Method for Assessing Objective Video Quality for Automatic License Plate Recognition Tasks

Mikołaj Leszczuk\textsuperscript{1}, Lucjan Janowski\textsuperscript{1}, Jakub Nawała\textsuperscript{1}, Atanas Boev\textsuperscript{2}

qoe@agh.edu.pl

\textsuperscript{1}AGH University of Science and Technology

\textsuperscript{2}Huawei Technologies Dusseldorf GmbH
Outline

1. Introduction

2. Acquisition of the Existing Source Reference Circuits (SRC)

3. Preparation of Hypothetical Reference Circuits (HRC)

4. Recognition Experiment

5. Quality Experiment

6. Results

7. Conclusions
Outline

1. Introduction

2. Acquisition of the Existing Source Reference Circuits (SRC)
Outline

1. Introduction
2. Acquisition of the Existing Source Reference Circuits (SRC)
3. Preparation of Hypothetical Reference Circuits (HRC)
Outline

1. Introduction
2. Acquisition of the Existing Source Reference Circuits (SRC)
3. Preparation of Hypothetical Reference Circuits (HRC)
4. Recognition Experiment
Outline

1. Introduction
2. Acquisition of the Existing Source Reference Circuits (SRC)
3. Preparation of Hypothetical Reference Circuits (HRC)
4. Recognition Experiment
5. Quality Experiment
Outline

1. Introduction
2. Acquisition of the Existing Source Reference Circuits (SRC)
3. Preparation of Hypothetical Reference Circuits (HRC)
4. Recognition Experiment
5. Quality Experiment
6. Results
Outline

1. Introduction
2. Acquisition of the Existing Source Reference Circuits (SRC)
3. Preparation of Hypothetical Reference Circuits (HRC)
4. Recognition Experiment
5. Quality Experiment
6. Results
7. Conclusions
Introduction

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.

However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).

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.

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 Automatic License Plate Recognition (ALPR) tasks.
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).
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.

However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).

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.
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.

However, they are not suitable for video sequences used for recognition tasks (Target Recognition Videos, TRV).

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.

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 Automatic License Plate Recognition (ALPR) tasks.
Acquisition of the Existing Source Reference Circuits (SRC) for ALPR

The ALPR Set

The source of the full data set for ALPR is the CCTV Source Reference Circuits (SRC video sequences) collected at the AGH University of Science and Technology, Krakow, Lesser Poland, by filming parking during high traffic hours.

The data set contains video sequences, containing approximately 15,500 frames in total.

Figure: Frame of the AGH data set for video quality assessment in plate recognition
The source of the full data set for ALPR is the CCTV Source Reference Circuits (SRC video sequences).

Figure: Frame of the AGH data set for video quality assessment in plate recognition
The ALPR Set

1. The source of the full data set for ALPR is the CCTV Source Reference Circuits (SRC video sequences).

2. Collected at the AGH University of Science and Technology, Krakow, Lesser Poland, by filming parking during high traffic hours.

Figure: Frame of the AGH data set for video quality assessment in plate recognition
The ALPR Set

1. The source of the full data set for ALPR is the CCTV Source Reference Circuits (SRC video sequences).

2. Collected at the AGH University of Science and Technology, Krakow, Lesser Poland, by filming parking during high traffic hours.

3. The data set contains video sequences, containing approximately 15,500 frames in total.

Figure: Frame of the AGH data set for video quality assessment in plate recognition
The ALPR Subset

Figure: The montage of selected SRC frames for ALPR
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 for ALPR
Preparation of Hypothetical Reference Circuits (HRC)

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 Astro cog – Own work, CC BY-SA 3.0
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
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
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
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
We select the following HRCs:

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

2. HRC related to lens elements (camera optics):

   Photographic lighting
   Lens elements (camera optics)
   Electronic (camera) sensor(s)
   Processing
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
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
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
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
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:**
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)

<table>
<thead>
<tr>
<th>HRC</th>
<th>Unit</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-Exposure</td>
<td>FFmpeg filter parameter</td>
<td>0</td>
<td>-0.6</td>
</tr>
<tr>
<td>Over-Exposure</td>
<td>FFmpeg filter parameter</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>Defocus (Blur)</td>
<td>ImageMagick filter parameter</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Gaussian Noise</td>
<td>FFmpeg filter parameter</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Motion Blur</td>
<td>ImageMagick filter parameter</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>JPEG</td>
<td>ImageMagick filter parameter</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Distortion Type</th>
<th>#HRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over/Under-Exposure (Photography)</td>
<td>12</td>
</tr>
<tr>
<td>Defocus (Blur)</td>
<td>6</td>
</tr>
<tr>
<td>Gaussian Noise</td>
<td>6</td>
</tr>
<tr>
<td>Motion Blur</td>
<td>6</td>
</tr>
<tr>
<td>JPEG</td>
<td>19</td>
</tr>
<tr>
<td>Motion Blur + Gaussian Noise</td>
<td>5</td>
</tr>
<tr>
<td>Over-Exposure + Gaussian Noise</td>
<td>5</td>
</tr>
<tr>
<td>Under-Exposure + Motion Blur</td>
<td>5</td>
</tr>
<tr>
<td>#PVS</td>
<td>6720</td>
</tr>
</tbody>
</table>
Recognition Experiment Overview

The average execution time of the ALPR computer vision algorithm per single video frame is 0.21 s. Importantly, execution times are evaluated using a PC with an Intel Core i5-8600K CPU.
The average execution time of the ALPR computer vision algorithm per single video frame is 0.21 s.
The average execution time of the ALPR computer vision algorithm per single video frame is 0.21 s.

Importantly, execution times are evaluated using a PC with an Intel Core i5-8600K CPU.
Quality Experiment Overview

Source video frame

HRCs (exposure distortion, blurring etc.)
Distorted by HRC #1
Distorted by HRC #2
Distorted by HRC #N

Quality Indicators (Blur VQI, BRISQUE etc.)

49.2 23.2 10.2 ...

A vector of results for the frame distorted by HRC #1

32.1 17.4 9.8 ...

A vector of results for the frame distorted by HRC #2

91.5 34.6 55.8 ...

A vector of results for the frame distorted by HRC #N
## Indicators

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Authors</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Commercial Black</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>2</td>
<td>Blockiness</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>3</td>
<td>Block Loss</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>4</td>
<td>Blur</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>5</td>
<td>Contrast</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>6</td>
<td>Exposure</td>
<td>VQ AGH</td>
<td>C/C++</td>
</tr>
<tr>
<td>7</td>
<td>Interlacing</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>8</td>
<td>Noise</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>9</td>
<td>Slicing</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>10</td>
<td>Spatial Activity</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>11</td>
<td>Temporal Activity</td>
<td></td>
<td>C/C++</td>
</tr>
<tr>
<td>No</td>
<td>Name</td>
<td>Authors</td>
<td>Language</td>
</tr>
<tr>
<td>----</td>
<td>-----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>12</td>
<td>BIQI</td>
<td></td>
<td>MATLAB</td>
</tr>
<tr>
<td>13</td>
<td>BRISQUE</td>
<td></td>
<td>MATLAB</td>
</tr>
<tr>
<td>14</td>
<td>NIQE</td>
<td>LIVE</td>
<td>MATLAB</td>
</tr>
<tr>
<td>15</td>
<td>OG-IQA</td>
<td></td>
<td>MATLAB</td>
</tr>
<tr>
<td>16</td>
<td>FFRIQUEE</td>
<td></td>
<td>MATLAB</td>
</tr>
<tr>
<td>17</td>
<td>IL-NIQE</td>
<td></td>
<td>MATLAB</td>
</tr>
<tr>
<td>18</td>
<td>CORNIA</td>
<td>UMIACS</td>
<td>MATLAB</td>
</tr>
<tr>
<td>19</td>
<td>HOSA</td>
<td>BUPT</td>
<td>MATLAB</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>F-measure</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>All metrics</td>
<td>0.779</td>
<td>0.776</td>
<td>0.777</td>
</tr>
<tr>
<td>Only ours</td>
<td>0.758</td>
<td>0.759</td>
<td>0.764</td>
</tr>
</tbody>
</table>

**Table:** General results we received for ALPR for 2 classes
Table: Confusion matrix for the test set, ALPR scenario, all metrics, and two classes

<table>
<thead>
<tr>
<th>Truth</th>
<th>Not more than 2 err.</th>
<th>Other cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not more than 2 err.</td>
<td>292</td>
<td>302</td>
</tr>
<tr>
<td>Other cases</td>
<td>138</td>
<td>628</td>
</tr>
</tbody>
</table>
Results

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

The numerical analysis is to check the sensitivity of the model to individual distortions.

As one can see, for the Gaussian Noise, Defocus, Motion Blur, and JPEG HRCs, the model shows quite similar error sensitivity – it is wrong in about 30% of the cases.

The exception is Exposure HRC, for which the model is much less mistaken, only for 11% cases.

**Figure:** Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in ALPR
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 ALPR
Results

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

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 ALPR
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 Gaussian Noise, Defocus, Motion Blur, and JPEG HRCs, the model shows quite similar error sensitivity – it is wrong in about 30% of the cases.

Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in ALPR
A more detailed analysis of the results obtained is also carried out.

The numerical analysis is to check the sensitivity of the model to individual distortions.

As one can see, for the Gaussian Noise, Defocus, Motion Blur, and JPEG HRCs, the model shows quite similar error sensitivity – it is wrong in about 30% of the cases.

The exception is Exposure HRC, for which the model is much less mistaken, only for 11% cases.

Figure: Share of erroneous predictions for a given Hypothetical Reference Circuits (HRC) in ALPR
We show in this study that the implementation of the new concept of an objective model to evaluate video quality for ALPR tasks is feasible.

The achieved value of the model accuracy (F-measure parameter) is 0.777.

When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.

Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.764).

It is worth mentioning the most typical problems encountered by the models during their work.

Our observations suggest that the characteristics of the initial scene are an important component that misleads the models.

VQI completely disregards this factor, which has a major impact on the accuracy of recognition.
We show in this study that the implementation of the new concept of an objective model to evaluate video quality for ALPR tasks is feasible.
We show in this study that the implementation of the new concept of an objective model to evaluate video quality for ALPR tasks is feasible.

The achieved value of the model accuracy (F-measure parameter) is 0.777.
Conclusions

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

2. The achieved value of the model accuracy (F-measure parameter) is 0.777.

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 ALPR tasks is feasible.

2. The achieved value of the model accuracy (F-measure parameter) is 0.777.

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 0.764).
We show in this study that the implementation of the new concept of an objective model to evaluate video quality for ALPR tasks is feasible.

The achieved value of the model accuracy (F-measure parameter) is 0.777.

When all potential VQIs are used (VQIs by AGH and other research teams), the best modelling results are obtained.

Nevertheless, it is worth noting that the restriction of AGH VQI does not lead to a significant decrease in prediction accuracy (F-measure of 0.764).

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 ALPR tasks is feasible.

2. The achieved value of the model accuracy (F-measure parameter) is 0.777.

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 0.764).

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 ALPR tasks is feasible.

2. The achieved value of the model accuracy (F-measure parameter) is 0.777.

3. When all potential VQIs are used (VQIs by AGH and other research teams), the best modeling 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 0.764).

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.
https://doi.org/10.1007/978-3-031-20215-5_13