Article

UHD Database Focus on Smart Cities and Smart Transport

Lukas Sevcik [*] , Miroslav Uhrina , Jaroslav Frnda

Article

UHD Database Focus on Smart Cities and Smart Transport

Lukas Sevcik * , **Miroslav Uhrina** and **Jaroslav Frnda**

University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia

*   Correspondence: lukas.sevcik@uniza.sk

Abstract: Smart City refers to a modern solution for organizing a city's services using cloud technologies to collect and evaluate large amounts of data, including data from camera systems. Smart city management covers several areas that can be implemented separately, but only their combination can realize the overall desired Smart City function. One of the core areas of smart city automation is smart city transport. Transportation is a crucial system in any city, and that's why it needs to be monitored. The primary objective of this publication is to generate top-notch 4K UHD video sequences that are solely dedicated to showcasing smart cities and their transportation systems. The resulting comprehensive database will be made accessible to all the professionals in the field who can utilize it for extensive research purposes. Additionally, all the reference video sequences will be transcoded into various quality settings by altering critical parameters like resolution, compression standard, and bit rate. The ultimate aim is to determine the best combination of video parameters and their respective settings based on the measured values. This in-depth evaluation will ensure that each video sequence is of the highest quality and provides an unparalleled experience for the service providers offering the service. The video sequences captured will be analyzed for quality assessments in smart cities or smart transport technologies. The database will also include objective and subjective ratings, along with information about the dynamics determined by spatial and temporal information. This will enable a comparison of the subjective evaluation of a selected sample of our respondents with the work of other researchers who may evaluate with a different sample of evaluators. The assumption of our future research is to predict the subjective quality based on the type of sequence determined by its dynamicity. Our database's reference sequences can be found at https://doi.org/10.5281/zenodo.10663664 [1], while the FTP server with IP address 158.193.214.161 hosts also encoded sequences and evaluations. Detailed information is described in Section 7.

Keywords: UHD database; smart cities; smart transport; subjective evaluation; VMAF; QoE

---

1. Introduction

It is crucial to have a reliable evaluation of the quality of video services offered online, including those of video-on-demand. This is because such services are rapidly expanding and gaining widespread popularity, as seen with the likes of YouTube, Netflix, and other streaming platforms. The technology used for streaming is also evolving, from the traditional non-adaptive form to adaptive streaming. Additionally, internet video streaming technology is moving from an adapted connection-oriented video transport protocol, such as the Real Time Messaging Protocol (RTMP), to adaptive streaming that utilizes HTTP.

The use of surveillance cameras has increased significantly, leading to challenges in storing, transmitting, and analyzing video data. As a result, there is a great need for a reliable system to manage large amounts of video data. Such a system should have efficient video compression, stable storage, and high-bandwidth Ethernet or Internet transmission capabilities. In modern cities, distributed cameras capture video from various scenarios. Video sensor systems can provide valuable data for improving traffic planning and management. Future intelligent road technologies will rely heavily on the quality and quantity of such data. Video-based detection systems are an essential

component of intelligent transport systems and offer a flexible way of acquiring data. They are also evolving rapidly, which makes them highly promising.

We must keep in mind that quality is a crucial factor in many industries. In the video industry, various databases are designed to assess video quality using objective or subjective metrics. We have compared various databases produced by different institutions. Ultra Video Group [2] shot 16 4K quality video sequences, capturing different spatio-temporal information and covering all four quadrants. They used well-known sequences, such as Beauty, Bosphorus, Jockey, or ReadySetGo, as a good basis for testing. The SJTU database [3] contains ten 4K video sequences, some of which could be used in the smart city monitoring industry, such as Bund Nightscape, Marathon, Runners, Traffic and Building, and Traffic Flow. However, this database is not designed specifically for this purpose. In [4], a database of video sequences suitable for mobile viewing was described, exploring the impact of adaptive streaming for mobile phones, where the quality can vary depending on the connection. In [5], a database representing different distortions was created and presented, taking into account not only temporal and spatial information, but also colorfulness, blurriness, or contrast.

The authors proposed a database called LIVE-NFLX-II in the paper [6]. This database contains subjective QoE responses for various design dimensions, such as different bit rate adaptation algorithms, network conditions, and video content. The content characteristics cover a wide range, including natural and animated video content, fast and slow scenes, light and dark scenes, and low and high texture scenes. In [7], the authors proposed and built a new database, the LIVE-Qualcomm mobile in-capture video quality database, which contains a total of 208 videos that model six common in-capture distortions. The coding structure, syntax, various tools, and settings relevant to coding efficiency have been described in [8]. In it, the perception of compression as well as spatial and temporal information was further investigated. The authors compiled an extensive database of video sequences whose quality was subjectively evaluated.

To enhance the accuracy of video quality prediction of NR, a comprehensive video quality assessment database was developed in [9]. This database comprises 585 videos featuring unique content, captured by a diverse group of users with varying levels of complexity and authentic distortions. The subjective video quality ratings were determined through crowdsourcing. To effectively analyze and utilize data in areas such as smart cities and smart traffic, it is crucial to expand existing databases. This means adding new sequences that solely contain snapshots of the city, traffic, or other relevant information.

2. Related work

Quality of Experience (QoE) is highly dependent on QoS parameters, as factors such as latency, jitter or packet loss are also important in video traffic. Although such factors are easily measurable, QoE cannot be easily quantified. Currently, one of the most popular network services is live video streaming, which is growing at a behemoth scale [10,11]. In [12], a detailed quantitative analysis of video quality degradation in a homogeneous HEVC video transcoder was presented, along with an analysis of the origin of these degradations and the impact of the quantization step on the transcoding. The differences between video transcoding and direct compression were also described. The authors also found a dependency between the quality degradation caused by transcoding and the bit rate changes of the transcoded bitrate.

In [13] the authors compared the available 4K video sequences. The compression standards H.264 (AVC), H.265 (HEVC), and VP9 were compared. Video sequences were examined using objective metrics PSNR, SSIM, MS-SSIM, and VMAF. Recently, many experts and researchers have provided quality performance analyses of well-known video codecs such as H.264/AVC, H.265/HEVC, H.266/VVC, AV1. Authors in [14] performed analysis between the HEVC and VVC codecs for test sequences with resolution ranging from 480p up to Ultra HD (UHD) resolution using the Peak Signal-to-Noise Ratio (PSNR) objective metric. In paper [15], the rate-distortion analysis of the same codecs using the PSNR, Structural Similarity Index (SSIM), and Video Multi-Method Assessment

Fusion (VMAF) quality metrics was provided. Authors in [16,17] assessed video quality of the HEVC, VVC, and AV1 compression standards for test sequences with resolution varying from 240p to UHD/4K resolutions and in [18,19] at Full HD (FHD) and Ultra HD (UHD) resolutions, respectively. The compression efficiency was calculated using the PSNR objective metric. In [18,19], for the quality evaluation also the Multi-Scale Structural Similarity Index (MS-SSIM) method was used. Paper [20] presents a comparative performance assessment of five video codecs - HEVC, VVC, AV1, EVC, and VP9. The experimental evaluation was done on three video datasets with three different resolutions - 768x432, 560x488, and 3840x2160 (UHD). Paper [21] deals with an objective performance evaluation of the HEVC, JEM, AV1, and VP9 codecs done by the PSNR metric. A large test set of 28 video sequences with different resolutions varying from 240p to Ultra HD (UHD) was generated. Paper [22] examines the compression performance of three codecs, namely HEVC, VVC, and AV1 measured by the PSNR and SSIM objective video quality metrics. In paper [23] authors compared the coding performance of the HEVC, EVC, VVC, and AV1 in terms of computational complexity.

In [24], the authors proposed a new methodology for video quality assessment using JND (Just-noticeable-difference). The publication focuses on describing the process of subjective tests. In [25], the authors presented an empirical study of the impact of packet loss-related errors on television viewing. They compared different delivery platforms and technologies. Video sequences and delivery quality information obtained from the service provider were used in the experiments. Sequence length, content and connection type were compared. In [26], 16 types of metrics were compared for quality assessment. Packet loss was simulated in the video encoding and the losses were then hidden using different techniques to conceal the errors. The purpose was to show that the subjective quality of a video cannot be predicted from the visual quality of the frame alone when some hidden error occurs. In [27], a new objective indicator XLR (piXel Loss Rate) was proposed. It evaluates the packet loss rate during video streaming. This method achieved comparable results with fully benchmarked metrics and a very high correlation with MOS. In [28], the authors provided an overview of packet loss in Wi-Fi networks, mainly for real-time multimedia.

In [29], an optimal packet classification method was proposed to classify packets that were given a different priority when the transmission conditions deteriorated. The network transmits the segments with the highest priority concerning perception quality when constrained network conditions occur. The results showed that the proposed method can achieve higher MOS scores compared to non-selective packet discarding. The authors in [30] stated that highly textured video content is difficult to compress because a trade-off between bit rate and perceived video quality is necessary. Based on this, they introduced a video texture dataset that was generated using a development environment. It was named the BVI-SynTex video dataset and was created from 196 video sequences grouped in three different texture types. It contains five-second Full HD video scenes with a frame rate of 60 fps and a depth of 8 bits.

Video analysis in smart cities is very useful when deploying various systems and investigating their performance. The description of technologies used for video analysis in smart cities can be found in publication [31]. With the help of this analysis, various situations have been studied, whether it is traffic control and monitoring, security, or entertainment. In [32], the authors evaluate the transmission of multimedia data at a certain throughput setting. They then evaluate the performance and describe the benefits of real-time transfer. This publication describes video surveillance in smart cities, and multimedia data transmission with cloud utilization. They discuss the impact of network connectivity on the transmission of multimedia data over a network. The algorithm dealing with image and video processing was presented by the authors in publication [33]. This solution suppresses noise to achieve accuracy in the traffic scene. This knowledge was then used in a smart lighting experiment in a smart city system.

The transmission of streaming video in automotive ad-hoc networks is addressed by the authors in [34]. These scenarios investigate the conditions that affect the quality of streaming, which is simulated in the NS3 environment. The publication [35] discusses the influence of resolution, number

of subjects per frame, and frame rate on the performance of metrics for object recognition in video sequences. The authors used videos taken from cameras placed at intersections, which captured the scene from above. Changing the cropping method or changing the frame width was described. The classification of municipal waste using an automated system was proposed in paper [36]. The suggested model classifies the waste into multiple categories using convolutional neural networks.

The authors proposed a blind video quality assessment (BVQA) method based on DNN to compare scenarios in the wild in [37]. Transfer learning methods with spatial and temporal information were used. They used DNN to account for the motion perception of video sequences for spatial features. They applied the results to six different VQA databases. In this work, the authors have used their knowledge from image quality assessment (IQA). The authors of the research paper [5] have developed a database of videos. This database is sampled and subjectively annotated and is intended to display authentic distortions. To ensure that the dataset is diverse in terms of content and multidimensional quality, six attributes were computed. These attributes include spatial information, temporal information, blur, contrast, color, and VNIQE. The paper introduces a new VQA database called KoNViD-lk.

In paper [38], the authors propose an EQASH (Enhanced Quality Adaptation Scheme for DASH). The proposed scheme adjusts the quality of the segments not only based on the network and playback buffer status but also based on the VBR characteristics of the contents. The proposed scheme also reduces the latency by employing the new server push feature in HTTP 2.0. According to a study [39], a video playback schedule that has a minimum number of low-resolution video segments provides the best QoE. The paper presents the M-Low linear time scheduling algorithm, which adjusts the video resolution and optimizes the QoE indices in the DASH streaming service. The authors of the study describe several QoE metrics, including minimization of resolution switching events, freeze-free playback, maximization of video playback bit rate, and minimization of low-resolution video segments.

The introduction of smart cities has revolutionized the way we live, work, and commute. Although the papers presented in the introduction [2–9] have created various databases, none of them have covered the content of smart cities directly. After conducting extensive research, it has become evident that this topic is highly sought after and the creation of a database that focuses on smart transportation and smart cities will be highly beneficial. The primary objective of this paper is to create a database that is unprecedented in its type. The database will contain images that capture smart transportation and smart cities. Additionally, the reference sequences will be transcoded into different quality settings. The final transcoded images will be evaluated subjectively and objectively to ensure that they meet the desired quality standards. This will enable researchers, developers and stakeholders to have access to high-quality images that can be used in various applications related to smart cities.

3. Motivation

The concept of a smart city has emerged in the last decade, as a combination of ideas on how information and communication technologies can enhance the functioning of cities. The goal is to improve efficiency, and competitiveness, and provide new ways to address poverty and social deprivation. The main idea is to coordinate and integrate technologies that can bring new opportunities to improve the quality of life. A smart city can take various forms, such as a virtual, digital, or information city. These perspectives emphasize the role of information and communication technologies in the future operation of cities.

The concept of Quality of Experience was introduced as an alternative to Quality of Service (QoS) to design more satisfactory systems and services by considering human perception and experience. Our research focuses on QoS as well as QoE and their interconnection using a mapping function, followed by prediction. We test the impact of various quality parameters such as resolution, bitrate, and compression standard on the resulting quality. In the case of QoE, we use subjective metrics for evaluation, while for QoS, objective metrics are used. We also simulate the impact of packet loss, delay, or network congestion using a simulation tool to understand their effects on quality.

Based on the results and evaluations obtained, we recommend an appropriate choice of parameters that will guarantee maximum quality for the end user while ensuring bandwidth efficiency for the provider. By combining these parameters, we can set the Variable BitRate (VBR) to stream the video as efficiently as possible. In a classical streaming scenario, the video is viewed at one specific resolution, which is predefined before each session is started using a connection-oriented transport layer protocol. Adaptive streaming, on the other hand, involves encoding the video at multiple discrete bit rates. Each bitstream or video with a specific resolution is then divided into sub-segments or chunks, each taking a few seconds to complete (typically around 2-15 seconds). For optimal video quality during playback, it's important to ensure that the end user's connection conditions and download speed are taken into consideration. VBR encoding can lead to inconsistencies in video block size, which can cause frequent re-caching and reduce the user's QoE, especially when network bandwidth is limited and fluctuating.

In this publication, we will discuss the impact of various quality settings, such as codec used, resolution, and bit rate, on overall quality. Both objective and subjective metrics will be used to determine the quality. Quality and appropriately set parameters are also important in the field of smart cities and traffic. The rapid expansion of cities in recent years has resulted in urban problems such as traffic congestion, public safety concerns, and crime monitoring. Smart city technologies leverage data sensing and big data analytics to gather information on human activities from entire cities. This data is analyzed to provide intelligent services for public applications.

4. Methodology

Video quality analysis focuses on packet loss in the network depending on the codec used, which causes artifacts in the video. We use QoE metrics to determine user satisfaction boundaries and, most importantly, the application of such QoS tools in the network to guarantee the minimum QoE expected by the user. The use of the Internet as an environment for multimedia delivery is quite common today, but it is not entirely obvious that the user will receive, in such an environment, a service with the desired quality. This makes QoE monitoring and the search for links between QoS and QoE all the more important today.

It is essential to evaluate the performance of systems for sending information from one source to another (data link) and ensure efficient information transfer. When evaluating the transmission quality of IPTV services, we focus on user satisfaction with the quality of media content. It is generally assumed that high performance and transmission quality result in high user satisfaction with the service. From a human perceptual point of view, quality is determined by the perceived composition, which involves a process of perception and judgment. During this process, the perceiver compares the perceived events with a previously unknown reference. The nature of the perceived composition may not necessarily be a stable characteristic of the object, as the reference may influence what is currently perceived. Quality is usually relative and occurs as an event in a particular spatial, temporal, and functional context.

Objective quality assessment is an automated process, as opposed to subjective assessment, which requires human involvement. There are three types of methods for objective video quality models, which can be classified based on the availability of information about the received signal, the original signal, or whether the signal is present at all (FF). In our evaluation, we use FF objective methods (SSIM, MS-SSIM, PSNR, and VMAF). A more detailed description can be found in our previous publications [40,41]. As a subjective metric, we used the non-referential ACR method, because in this case the video is compared only based on the seen video sequence and not by comparison with a reference. In a real environment, when receiving a signal from a service provider, the end user also receives only the received signal and does not compare it with the reference original. The quality is defined by a 5-degree MOS scale. This standard [42] provides a methodology for the subjective assessment of the quality of voice and video services from the end-user's perspective. This metric summarises ratings that are averaged on a scale from 1, which is the worst quality, to 5, which represents excellent quality. For more information, see our publication [41].

5. The methods of proposed model

Our primary goal will be to create video sequences in Ultra HD 4K resolution, which will contain various shots that map the traffic and the city. The created database of video sequences will cover both static and dynamic sequences. The created video sequences will then be transcoded to the necessary quality settings and objectively and subjectively rated. Furthermore, they will be accessible for subjective evaluation by another group. Each sequence will be identifiable by different parameters, e.g. SI (spatial information) and TI (temporal information). Subsequently, using a neural network, an appropriate bitrate can be allocated to each video sequence to achieve the desired quality.

To begin with, we had to take numerous shots from which we selected reference video sequences. These were chosen to cover as much space as possible in the SI and TI quadrants. Their description can be found in Section 6.1. The next step was to encode these reference sequences into a combination of Full HD and Ultra HD resolutions, using H.264 (AVC) and H.265 (HEVC) compression standards and bitrates of 5, 10, and 15 Mbps using FFmpeg. FFmpeg is a collection of programs and libraries that enable the processing, editing, or playing of multimedia files. It is operated via the command line. In our case, the multimedia content had to be first encoded into a defined codec, which is a compression algorithm. Then, it was decoded to enable its use. With transcoding, it is possible to convert multimedia files to a different file container, or codec, or to use different framerates.

The selection and evaluation process are illustrated in Figure 1. The encoding process can be found in Section 6.2. After transcoding, the sequences are characterized again using SI and TI information. The sequences are evaluated using objective metrics such as SSIM, MS-SSIM, PSNR, and VMAF (see Section 6.3 for a description). The subjective metric ACR evaluation is described in Section 6.4.
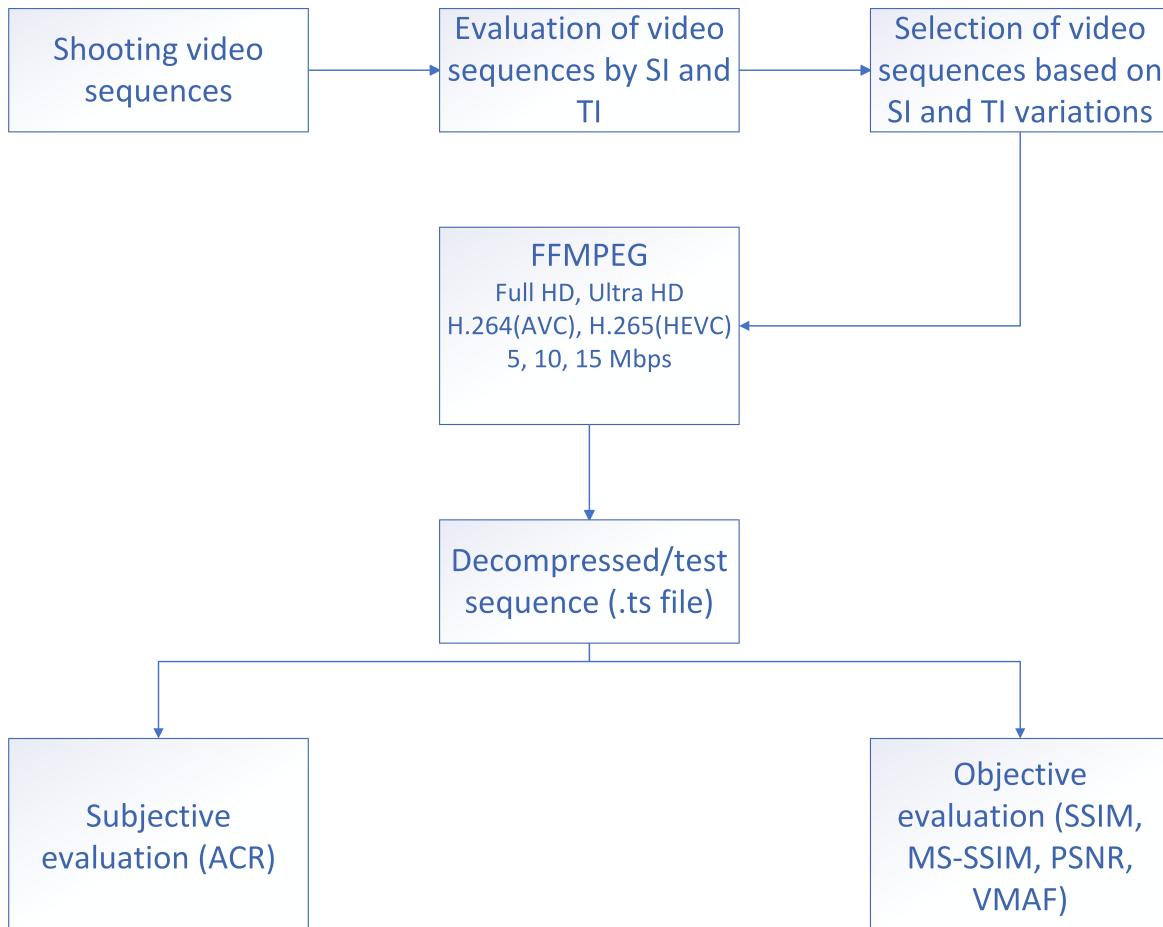

Figure 1. The selection and evaluation scheme of the video sequences.

6. Results

In this chapter, we describe how the database was created, the encoding of the resulting video sequences, their characteristics, and then the objective and subjective evaluation of each sequence.

6.1. Description of the dataset

The video sequences were captured using a DJI Mavic Air 2 drone. Unfortunately, this device doesn't allow shooting video sequences in uncompressed .yuv format. However, UHD (3840×2160) format is available. The parameters that were chosen for shooting can be found in Table 1. 4K with a resolution of 4096×2160 is not yet used commercially and therefore, UHD is preferred due to its 16:9 image ratio. This is why UHD resolution was used for our recording. The aim of this paper is to create a database of 4K video sequences that cover scenes from traffic monitoring and cityscapes. These sequences will be encoded to the necessary quality parameters and rated either objectively or subjectively. The video sequences we created offer a wide variety of dynamicity, whether it's the dynamics of the objects in the video or the dynamics of the camera.

The following video sequences have been created focusing on transport:

- Dynamic road traffic-dynamic camera motion - frequent traffic at higher vehicle speeds (name in our database: Sc1)
- Dynamic road traffic-static camera motion (Sc2)
- Parking lot-dynamic camera motion - less dynamic movement of cars in a parking lot (Sc8)
- Parking lot-static camera motion (Sc3)
- Road traffic-busy traffic at lower vehicle speeds (Sc4)
- Traffic roundabout-dynamic camera motion - traffic on a roundabout (Sc5)
- Traffic roundabout with a parking lot - a dynamic part of the scene with slow movement in the parking lot (Sc6)
- Traffic roundabout-static camera motion-traffic on a roundabout (Sc10)
- Train station-train leaving the station (Sc7)
- Dynamic train-train in dynamic motion (Sc9)
- Trolleybus-trolleybus arriving at a public transport stop
- Dynamic trolleybus-trolleybus in dynamic driving
- The University town-university town (movement of people)
- Waving flags-flags flying in the university town

You can view a preview of the reference sequences in Figure 2.

Each video sequence was evaluated based on its SI and TI values. The resulting parameter value is the maximum value across all frames. The temporal information is derived from changes in brightness of the same pixels in successive frames, while the spatial information is obtained from the Sobel filter for the luminance component and the subsequent calculation of standard deviation of the pixels. More details can be found in our publication [41]. Table 2 shows the characterization of the reference sequences based on SI and TI values, which highlights the diversity of the individual sequences in terms of spatial and temporal information.

Table 1. Recording parameters.

Resolution	Ultra HD (3840×2160)
Compression standard	H.265/HEVC
Bitrate	120 Mbps
Video Frame Rates	50 fps (Frames per Second)
Subsampling	4:2:0
Bit depth	8b
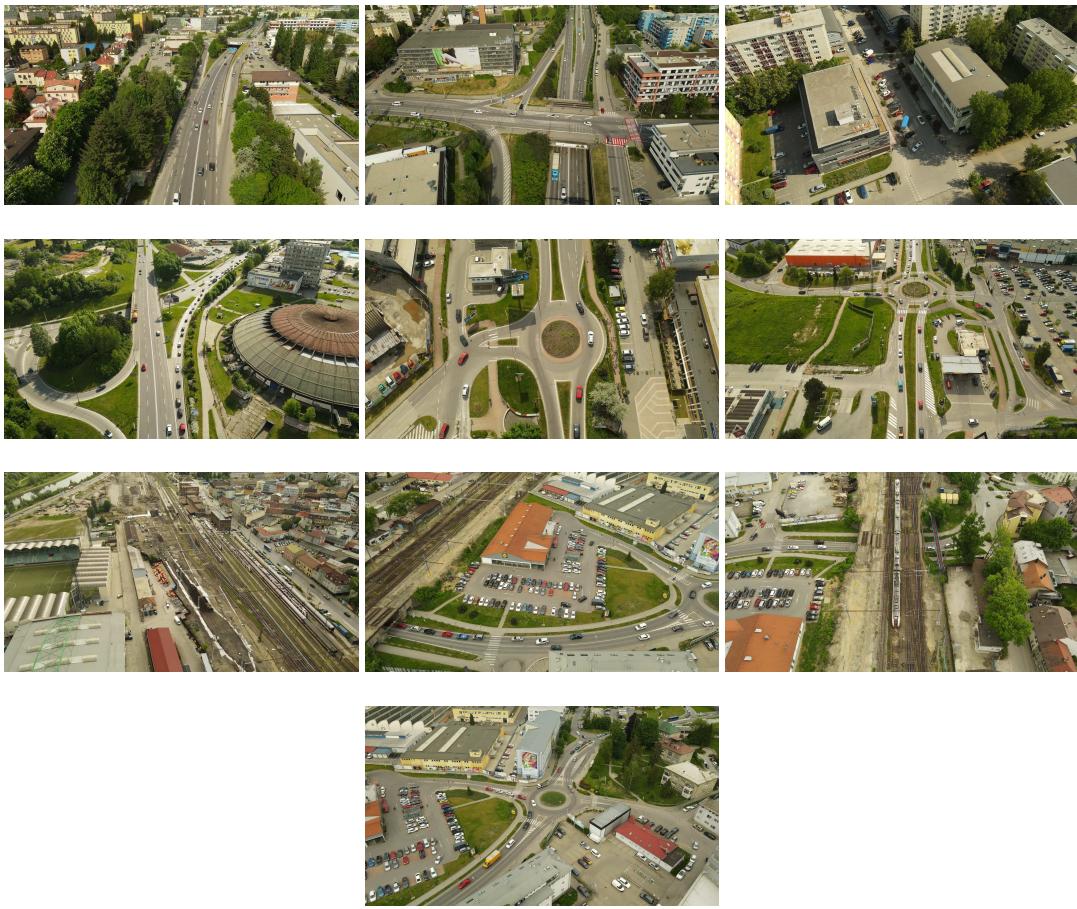
Figure 2. Previews of created sequences.

Table 2. SI and TI information of the reference videosequences.

	max SI	max TI
Sc1	90.8834	35.9334
Sc2	83.8965	6.29607
Sc3	85.1865	5.19369
Sc4	88.624	19.3205
Sc5	72.303	22.8181
Sc6	74.9079	11.1935
Sc7	96.9839	16.6872
Sc8	78.7106	5.47847
Sc9	87.0767	24.1257
Sc10	76.3554	5.06084

6.2. Encoding of the reference videosequences

The first ten reference sequences from the list above were further encoded to Full HD (1920×1080) resolution and H.264/AVC compression standard. These are labeled "Sc_x" so that we can name them for each variation. These reference sequences were selected precisely based on the characterization of temporal and spatial information. The quality of the encoded content is determined by the amount of data lost during the compression and decompression of the content. In real broadcasting, a bitrate of 10 Mbps is often used for HD resolution, and some stations use bitrates up to around 15 Mbps. This bitrate is also taken into account for UHD deployments. Therefore, each sequence in both resolutions and compression standards has been encoded with bitrate values of 5, 10, and 15 Mbps.

Table 3 shows the parameters that we used in encoding the video sequences. This combination produced twelve variations for each one which means 120 sequences. We evaluated each with objective metrics. Seven of them were also evaluated subjectively (Sc1-5, Sc9-10). The seven sequences for subjective evaluation were selected to comply with the recommendations [43]. With the combination of seven sequence types with twelve coding variations, we can evaluate one group continuously without a long pause. If we selected more sequence types, we would have to split the subjective evaluation of one group of evaluators, which would imply a higher time effort. When selecting these seven video sequences, we also considered the calculation of spatial and temporal information of the sequences.

Table 3. Parameters of encoded sequences.

Resolution	Full HD, Ultra HD
Compression standard	H.264/AVC, H.265/HEVC
Bitrate [Mbps]	5, 10, 15
Frames per Second	50 fps
Subsampling	4:2:0
Bit depth	8b

The created database will be available to the general scientific public. The created video sequences can be further used for the needs of the analysis of the appropriate qualitative setting in order to provide the highest possible quality while saving as many network resources as possible. Thus, it will be possible to further work with the database, to shoot new sequences, which will then be evaluated, either by objective or subjective tests. This will give a detailed view of the performance of streaming video over IP-based networks. Video sequences offer the possibility to test what other parameters can characterize a given sequence, or how individual video parameters affect the quality. It will also be possible to see at which bitrate each scene achieves the highest end-user satisfaction in terms of quality and thus define boundaries for each scene based on selected content information. A suitable bitrate would be assigned for each boundary so that it satisfies the highest quality at the individual resolution. This will allow technologies and applications in the smart cities and smart traffic sector to use the available resources efficiently.

We encoded the created reference sequences with changing quality parameters using FFmpeg. Coding example for 15 Mbps:

ffmpeg -i input_sequence -vf scale=resolution -c:v codec -b:v 15000k -maxrate 15000k -bufsize 15000k -an -pix_fmt yuv420p -framerate 50 SeqName.ts

Description of the individual parameters used in the command:

-i is used to import video from the selected file.
-vf scale is used to specify the resolution of the video. In our case, this parameter was changed for Full HD resolution (1920x1080) and UHD resolution (3840x2160).
-c:v is used to change the video codec, we used two codecs, H.264/AVC, which is written libx264, and the H.265/HEVC codec, which is written libx265.
-b:v is used to select the bitrate, we varied this parameter to 5, 10, and 15Mbps.
-maxrate is used to set the maximum bit rate tolerance. It requires buffsize in the settings.
-bufsize is used to choose the buffer.
-an is a parameter that removes the audio track from the video.
pix_fmt is the parameter to select the subsampling.
-framerate is used to set the number of frames per second.

The last parameter is the video output where we set the video name and its format.

The output sequence has been encoded into a .ts container so that we can test the impact of packet loss in the future. You can use programs like Media Info and Bitrate Viewer to check the individual transcoded parameters. Media Info will display all the parameters and settings of the video, while Bitrate Viewer is used to display the bitrate in exact time.

We have included a five-second pause between each sequence to ensure that the evaluators don't overthink the evaluation and it remains spontaneous. The video contains a grey background, so the image is not distorted and does not draw the eyes of the evaluators. We have inserted text in the grey background that describes the rating so that the raters know in which part of the evaluation process to do it.

To ensure an accurate evaluation, a maximum of three people participated simultaneously, and they had a direct and undistorted view of the TV set. The video sequences were evaluated on a Toshiba 55QA4163DG TV set placed 1.1 meters from the raters, in compliance with the standard [43]. The distance between the viewer and the monitor should be 1.5 times the height of the monitor. Each evaluator had access to a questionnaire where they recorded early evaluations of a given sequence. A total of 30 human raters participated in the evaluation, rating 84 video sequences. The MOS rating scale of 1 to 5 was used for the evaluation, where 1 represents the worst quality, while 5 is the best.

In the following sections, we will analyze the outcomes of both objective (see Section 6.3) and subjective (see Section 6.4) evaluations. Please note that the results presented here are based on selected samples only, while all other numerical or graphical data can be obtained upon request. Moreover, we are currently working on creating a website where the entire database will be published and available for free.

6.3. Objective quality evaluation

In the case of objective evaluation, we selected one video sequence to present the results, namely the Traffic roundabout with a parking lot (Sc6). For this sequence, we present the evaluation progress frame by frame of the individual objective metrics (SSIM, MS-SSIM, PSNR, and VMAF) for 15 Mbps bitrate in both resolutions and codecs. Results are presented by normalized value range <0, 1>. Here it is possible to compare the overall correlation of the evaluated metrics.

The results of Ultra HD resolution for the H.265 (HEVC) codec can be seen in Figure 3, for the H.264 (AVC) codec in Figure 4. Full HD resolution can be viewed in Figure 5 for the H.265 (HEVC) compression standard and in Figure 6 for the H.264 (AVC) compression standard. With such a high bitrate, the H.265 compression standard achieves better results compared to H.264 for both resolutions.
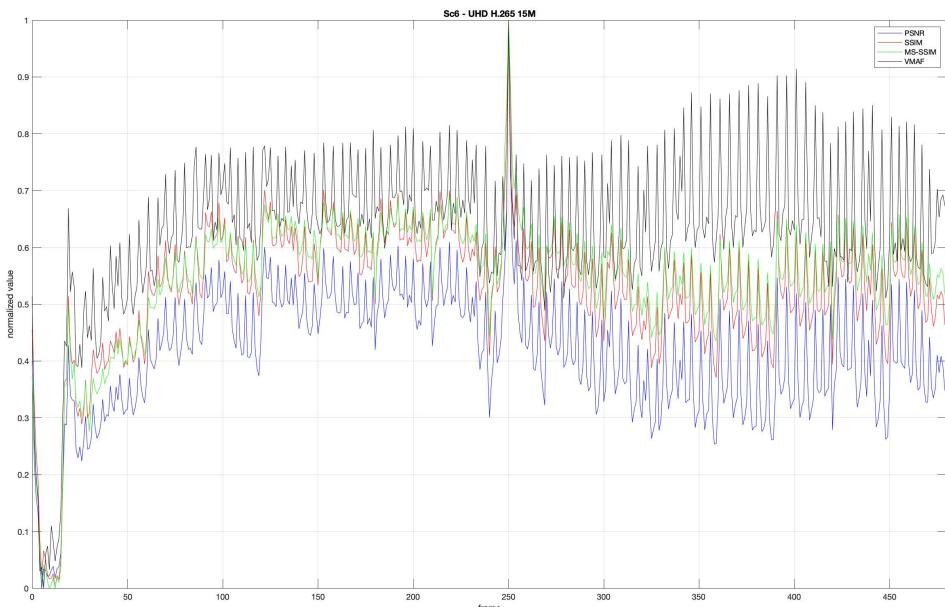
Figure 3. Traffic roundabout with a parking lot (Sc6): Ultra HD, H.265, 15Mbps norm.

Figure 4. Traffic roundabout with a parking lot (Sc6): UHD, H.264, 15Mbps norm.
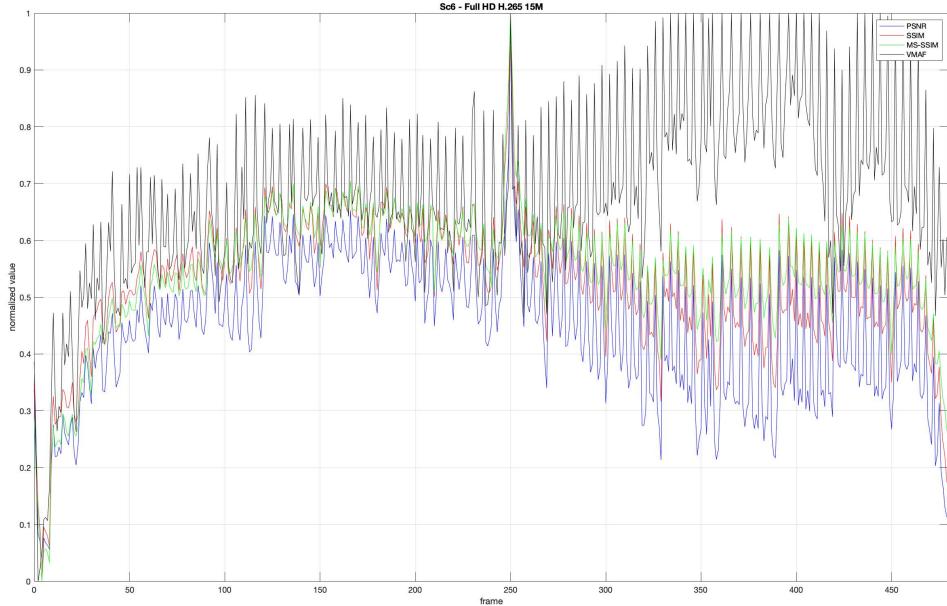
**Figure 5.** Traffic roundabout with a parking lot (Sc6): Full HD, H.265, 15Mbps norm.

Full HD resolution achieves a better rating. With increasing bitrate the difference is smaller. When comparing Figures 3 and 4, we can conclude that the ratings correlate with each other and there are noticeable equal rating shifts in both compression standards. At Full HD resolution we can observe a larger variation between the compression standard H.265 in Figure 5 and for H.264 in Figure 6. Mapping the results of the different objective metrics confirms the high correlation between the methods used and brings the comparison of these metrics closer to the citers. We can also see that the VMAF scores oscillate more than the results of other metrics.
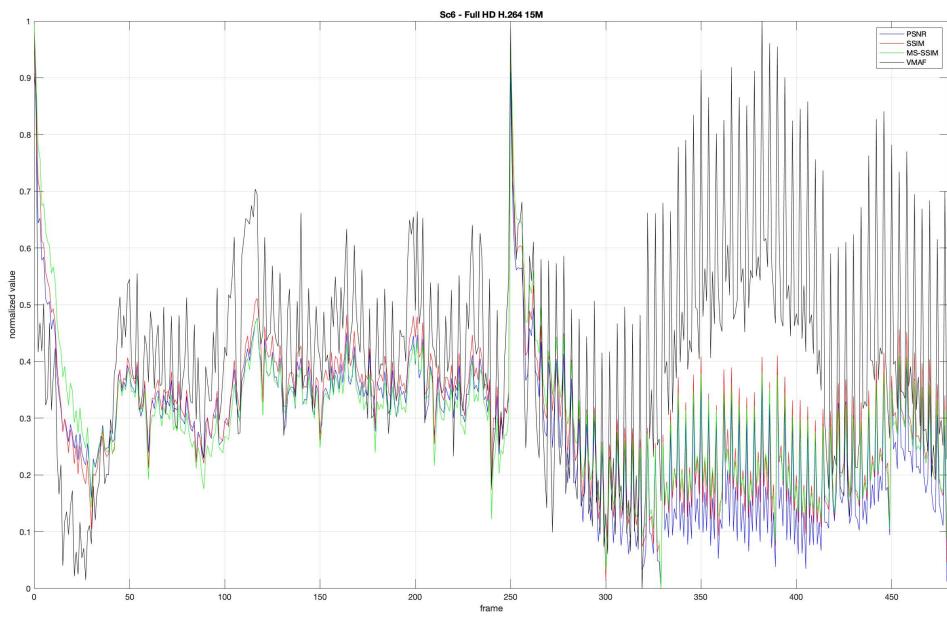
Figure 6. Traffic roundabout with a parking lot (Sc6): Full HD, H.264, 15Mbps norm.

We are presenting the final results for each sequence (Sc1-Sc10) in the form of mean values of VMAF and PSNR metrics for the 15 Mbps bitrates. As expected, the H.265 codec achieves better results, and we can also see an improvement in the results with an increasing bitrate value. The results of the objective evaluation of all sequences for Ultra HD resolution in combination with the H.265 (HEVC) codec are shown in Figure 7, and for the H.264 (AVC) codec in Figure 8. We can observe that the results for the other sequences also confirm that the H.265 (HEVC) compression standard has a better rating. For some sequences, the difference is more pronounced, which is due to the dynamics of the scene.
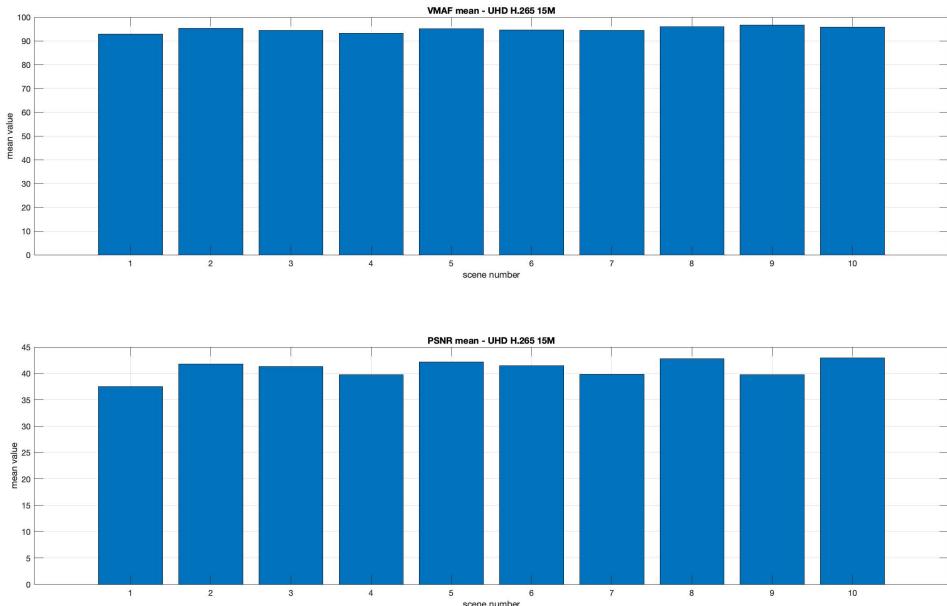

**Figure 7.** Mean values of VMAF and PSNR for UHD, H.265, 15 Mbps.
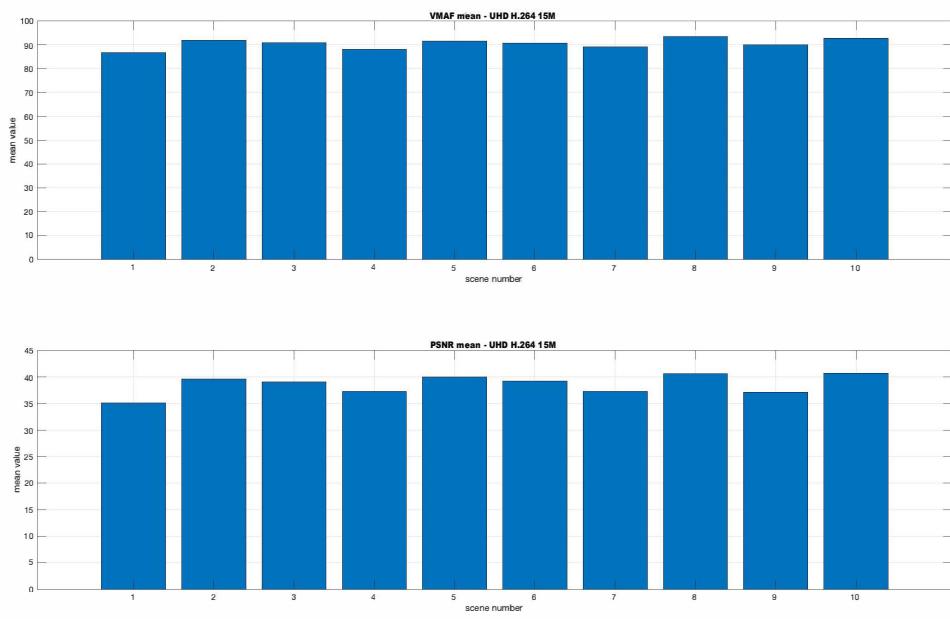
Figure 8. Mean values of VMAF and PSNR for UHD, H.264, 15 Mbps.

At Full HD resolution, the differences between H.265 (HEVC), which can be seen in Figure 9, and H.264 (AVC), shown in Figure 10, are smaller. In both cases, Full HD resolution achieves higher values than Ultra HD resolution.
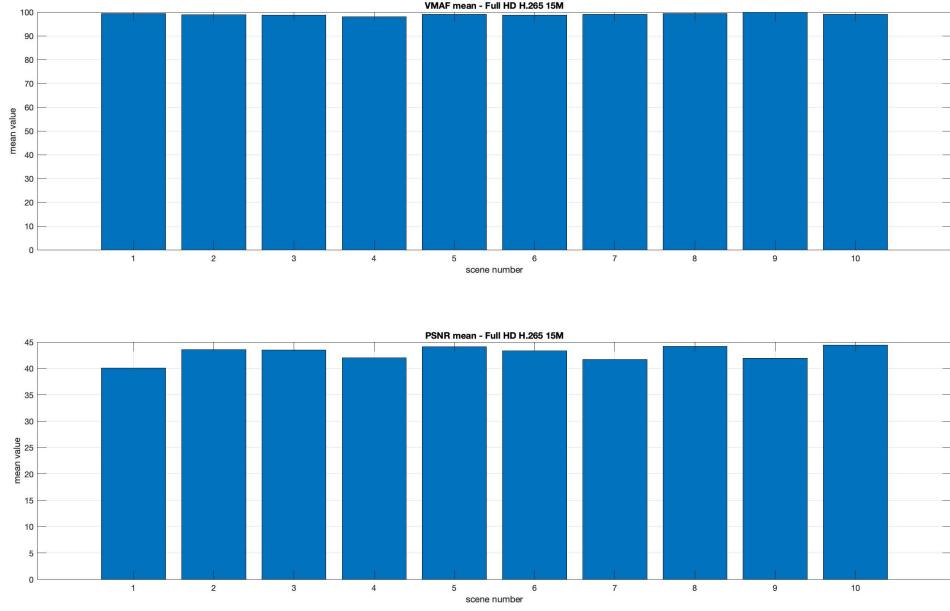

Figure 9. Mean values of VMAF and PSNR for Full HD, H.265, 15 Mbps.
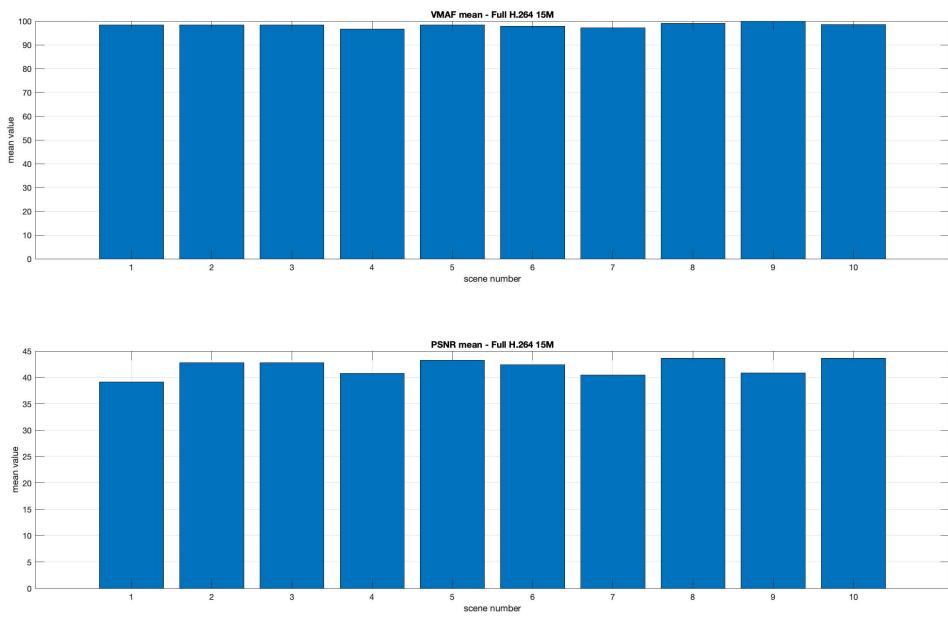
Figure 10. Mean values of VMAF and PSNR for Full HD, H.264, 15 Mbps.

6.4. Subjective quality evaluation

In this chapter, we present the results of the subjective evaluation of selected seven reference sequences. These sequences were recoded into various qualitative parameters. We calculated the average ratings from 30 users for each type of coded sequence from the references Sc1 (Dynamic road traffic-dynamic camera motion) and Sc2 (Dynamic road traffic-static camera motion), as shown in Figure 11a. The results for Sc3 (Parking lot-static camera motion) and Sc4 (Road traffic) can be seen in Figure 11b, while the figures for Sc5 (Traffic roundabout-dynamic camera motion) and SC10 (Traffic roundabout-static camera motion) are presented in Figure 11c.
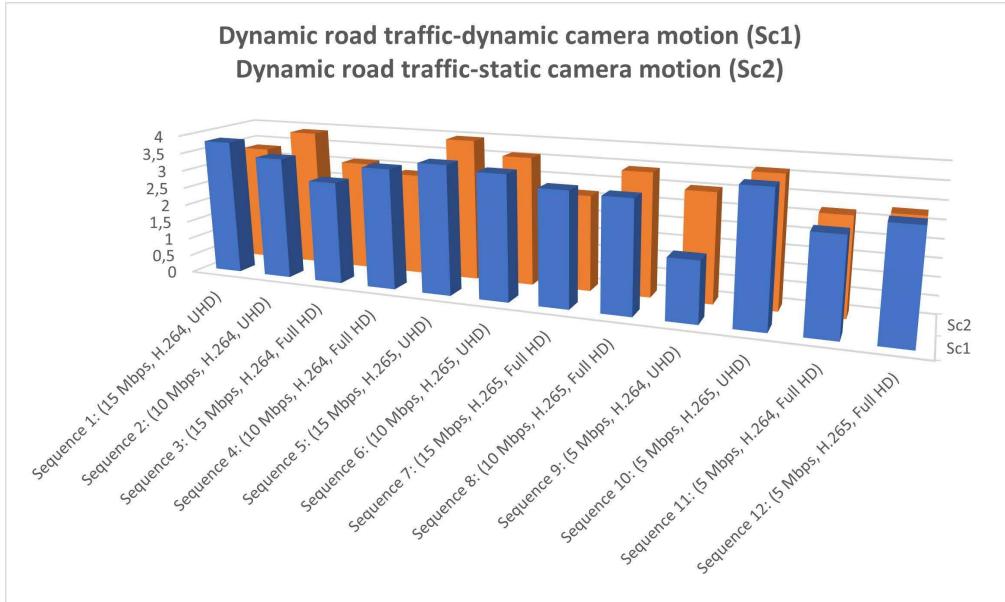
(a) Subjective evaluation of the Sc1 and Sc2.

Figure 11. *Cont.*
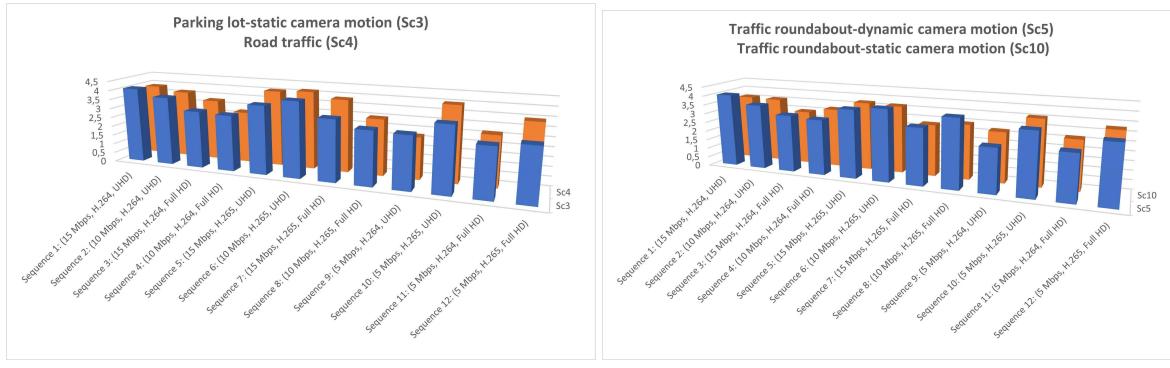
**(b) Subjective evaluation of the Sc3 and Sc4.**   **(c) Subjective evaluation of the Sc5 and Sc10.**

Figure 11. Average values of subjective evaluation.

In Table 4, you can find the complete results of the transcoded sequences from the Dynamic Train (Sc9) reference sequence. Table 4 includes the average result as well as the exact number of occurrences for each MOS scale value.

Table 4. Dynamic train - train in dynamic motion (Sc9).

| | MOS score | | | | | Average value |
	1	2	3	4	5	
Sequence 1: (15 Mbps, H.264, UHD)	0 times	6 times	11 times	9 times	4 times	3,37
Sequence 2: (10 Mbps, H.264, UHD)	2	7	14	7	0	2,87
Sequence 3: (15 Mbps, H.264, Full HD)	10	8	7	4	1	2,27
Sequence 4: (10 Mbps, H.264, Full HD)	1	5	15	8	1	3,1
Sequence 5: (15 Mbps, H.265, UHD)	0	3	7	11	9	3,87
Sequence 6: (10 Mbps, H.265, UHD)	2	6	10	9	3	3,17
Sequence 7: (15 Mbps, H.265, Full HD)	4	6	11	8	1	2,87
Sequence 8: (10 Mbps, H.265, Full HD)	2	6	10	9	3	3,17
Sequence 9: (5 Mbps, H.264, UHD)	20	8	1	1	0	1,43
Sequence 10: (5 Mbps, H.265, UHD)	3	7	12	6	2	2,9
Sequence 11: (5 Mbps, H.264, Full HD)	7	7	9	7	0	2,53
Sequence 12: (5 Mbps, H.265, Full HD)	1	11	8	8	2	2,97

6.5. Correlation between Objective and Subjective assesment

There are various metrics to express the correlation between subjective and objective assessment. The two most commonly used statistical metrics to measure performance are Root Mean Square Error (RMSE) and Pearson's correlation coefficient. A high correlation value (usually greater than 0.8) is considered to be effective. To measure correlation, we used three sequences (Sc1-Dynamic road traffic, Sc9-Dynamic train, and Sc10-Traffic roundabout) in UHD resolution for comparison. The results show that there is a strong correlation between the subjective evaluation by the respondents and the objective evaluation. You can see the correlation between these evaluations in Table 5.

Table 5. Correlation between subjective and objective evaluations

| | Sc1 | | Sc9 | | Sc10 | |
	MOS	SSIM	MOS	SSIM	MOS	SSIM
Sequence 1: (15 Mbps, H.264, UHD)	3,8	0,929	2,27	0,968	3,6	0,962
Sequence 2: (10 Mbps, H.264, UHD)	3,43	0,903	3,1	0,961	3,57	0,955
Sequence 5: (15 Mbps, H.265, UHD)	3,6	0,942	2,87	0,974	3,73	0,971
Sequence 6: (10 Mbps, H.265, UHD)	3,47	0,929	3,17	0,969	3,67	0,967
Sequence 9: (5 Mbps, H.264, UHD)	1,67	0,82	2,53	0,938	2,73	0,928
Sequence 10: (5 Mbps, H.265, UHD)	3,63	0,894	2,97	0,953	3,57	0,956
Pearson correlation coefficient	\multicolumn{2}{c}{**0,917**}		\multicolumn{2}{c}{**0,981**}		\multicolumn{2}{c}{**0,968**}

7. Discussion

We need to consider the purpose and space of capturing individual moments when monitoring smart city footage. Depending on the importance of the captured part of the city, we can define the necessary quality of the recording. If we need to address security, we can use high-resolution security cameras such as Internet Protocol cameras (IP cameras) that can produce 4K resolution or better. However, for monitoring a certain event, checking the traffic, or monitoring a location with a static background, we don't need the best-resolution video. In this case, wireless cameras can be used, but their quality may not match the reality of the viewed footage. The quality of the footage may be limited by insufficient WiFi signal or a monitor/display with a lower resolution on which you would like to view the video footage. The selection of an individual system for deployment involves several important aspects. Our recommendation for setting the quality parameters can help determine appropriate parameters. We can define sufficient quality for different types of video sequences based on the deployment requirements. To achieve this, we had to create a large set of video sequences, some of which had to be recorded multiple times due to bad weather conditions or image interference. The final shots were of high quality, with different object dynamics in the scenes and dynamic camera movement.

We have created a database of 4K video sequences that cover scenes from traffic or city monitoring. Our goal is to expand this database with video sequences shot with different devices such as classic cameras, drones, mobile phones, and GoPro cameras. This will help us compare whether the quality is also affected by the camera on which the video sequences are shot. In the future, we plan to extend the encoded sequences with H.266 and AV1 codecs and bitrates of 1, 3, 7, and 20 Mbps to compare the ratings of other combinations of quality parameters. We are also considering using other metrics for objective evaluation and a larger sample for subjective evaluation.

We are looking for partners who can provide us with video sequences to improve our monitoring system. Our team is interested in collaborating with the city of Zilina to identify video sequences that could be used to enhance the system. We would also appreciate it if they could provide us with some of their own recordings. Furthermore, we are looking for a reliable security systems company to partner with and expand our database in the future. In addition, we are interested in working with partners who can help us film 8K sequences and expand our laboratory with 8K imaging units to perform subjective tests. Although we have reached out to other universities in Slovakia and the Czech Republic, the possibilities are currently limited.

The reference sequences of our database are available at [1]. All of them as well as the encoded sequences can be downloaded by researchers from the server using FTP (File Transfer Protocol). The FTP server is configured to allow passwordless access to users at IP address 158.193.214.161 via any FTP client. Once connected, the user has access to both reference and transcoded sequences. The "reference sequences" section contains the names defined in the description of the dataset, while the "encoded sequences" section contains sub-sequences for each resolution (Full HD, Ultra HD). The transcoded sequences' names are defined by the key: original sequence name_compression standard_bitrate. We have a test web page that is currently being finalized, which will contain these sequences, their descriptions, and a contact form where users can leave comments or advice. Until the website is launched, you can contact us by email for more information or to provide feedback.

Modern technology is rapidly developing in all areas of our society. However, the potential advantages and disadvantages of these technologies are often not sufficiently discussed. Although they can make our lives easier and more efficient, they can also have a negative impact on social relationships. An example is the use of industrial cameras in public spaces. CCTV cameras are used in public spaces primarily for monitoring and crime prevention. However, this type of surveillance raises human rights concerns that are often overlooked in discussions about the use of modern technology. CCTV is intended for places where increased security and public surveillance are needed, and smart technologies are used to create a safer environment. Video recordings do not target individuals or their

personal belongings, but rather are used for research purposes. Anyone who downloads sequences from our store agrees to this statement.

8. Conclusion

The purpose of this publication's research was to create 4K UHD video sequences to capture traffic conditions in the city and monitor specific areas. The footage was intended to be used to analyze quality requirements and provide recommendations for implementing technologies such as smart cities or smart traffic. To begin, we determined the types of video sequences that could be applicable in the smart cities or traffic sector. We selected video sequences that provided slower but also more dynamic shots, as well as video sequences where the camera movement was both static and dynamic, changing the characteristics of the footage. We identified individual video scenes through spatial and temporal information, knowing that camera movement also affects these values, producing a different type of video sequence. For transportation, we chose Zilina's available means of public transportation, specifically the trolleybus coming and going from the public transport stop, as well as its dynamic driving. We also recorded the traffic situation at lower and higher speeds, including busy roads, roundabouts, and parked vehicles. We focused on rail transport as well, recording slower trains arriving or leaving the station and faster moving trains. Selecting video sequences for smart cities was more difficult, as we needed to cover different dynamics. We chose a sequence that monitored the movement of people in a university town and flags flying as a demonstration of something that could be recorded. Monitoring systems record various situations, whether in the context of security or sensing different situations where the system helps to evaluate the appropriate response.

We used both objective and subjective methods to evaluate the tests conducted and based on the measurements obtained, we plan to propose a QoS model for estimating triple-play services in our future work. Our next focus is to assess the quality of video data delivery in various scenarios by simulating different values of packet loss and delay in the network. The results of this study will help us determine whether it is better for video quality to receive packets in the wrong order or to lose them entirely.

We plan to expand our database to include video sequences recorded with various devices, including mobile phones, GoPro cameras, and conventional 4K cameras. This comprehensive database will allow us to compare the resulting quality of the videos captured by different devices. These comparisons will help improve stream services. We will also develop a prediction model that can calculate the resulting video quality based on the network's state and behavior. This model can be used by ISPs during the network architecture design process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sevcik, L. UHD database focus on smart cities and smart transport, 2024. doi:10.5281/ZENODO.10663664.
2. Mercat, A.; Viitanen, M.; Vanne, J. UVG dataset: 50/120fps 4K sequences for video codec analysis and development. Proceedings of the 11th ACM Multimedia Systems Conference. ACM, 2020, MMSys '20. doi:10.1145/3339825.3394937.
3. Song, L.; Tang, X.; Zhang, W.; Yang, X.; Xia, P. The SJTU 4K video sequence dataset. 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). IEEE, 2013. doi:10.1109/qomex.2013.6603201.
4. Ghadiyaram, D.; Pan, J.; Bovik, A.C. A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos. *IEEE Transactions on Circuits and Systems for Video Technology* **2019**, *29*, 183–197. doi:10.1109/tcsvt.2017.2768542.
5. Hosu, V.; Hahn, F.; Jenadeleh, M.; Lin, H.; Men, H.; Sziranyi, T.; Li, S.; Saupe, D. The Konstanz natural video database (KoNViD-1k). 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2017. doi:10.1109/qomex.2017.7965673.

6. Bampis, C.G.; Li, Z.; Katsavounidis, I.; Huang, T.Y.; Ekanadham, C.; Bovik, A. Towards Perceptually Optimized End-to-end Adaptive Video Streaming. *arXiv: Image and Video Processing* **2018**.

7. Ghadiyaram, D.; Pan, J.; Bovik, A.C.; Moorthy, A.K.; Panda, P.; Yang, K.C. In-Capture Mobile Video Distortions: A Study of Subjective Behavior and Objective Algorithms. *IEEE Transactions on Circuits and Systems for Video Technology* **2018**, *28*, 2061–2077. doi:10.1109/tcsvt.2017.2707479.

8. Duanmu, Z.; Ma, K.; Wang, Z. Quality-of-Experience for Adaptive Streaming Videos: An Expectation Confirmation Theory Motivated Approach. *IEEE Transactions on Image Processing* **2018**, *27*, 6135–6146. doi:10.1109/tip.2018.2855403.

9. Sinno, Z.; Bovik, A.C. Large-Scale Study of Perceptual Video Quality. *IEEE Transactions on Image Processing* **2019**, *28*, 612–627. doi:10.1109/tip.2018.2869673.

10. Long, C.; Cao, Y.; Jiang, T.; Zhang, Q. Edge Computing Framework for Cooperative Video Processing in Multimedia IoT Systems. *IEEE Transactions on Multimedia* **2018**, *20*, 1126–1139. doi:10.1109/tmm.2017.2764330.

11. Li, M.; Chen, H.L. Energy-Efficient Traffic Regulation and Scheduling for Video Streaming Services Over LTE-A Networks. *IEEE Transactions on Mobile Computing* **2019**, *18*, 334–347. doi:10.1109/tmc.2018.2836421.

12. Grajek, T.; Stankowski, J.; Karwowski, D.; Klimaszewski, K.; Stankiewicz, O.; Wegner, K. Analysis of Video Quality Losses in Homogeneous HEVC Video Transcoding. *IEEE Access* **2019**, *7*, 96764–96774. doi:10.1109/access.2019.2928423.

13. Ramachandra Rao, R.R.; Goring, S.; Robitza, W.; Feiten, B.; Raake, A. AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1. 2019 IEEE International Symposium on Multimedia (ISM). IEEE, 2019. doi:10.1109/ism46123.2019.00012.

14. Bouaafia, S.; Khemiri, R.; Sayadi, F.E. Rate-Distortion Performance Comparison: VVC vs. HEVC. 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD). IEEE, 2021. doi:10.1109/ssd52085.2021.9429377.

15. Mercat, A.; Makinen, A.; Sainio, J.; Lemmetti, A.; Viitanen, M.; Vanne, J. Comparative Rate-Distortion-Complexity Analysis of VVC and HEVC Video Codecs. *IEEE Access* **2021**, *9*, 67813–67828. doi:10.1109/access.2021.3077116.

16. García-Lucas, D.; Cebrián-Márquez, G.; Cuenca, P. Rate-distortion/complexity analysis of HEVC, VVC and AV1 video codecs. *Multimedia Tools and Applications* **2020**, *79*, 29621–29638. doi:10.1007/s11042-020-09453-w.

17. Topiwala, P.; Krishnan, M.; Dai, W. Performance comparison of VVC, AV1 and EVC. Applications of Digital Image Processing XLII; Tescher, A.G.; Ebrahimi, T., Eds. SPIE, 2019. doi:10.1117/12.2530559.

18. Nguyen, T.; Wieckowski, A.; Bross, B.; Marpe, D. Objective Evaluation of the Practical Video Encoders VVenC, x265, and aomenc AV1. 2021 Picture Coding Symposium (PCS). IEEE, 2021. doi:10.1109/pcs50896.2021.9477482.

19. Nguyen, T.; Marpe, D. Compression efficiency analysis of AV1, VVC, and HEVC for random access applications. *APSIPA Transactions on Signal and Information Processing* **2021**, *10*. doi:10.1017/atsip.2021.10.

20. Valiandi, I.; Panayides, A.S.; Kyriacou, E.; Pattichis, C.S.; Pattichis, M.S., A Comparative Performance Assessment of Different Video Codecs. In *Lecture Notes in Computer Science*; Springer Nature Switzerland, 2023; p. 265–275. doi:10.1007/978-3-031-44240-7_26.

21. Nguyen, T.; Marpe, D. Future Video Coding Technologies: A Performance Evaluation of AV1, JEM, VP9, and HM. 2018 Picture Coding Symposium (PCS). IEEE, 2018. doi:10.1109/pcs.2018.8456289.

22. Pourazad, M.T.; Sung, T.; Hu, H.; Wang, S.; Tohidypour, H.R.; Wang, Y.; Nasiopoulos, P.; Leung, V.C. Comparison of Emerging Video Compression Schemes for Efficient Transmission of 4K and 8K HDR Video. 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). IEEE, 2021. doi:10.1109/meditcom49071.2021.9647504.

23. Grois, D.; Giladi, A.; Choi, K.; Park, M.W.; Piao, Y.; Park, M.; Choi, K.P. Performance Comparison of Emerging EVC and VVC Video Coding Standards with HEVC and AV1. SMPTE 2020 Annual Technical Conference and Exhibition. IEEE, 2020. doi:10.5594/m001916.

24. Haiqiang Wang, I.K. VideoSet: A Large-Scale Compressed Video Quality Dataset Based on JND Measurement, 2016. doi:10.21227/H2H01C.

25. Karthikeyan, V.; Allan, B.; Nauck, D.D.; Rio, M. Benchmarking Video Service Quality: Quantifying the Viewer Impact of Loss-Related Impairments. *IEEE Transactions on Network and Service Management* **2020**, *17*, 1640–1652. doi:10.1109/tnsm.2020.2998905.

26. Kazemi, M.; Ghanbari, M.; Shirmohammadi, S. The Performance of Quality Metrics in Assessing Error-Concealed Video Quality. *IEEE Transactions on Image Processing* **2020**, *29*, 5937–5952. doi:10.1109/tip.2020.2984356.

27. Diaz, C.; Perez, P.; Cabrera, J.; Ruiz, J.J.; Garcia, N. XLR (piXel Loss Rate): A Lightweight Indicator to Measure Video QoE in IP Networks. *IEEE Transactions on Network and Service Management* **2020**, *17*, 1096–1109. doi:10.1109/tnsm.2020.2980752.

28. Silva, C.A.G.D.; Pedroso, C.M. MAC-Layer Packet Loss Models for Wi-Fi Networks: A Survey. *IEEE Access* **2019**, *7*, 180512–180531. doi:10.1109/access.2019.2958260.

29. Neves, F.; Soares, S.; Assuncao, P.A.A. Optimal voice packet classification for enhanced VoIP over priority-enabled networks. *Journal of Communications and Networks* **2018**, *20*, 554–564. doi:10.1109/jcn.2018.000088.

30. Katsenou, A.V.; Dimitrov, G.; Ma, D.; Bull, D.R. BVI-SynTex: A Synthetic Video Texture Dataset for Video Compression and Quality Assessment. *IEEE Transactions on Multimedia* **2021**, *23*, 26–38. doi:10.1109/tmm.2020.2976591.

31. Badidi, E.; Moumane, K.; Ghazi, F.E. Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review. *IEEE Access* **2023**, *11*, 80543–80572. doi:10.1109/access.2023.3300658.

32. Chen, Y.Y.; Lin, Y.H.; Hu, Y.C.; Hsia, C.H.; Lian, Y.A.; Jhong, S.Y. Distributed Real-Time Object Detection Based on Edge-Cloud Collaboration for Smart Video Surveillance Applications. *IEEE Access* **2022**, *10*, 93745–93759. doi:10.1109/access.2022.3203053.

33. Yun, Q.; Leng, C. Intelligent Control of Urban Lighting System Based on Video Image Processing Technology. *IEEE Access* **2020**, *8*, 155506–155518. doi:10.1109/access.2020.3019284.

34. Smida, E.B.; Fantar, S.G.; Youssef, H. Video streaming challenges over vehicular ad-hoc networks in smart cities. 2017 International Conference on Smart, Monitored and Controlled Cities (SM2C). IEEE, 2017. doi:10.1109/sm2c.2017.8071838.

35. Duan, Z.; Yang, Z.; Samoilenko, R.; Oza, D.S.; Jagadeesan, A.; Sun, M.; Ye, H.; Xiong, Z.; Zussman, G.; Kostic, Z. Smart City Traffic Intersection: Impact of Video Quality and Scene Complexity on Precision and Inference. 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). IEEE, 2021. doi:10.1109/hpcc-dss-smartcity-dependsys53884.2021.00226.

36. Malik, M.; Prabha, C.; Soni, P.; Arya, V.; Alhalabi, W.A.; Gupta, B.B.; Albeshri, A.A.; Almomani, A. Machine Learning-Based Automatic Litter Detection and Classification Using Neural Networks in Smart Cities. *International Journal on Semantic Web and Information Systems* **2023**, *19*, 1–20. doi:10.4018/ijswis.324105.

37. Li, B.; Zhang, W.; Tian, M.; Zhai, G.; Wang, X. Blindly Assess Quality of In-the-Wild Videos via Quality-Aware Pre-Training and Motion Perception. *IEEE Transactions on Circuits and Systems for Video Technology* **2022**, *32*, 5944–5958. doi:10.1109/tcsvt.2022.3164467.

38. Lee, S.; Roh, H.; Lee, N. Enhanced quality adaptation scheme for improving QoE of MPEG DASH. 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017. doi:10.1109/ictc.2017.8191002.

39. Chang, S.H.; Wang, K.J.; Ho, J.M. Optimal DASH Video Scheduling over Variable-Bit-Rate Networks. 2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP). IEEE, 2018. doi:10.1109/paap.2018.00015.

40. Mizdos, T.; Barkowsky, M.; Uhrina, M.; Pocta, P. How to reuse existing annotated image quality datasets to enlarge available training data with new distortion types. *Multimedia Tools and Applications* **2021**, *80*, 28137–28159. doi:10.1007/s11042-021-10679-5.

41. Sevcik, L.; Voznak, M. Adaptive Reservation of Network Resources According to Video Classification Scenes. *Sensors* **2021**, *21*, 1949. doi:10.3390/s21061949.

42. ITU-T. Recommendation ITU-T P.800.1 - Mean opinion score (MOS) terminology. *[online] Available: https://www.itu.int/rec/T-REC-P.800.1* **2016**.
43. ITU-T. Recommendation ITU-T P.1204.5 - Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to transport and received pixel information. *[online] Available: https://www.itu.int/rec/T-REC-P.910-200804-I/en* **2023**.