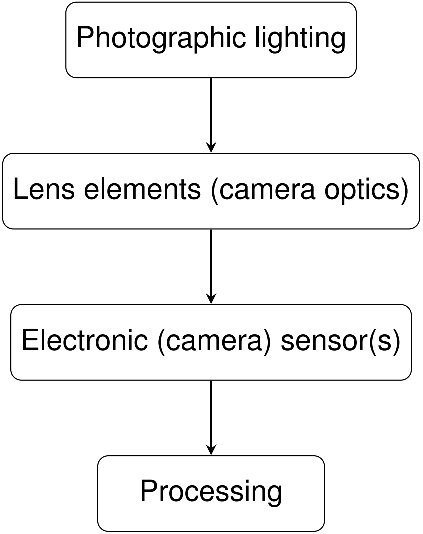
Preparation of Hypothetical Reference Circuits (HRC)



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

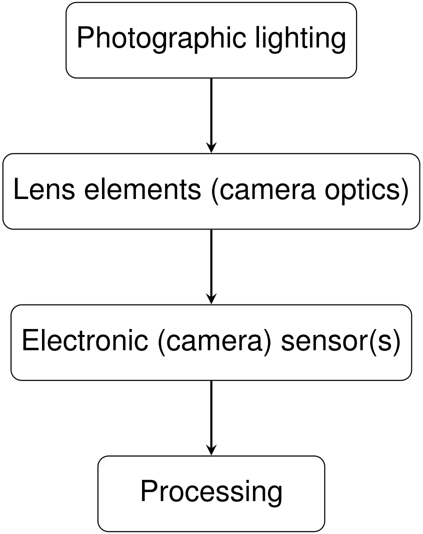
1. HRC related to photographic lighting:



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

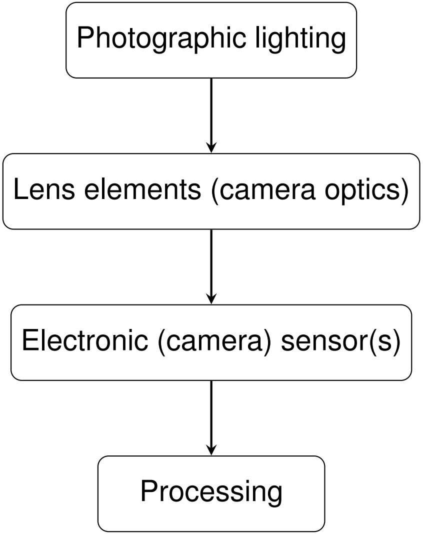
1. HRC related to photographic lighting:
   (1) Image under/overexposure



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):

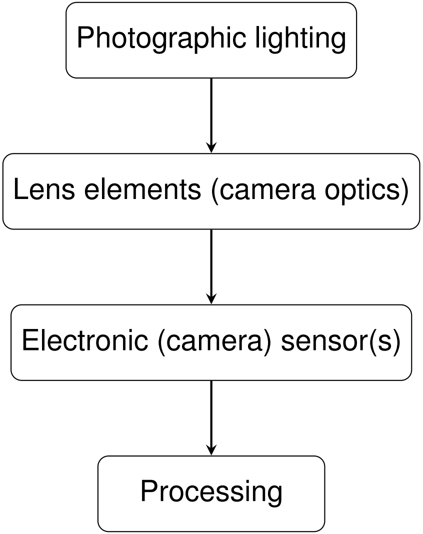


Preparation of Hypothetical Reference Circuits (HRC)

AGH

We select the following HRCs:

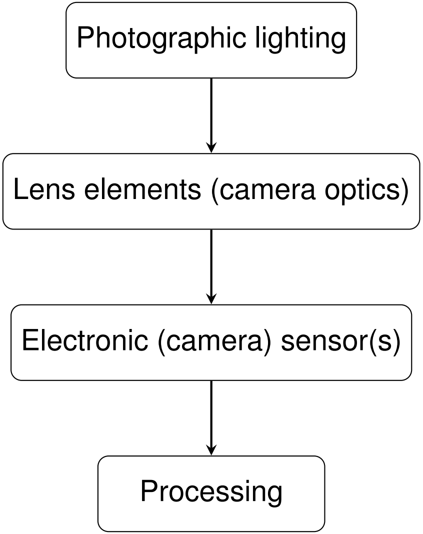
1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

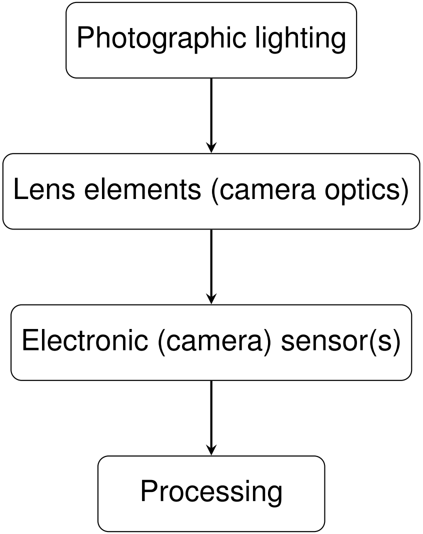
1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)
3. HRC related to electronic (camera) sensor(s):



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

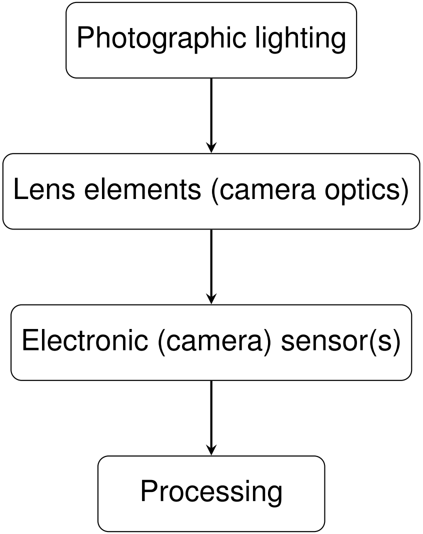
1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)
3. HRC related to electronic (camera) sensor(s):
   (3) Gaussian noise



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

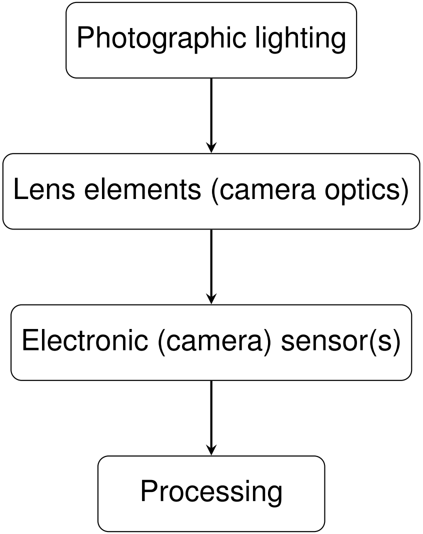
1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)
3. HRC related to electronic (camera) sensor(s):
   (3) Gaussian noise
   (4) Motion blur



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

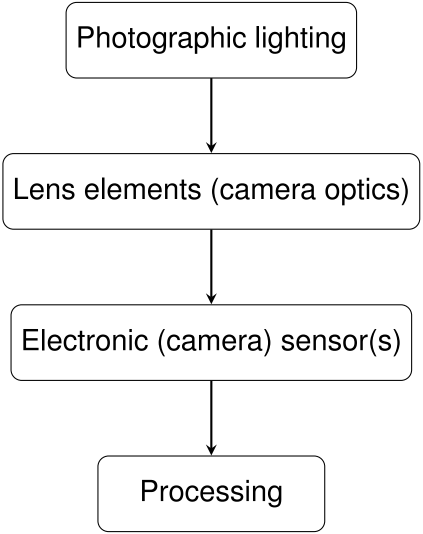
1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)
3. HRC related to electronic (camera) sensor(s):
   (3) Gaussian noise
   (4) Motion blur
4. HRC related to processing:



Preparation of Hypothetical Reference Circuits (HRC)

We select the following HRCs:

1. HRC related to photographic lighting:
   (1) Image under/overexposure
2. HRC related to lens elements (camera optics):
   (2) Defocus (blur)
3. HRC related to electronic (camera) sensor(s):
   (3) Gaussian noise
   (4) Motion blur
4. HRC related to processing:
   (5) JPEG compression



Preparation of Hypothetical Reference Circuits (HRC)

HRC	Unit	Min	Max
Under-Exposure	FFmpeg filter parameter	0	-0.6
Over-Exposure	FFmpeg filter parameter	0	0.6
Defocus (Blur)	ImageMagick filter parameter	0	6
Gaussian Noise	FFmpeg filter parameter	0	48
Motion Blur	ImageMagick filter parameter	0	18
JPEG	ImageMagick filter parameter	0	100

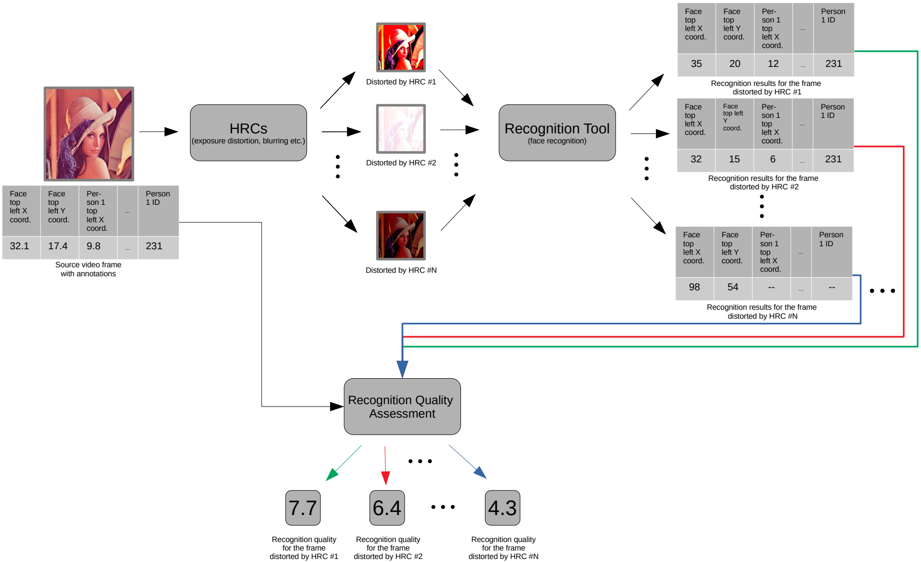
Table: Thresholds for specific Hypothetical Reference Circuits (HRC) – distortions (listed in rows)

Preparation of Hypothetical Reference Circuits (HRC)

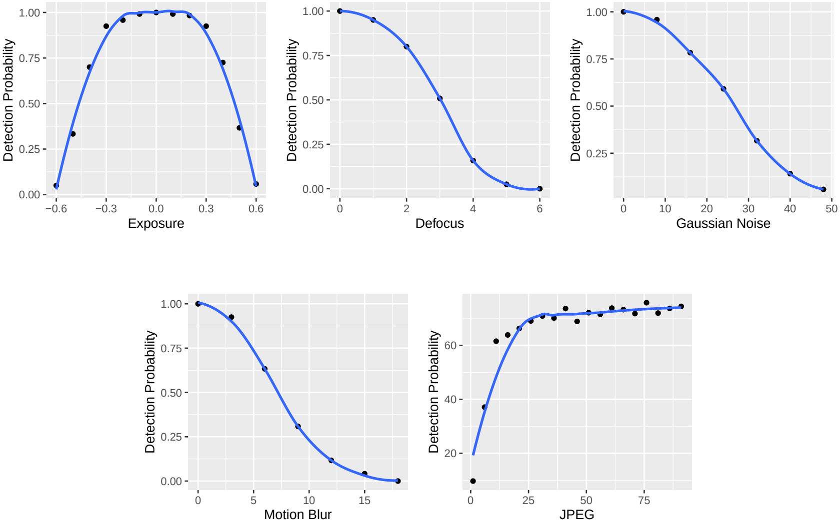
HRC	**#HRC**
Over/Under-Exposure (Photography)	12
Defocus (Blur)	6
Gaussian Noise	6
Motion Blur	6
JPEG	19
Motion Blur + Gaussian Noise	5
Over-Exposure + Gaussian Noise	5
Under-Exposure + Motion Blur	5
#PVS	**6720**

Table: Hypothetical Reference Circuits (HRC) – distortions

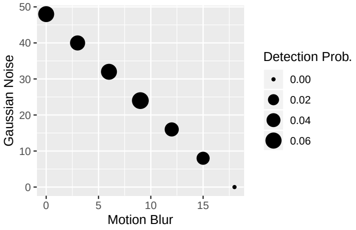
Recognition Experiment Overview

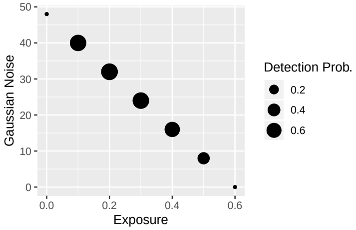


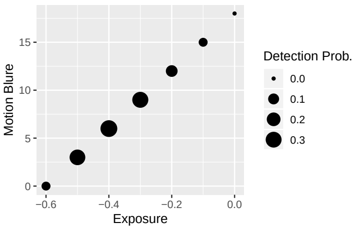
Face Recognition System



Face Recognition System







Face Recognition Time

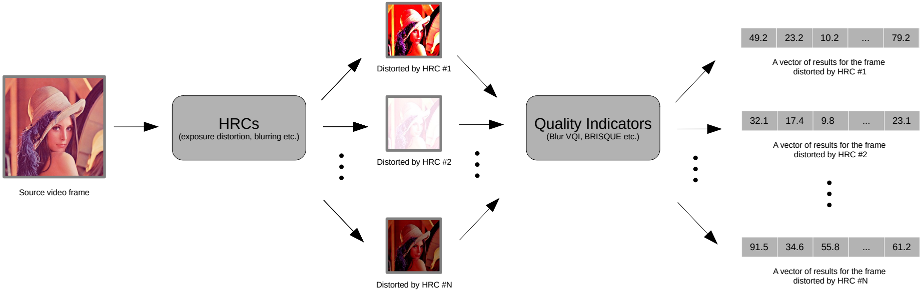
Face Recognition Time

1. The average execution time of the face recognition computer vision algorithm per single video frame is 0.068 s.

Face Recognition Time

1. The average execution time of the face recognition computer vision algorithm per single video frame is 0.068 s.
2. Importantly, execution times are evaluated using a PC with an Intel Core i5-8600K CPU.

Quality Experiment Overview



Indicators

No	Name	Authors	Language
1	Commercial Black		C/C++
2	Blockiness		C/C++
3	Block Loss		C/C++
4	Blur		C/C++
5	Contrast		C/C++
6	Exposure	VQ AGH	C/C++
7	Interlacing		C/C++
8	Noise		C/C++
9	Slicing		C/C++
10	Spatial Activity		C/C++
11	Temporal Activity		C/C++

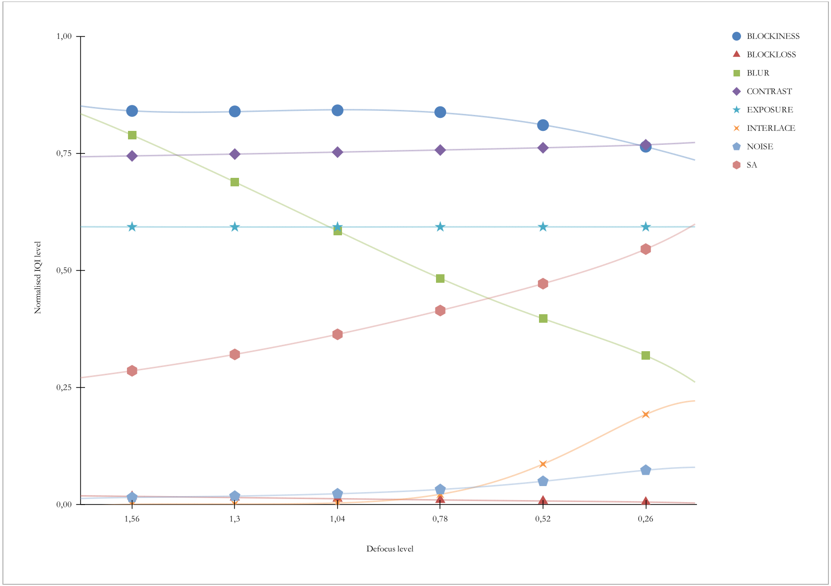
Indicators

No	Name	Authors	Language
12	BIQI	LIVE	MATLAB
13	BRISQUE		MATLAB
14	NIQE		MATLAB
15	OG-IQA		MATLAB
16	FFRIQUEE		MATLAB
17	IL-NIQE		MATLAB
18	CORNIA	UMIACS	MATLAB
19	HOSA	BUPT	MATLAB

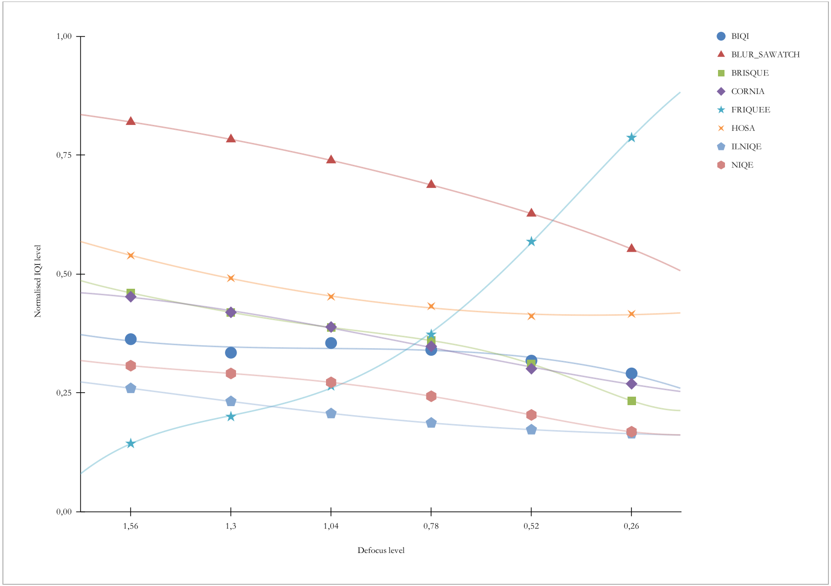
VQI Execution Time

Algorithm	Time [s]
BIQI	1.60
BRISQUE	1.67
NIQE	3.92
OG-IQA	5.72
FRIQUEE	40.79
IL-NIQE	10.70
CORNIA	7.71
HOSA	0.43
VQ AGH VQIs	0.12
Total	72.66

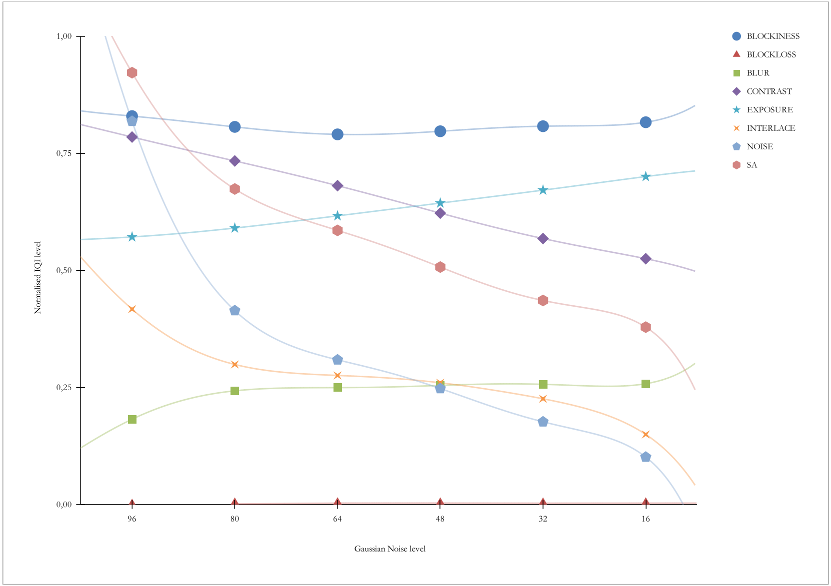
“Our” indicators vs. Defocus [σ/pixels]



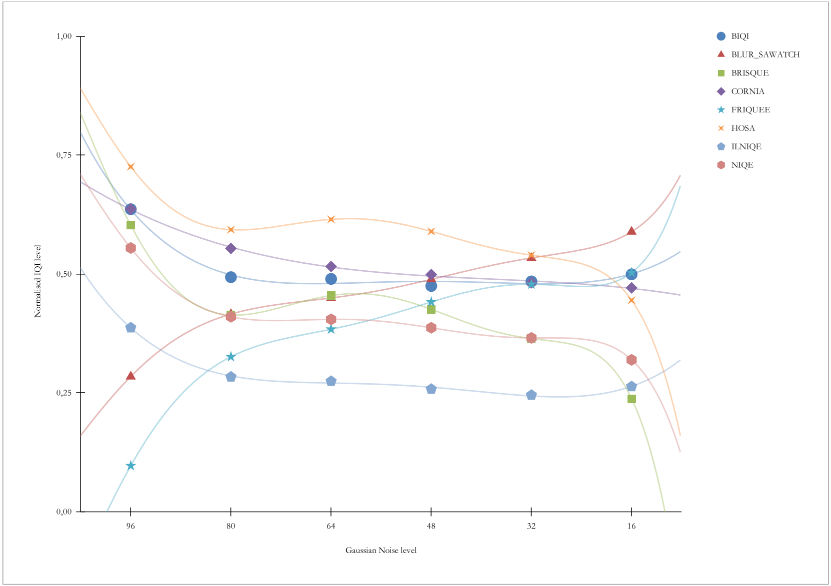
"Other" indicators vs. Defocus [σ/pixels]



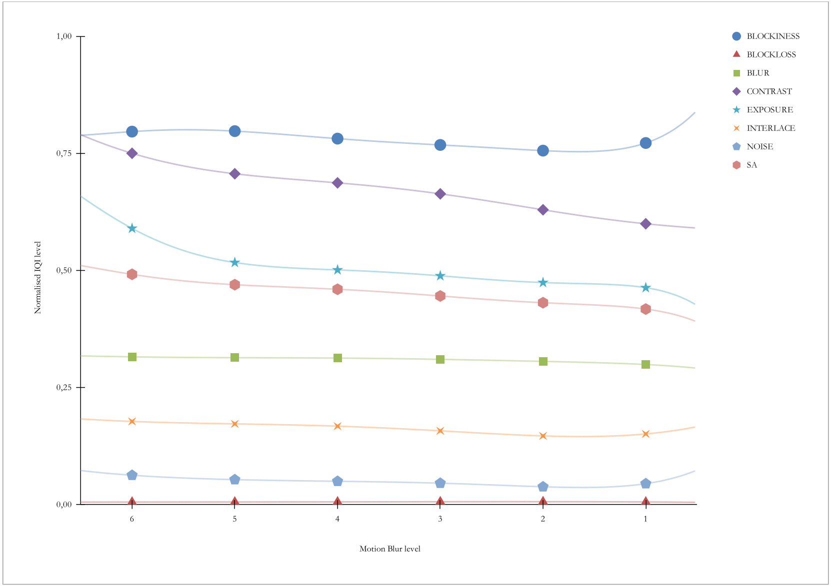
"Our" indicators vs. Gaussian Noise [σ/pixels]



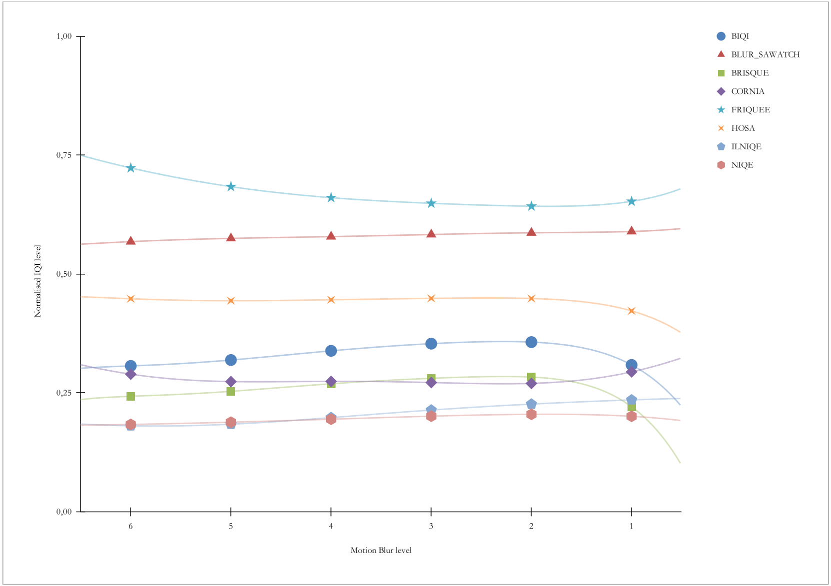
“Other” indicators vs. Gaussian Noise [σ/pixels]



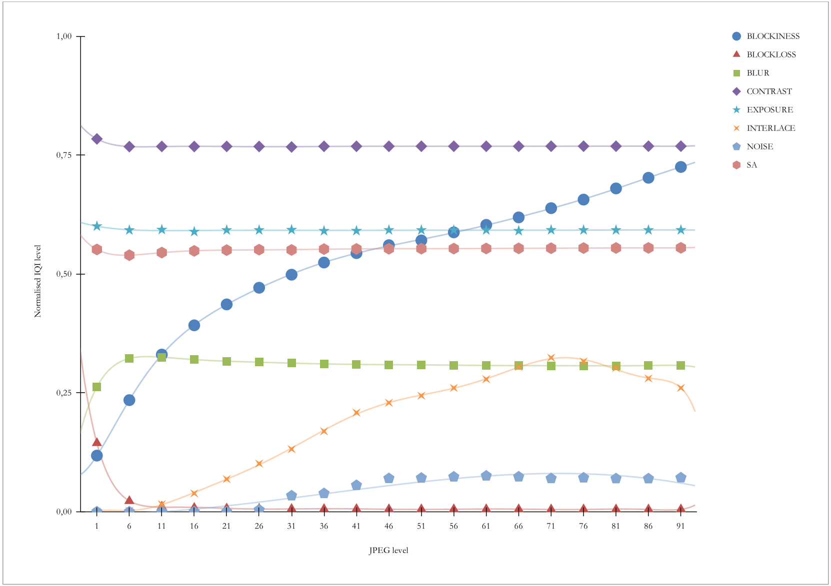
“Our” indicators vs. Motion Blur [σ/degrees]



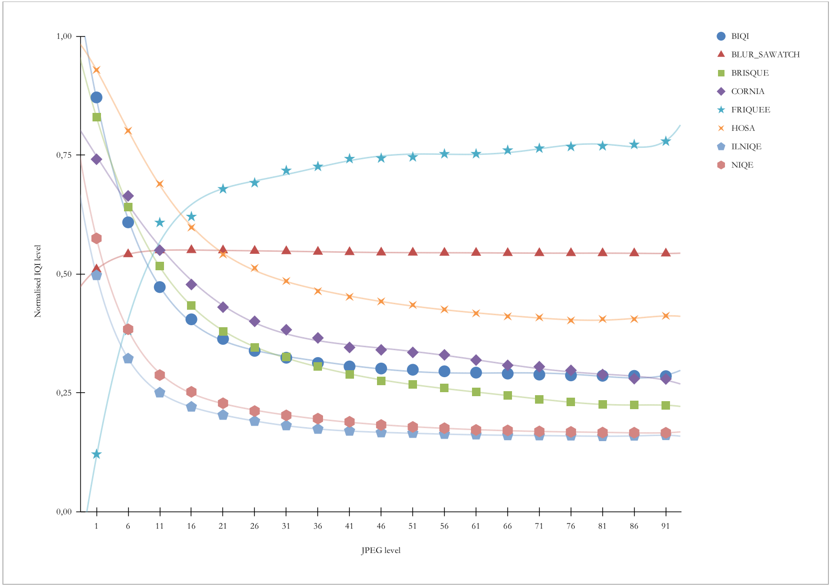
"Other" indicators vs. Motion Blur [σ/degrees]



“Our” indicators vs. JPEG [quality units]



“Other” indicators vs. JPEG [quality units]



Results

	Precision	Recall	F-measure
All metrics	0.893	0.846	0.869
Only ours	0.870	0.791	0.829

Table: General results (Precision, Recall, F-measure) we receive for face recognition, using all metrics and only our ones

Results

All metrics	Algorithm claims it can recognise a face	Algorithm claims the face cannot be recognised
Face was recognised	$tp = 562$	$fn = 102$
Face was not recognised	$fp = 67$	$tn = 439$

Only ours	Algorithm claims it can recognise a face	Algorithm claims the face cannot be recognised
Face was recognised	$tp = 440$	$fn = 116$
Face was not recognised	$fp = 66$	$tn = 548$

Table: Detailed results (true positives – t_p, false positives – f_p, true negatives – t_n, and false negatives – f_n) we receive for face recognition

Results

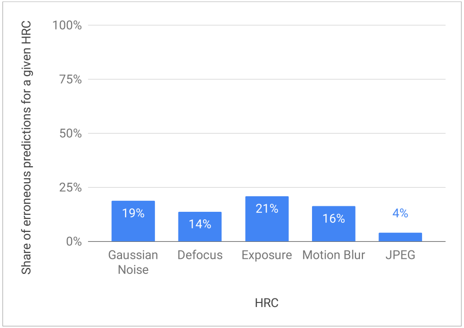


Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in face recognition

Results

1. A more detailed analysis of the results obtained is also carried out.

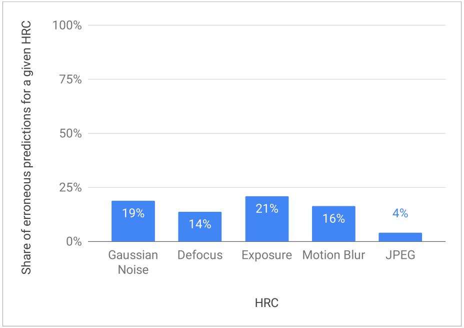


Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in face recognition

Results

1. A more detailed analysis of the results obtained is also carried out.
2. The numerical analysis is to check the sensitivity of the model to individual distortions.

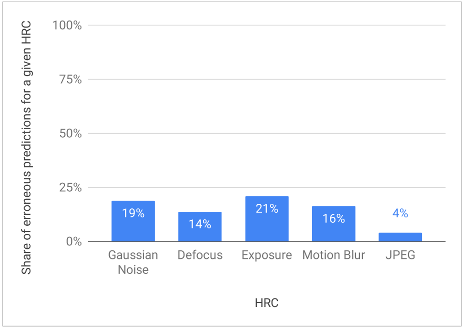


Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in face recognition

Results

1. A more detailed analysis of the results obtained is also carried out.
2. The numerical analysis is to check the sensitivity of the model to individual distortions.
3. As one can see, for the first four HRCs, the model shows a fairly similar error sensitivity – it is wrong in about a dozen or twenty percent of cases.

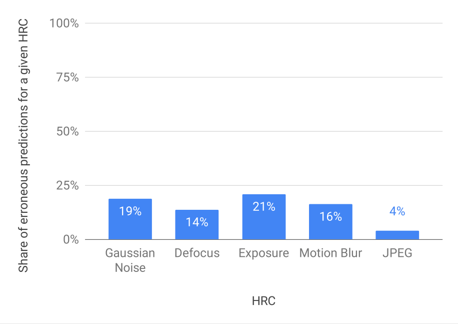


Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in face recognition

Results

1. A more detailed analysis of the results obtained is also carried out.
2. The numerical analysis is to check the sensitivity of the model to individual distortions.
3. As one can see, for the first four HRCs, the model shows a fairly similar error sensitivity – it is wrong in about a dozen or twenty percent of cases.
4. The exception is JPEG HRC, for which the model is much less mistaken – only for 4% of the cases.

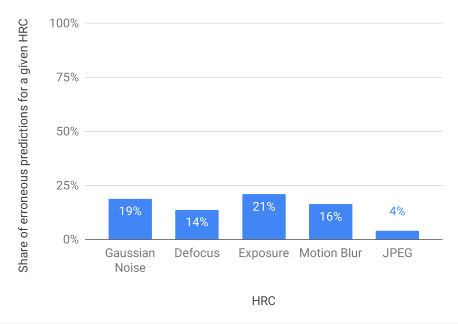


Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in face recognition

Conclusions

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.
3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.
3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.
4. Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.83).

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.
3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.
4. Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.83).
5. It is worth mentioning the most typical problems encountered by the models during their work.

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.
3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.
4. Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.83).
5. It is worth mentioning the most typical problems encountered by the models during their work.
6. Our observations suggest that the characteristics of the initial scene are an important component that misleads the models.

Conclusions

1. We show in this study that the implementation of the new concept of an objective model to evaluate video quality for face recognition tasks is feasible.
2. The achieved value of the model accuracy (F-measure parameter) is 0.87.
3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.
4. Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.83).
5. It is worth mentioning the most typical problems encountered by the models during their work.
6. Our observations suggest that the characteristics of the initial scene are an important component that misleads the models.
7. VQI completely disregards this factor, which has a major impact on the accuracy of recognition.

Publication

AGH

Mikołaj Leszczuk, Lucjan Janowski, Jakub Nawała, and Atanas Boev, "Objective Video Quality Assessment Method for Face Recognition Tasks", Electronics 2022, 11(8), 1167,
https://doi.org/10.3390/electronics11081167