INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 2
MPEG TECHNICAL REQUIREMENTS

ISO/IEC JTC 1/SC 29/WG 2 **M57456**
**Online – July 2021**

**Title:** **[VCM] New results on analysis of influence of HEVC and VVC coding on the SIFT keypoints extracted from the decoded video**

**Source:** **WG 2 MPEG Technical requirements**

**Author(s):** **Sławomir Różek, Marek Domański, Sławomir Maćkowiak, Olgierd Stankiewicz, Jakub Stankowski**
**Poznan University of Technology, Poznań, Poland**

**Status:** **Input**

## Abstract

This document provides new, updated results of analysis of the number and parameters fidelity of the SIFT keypoints extracted from decoded video. Both video coding techniques, HEVC and VVC have been used and the parameters of extracted keypoints have been compared and matched with these extracted from uncompressed sequences. According to the matching results, keypoints have been divided into four categories, which allow for partial transmission of features for VCM. The experimental results provide the relation between quantization parameter and the number of keypoints retrieved from the decoded video and their parameters correctness. The method of differential keypoints transmission was assumed, and the estimation of required data was done.

## 1. Introduction

Prospective technology of Video Coding for Machines allows for parallel transmission of video and features extracted from original video [1]. However such method can introduce some redundancy of information into bitstream. In fact some of the features can be obtained also from decoded video. The potential mismatch in parameters between original features and these extracted from decompressed video can be send as a correction – enhancement data for features. Such approach have been considered in [2]. To design such efficient mechanism of differential transmission of features, the statistics of features extracted from decoded video must be well known. A preliminary analysis of that have been done in [3] and [2]. At the current stage we investigate more precisely how the video compression (HEVC [4][5] and VVC [6][7]) impacts on SIFT[8] keypoints location, their parameters and usability in partial transmission of features. Also some criteria in this experiment have been slightly changed with reference to previous experiments. In this research two HD sequences have been used (*Poznan Carpark*, *Poznan Street*) [9], which have been proposed to be test data for VCM [10].

## 2. Video encoder configuration and encoding results

### 2.1. HEVC encoder parameters

HM software [6]: Encoder Version [16.20] (including RExt)[Linux][GCC 9.2.1][64 bit]

```
Real     Format                           : 1920x1088 25Hz
Internal Format                           : 1920x1088 25Hz
Profile                                   : main
CU size / depth / total-depth             : 64 / 4 / 4
RQT trans. size (min / max)               : 4 / 32
Max RQT depth inter                       : 3
Max RQT depth intra                       : 3
Min PCM size                              : 8
Motion search range                       : 384
Intra period                              : 32
Decoding refresh type                     : 1
QP                                        :  from 17 to 47
GOP size                                  : 16
Input bit depth                           : (Y:8, C:8)
MSB-extended bit depth                    : (Y:8, C:8)
Internal bit depth                        : (Y:8, C:8)
PCM sample bit depth                      : (Y:8, C:8)
Intra reference smoothing                 : Enabled
Input ChromaFormatIDC                     =   4:2:0
Output (internal) ChromaFormatIDC         =   4:2:0
```

The following encoder tool parameters were set:

TOOL CFG: IBD:0 HAD:1 RDQ:1 RDQTS:1 RDpenalty:0 LQP:0 SQP:0 ASR:1 MinSearchWindow:96 RestrictMESampling:0 FEN:1 ECU:0 FDM:1 CFM:0 ESD:0 RQT:1 TransformSkip:1 TransformSkipFast:1 TransformSkipLog2MaxSize:2 Slice: M=0 SliceSegment: M=0 CIP:0 SAO:1 PCM:0 TransQuantBypassEnabled:0 WPP:0 WPB:0 PME:2 WaveFrontSynchro:0 WaveFrontSubstreams:1 ScalingList:0 TMVPMode:1 AQpS:0 SignBitHidingFlag:1 RecalQP:0

## 2.2. VVC encoder parameters

VVCSoftware [9]: VTM Encoder Version 11.0 [Linux][GCC 9.3.0][64 bit] [SIMD=AVX2]

| | |
|---|---|
| Real Format | : 1920x1088 25Hz |
| Internal Format | : 1920x1088 25Hz |
| Profile | : main_10 |
| CTU size / min CU size | : 128 / 4 |
| Motion search range | : 384 |
| Intra period | : 32 |
| Decoding refresh type | : 1 |
| DRAP period | : 0 |
| QP | : from 17 to 47 |
| GOP size | : 32 |
| Input bit depth | : (Y:8, C:8) |
| MSB-extended bit depth | : (Y:8, C:8) |
| Internal bit depth | : (Y:10, C:10) |
| Intra reference smoothing | : Enabled |
| Input ChromaFormatIDC | = 4:2:0 |
| Output (internal) ChromaFormatIDC | = 4:2:0 |

The following encoder tool parameters were set:

TOOL CFG: IBD:1 HAD:1 RDQ:1 RDQTS:1 RDpenalty:0 LQP:0 SQP:0 ASR:1 MinSearchWindow:96 RestrictMESampling:0 FEN:1 ECU:0 FDM:1 ESD:0 TransformSkip:1 TransformSkipFast:1 TransformSkipLog2MaxSize:5 ChromaTS:1 BDPCM:0 Tiles: 1x1 Slices: 1 MCTS:0 SAO:1 ALF:1 CCALF:1 WPP:0 WPB:0 PME:2 WaveFrontSynchro:0 WaveFrontSubstreams:1 ScalingList:0 TMVPMode:1 DQ:1 SignBitHidingFlag:0 RecalQP:0 TOOL CFG: LFNST:1 MMVD:1 Affine:1 AffineType:1 PROF:1 SbTMVP:1 DualITree:1 IMV:1 BIO:1 LMChroma:1 HorCollocatedChroma:1 VerCollocatedChroma:0 MTS: 1(intra) 0(inter) SBT:1 ISP:1 SMVD:1 CompositeLTReference:0 Bcw:1 BcwFast:1 LADF:0 CIIP:1 Geo:1 AllowDisFracMMVD:1 AffineAmvr:1 AffineAmvrEncOpt:1 DMVR:1 MmvdDisNum:6 JointCbCr:1 ACT:0 PLT:0 IBC:0 HashME:0 WrapAround:0 VirtualBoundariesEnabledFlag:0 VirtualBoundariesPresentInSPSFlag:1 vertical virtual boundaries:[ ] horizontal virtual boundaries:[ ] Reshape:1 (Signal:SDR Opt:0 CSoffset:6) MRL:1 MIP:1 EncDbOpt:0 FAST TOOL CFG: LCTUFast:1 FastMrg:1 PBIntraFast:1 IMV4PelFast:1 MTSMaxCand: 4(intra) 4(inter) ISPFast:0 FastLFNST:0 AMaxBT:1 E0023FastEnc:1 ContentBasedFastQtbt:0 UseNonLinearAlfLuma:1 UseNonLinearAlfChroma:1 MaxNumAlfAlternativesChroma:8 FastMIP:0 FastLocalDualTree:1 NumSplitThreads:1 NumWppThreads:1+0 EnsureWppBitEqual:0 RPR:0 TemporalFilter:1

### 2.3. Encoding results

The relation between bitrate and QP parameter for both codecs and sequences is shown in Figure 2.1.

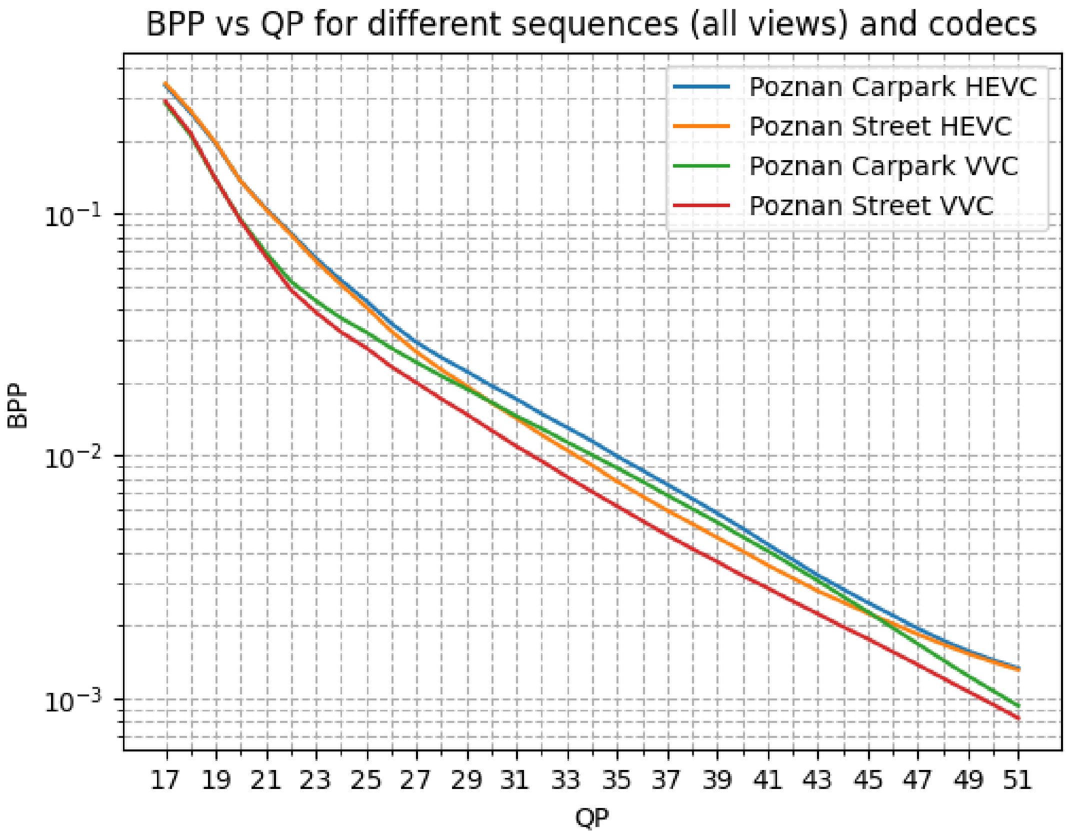

Fig. 2.1. Bitrate over QP parameter for both sequences and codecs.

# 3. The influence of HEVC and VVC compression on the ability to determine the SIFT keypoints in the decoded video

## 3.1. Influence of HEVC and VVC compression rates on the number of SIFT keypoints in decoded video

This part of experiment shows how video encoding impacts on the general number of SIFT keypoints detected in decoded video. The *Poznan Street* and *Poznan Carpark* sequences were encoded at 1920x1088 resolution using both HEVC and VVC encoders with seven different quantization factors (QP=17, 22, 27, 32, 37, 42 and 47) and then decoded. SIFT feature detector/extractor from OpenCV (version 4.3.0) and Python have been used to determine characteristic points in each decoded frame. The number of layers in an octave was left the original equal to 3. We left the sigma parameter at the default value, i.e. 1.6.

The keypoints extracted from uncompressed *Poznan Street* and *Poznan Carpark* sequences are used as a reference for comparison. Results were accumulated and averaged for 250 frames of sequence. The block diagram illustrating the experiment is presented in Fig. 3.1.

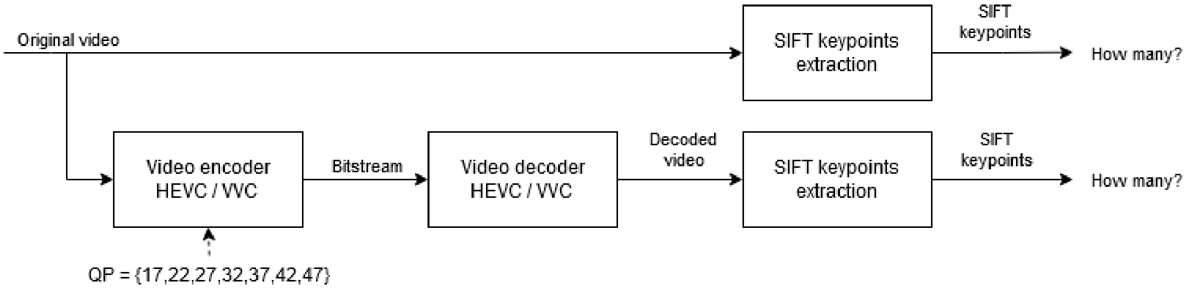

Fig. 3.1. Block diagram of the first experiment.

Figure 3.2. depicts the absolute numbers of extracted SIFT keypoints for both sequences and both codecs versus the quantization parameter (QP). Dashed, horizontal lines indicate numbers of keypoints found in uncompressed sequences.

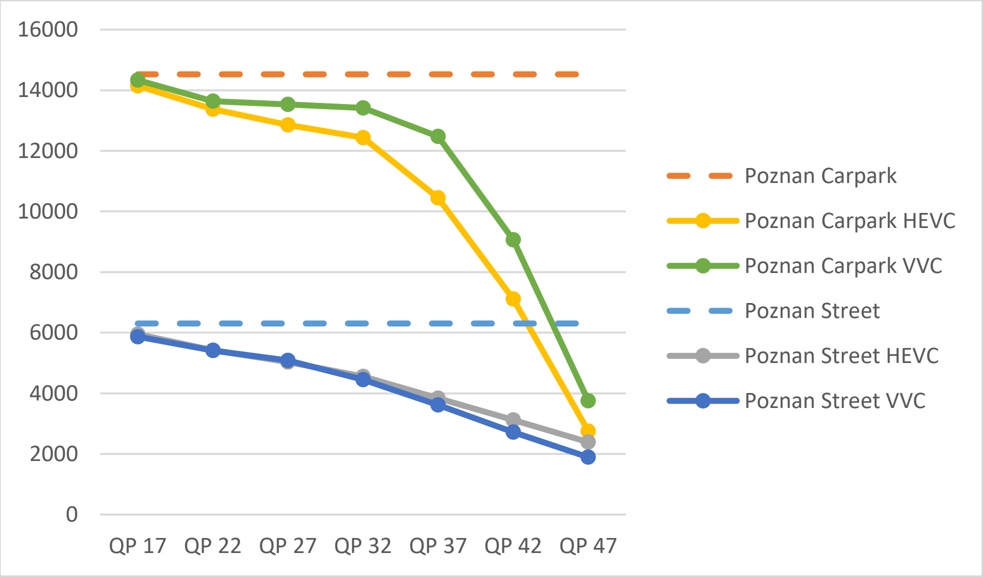

Fig. 3.2. The numbers of SIFT keypoints extracted from original and decoded sequences.

As can be seen when the QP increases, the number of keypoints extracted from decoded image decreases. However, from the graph one can conclude that roughly up to the QP value of 32, the number of points remains fairly constant and then rapidly decreases. This observation remains valid for both HEVC and VVC compression.

### 3.2. Comparison of SIFT keypoints between the original image and the decoded image

Next part of experiment investigates how the keypoints extracted from decoded sequences correspond to the original ones (Fig. 3.3.).

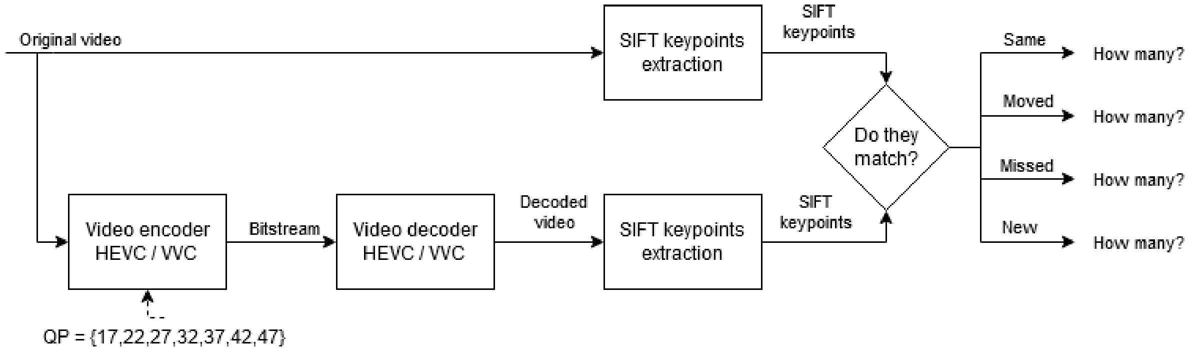

Fig. 3.3. Block diagram of the second experiment.

The main criteria of the keypoint matching was the location and octave index (see [3]). Regarding that four categories of keypoints have been introduced:

- *Same* – the keypoints with the same location as in original video (no shift permitted) and detected in the same octave.
- *Moved* – the keypoints detected in the same octave, but shifted by one sampling period (in original resolution) in one or two directions; the keypoints inside 3×3 window around original keypoint.
- *Missed* – the keypoints which are present in original frames, but have no corresponding ones in the decoded sequence; the keypoints lost due to compression.
- *New* – the keypoints found in the decoded frames, but without corresponding ones in the original sequence.

So, the keypoints from original sequence contain *Same*, *Moved* and *Missed* classes. On the other hand, the keypoints extracted from decoded frames contain *Same*, *Moved* and *New* classes. It is also worth to add that due to strict conditions for *Moved* class, some of the keypoints may have corresponding ones, but moved outside 3×3 window, so they are classified as *Missed* and *New*, instead of *Moved*.

In the case of multiple match between original keypoints and these from decompressed video, the pairs with most similar angles were chosen.

Figures 3.4-3.7 shows the counts of different categories as a function of quantization parameter for different sequences and codecs.

From the graphs can be seen that the *Same* keypoints constitute at most one third of all keypoints detected in compressed video. Most of the original keypoints are *Missed*, and these number increase with QP.

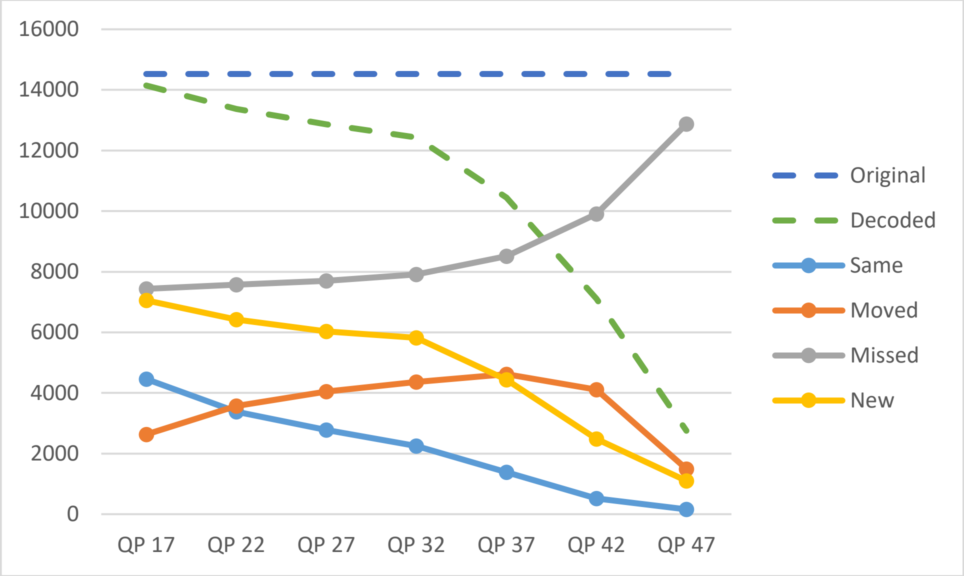

Fig. 3.4. Counts of different categories of keypoints for *Poznan Carpark* sequence, HEVC encoder.

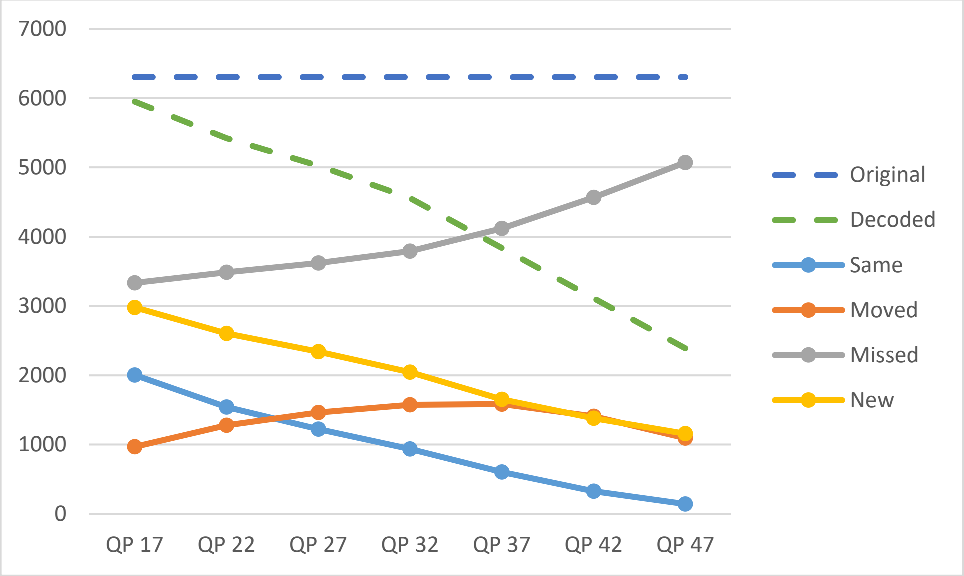

Fig. 3.5. Counts of different categories of keypoints for *Poznan Street* sequence, HEVC encoder

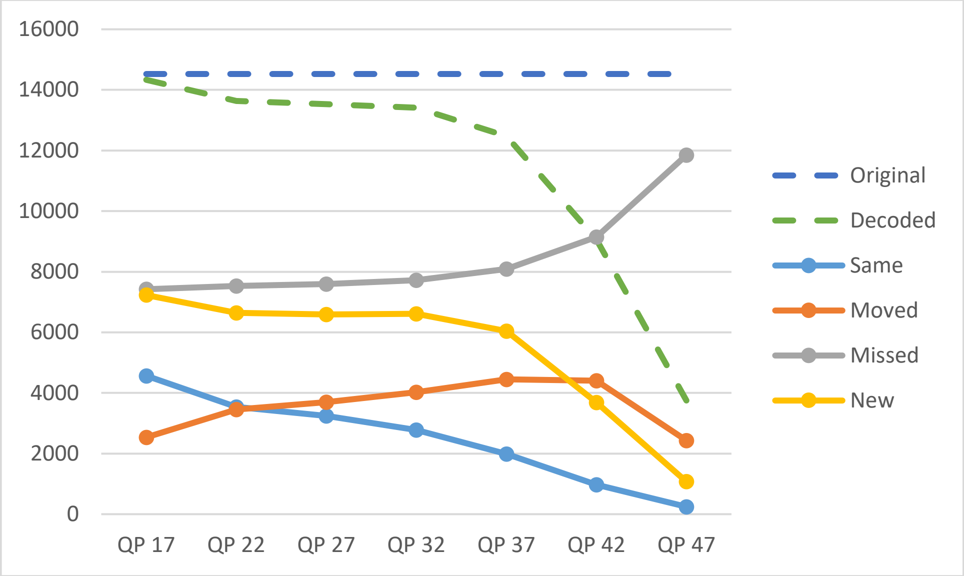

Fig. 3.6. Counts of different categories of keypoints for *Poznan Carpark* sequence, VVC encoder.

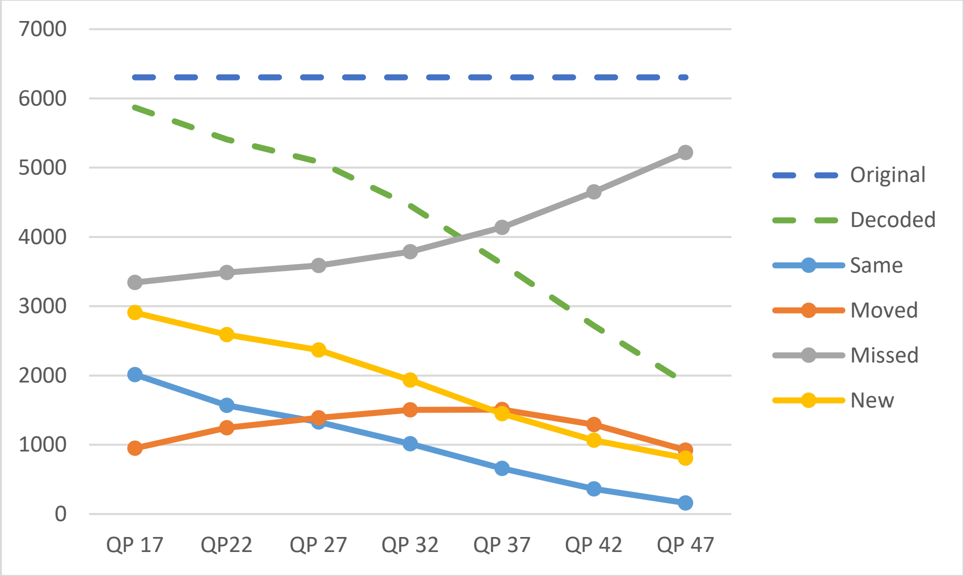

Fig. 3.7. Counts of different categories of keypoints for *Poznan Street* sequence, VVC encoder.

### 3.3. Analysis of the corectness of keypoints parameters in some categories

In the next step, the correctness of *orientation (angle)* and *size* parameters for *Same* and *Moved* keypoints have been investigated. The main assumption was that the *orientation* parameter of the keypoint is correct (*ok angle*), when the difference does not exceed ±5 degrees. The *size* parameter was correct (*ok size*) when the error of this parameter does not exceed ±5%. The counts for *Same* keypoints category is depicted on the figures 3.9-3.12.

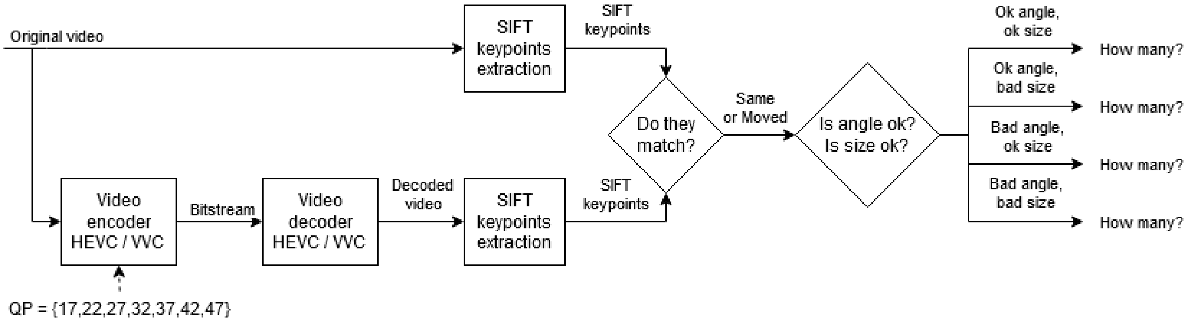

Fig. 3.9. Block dagram of the third experiment.

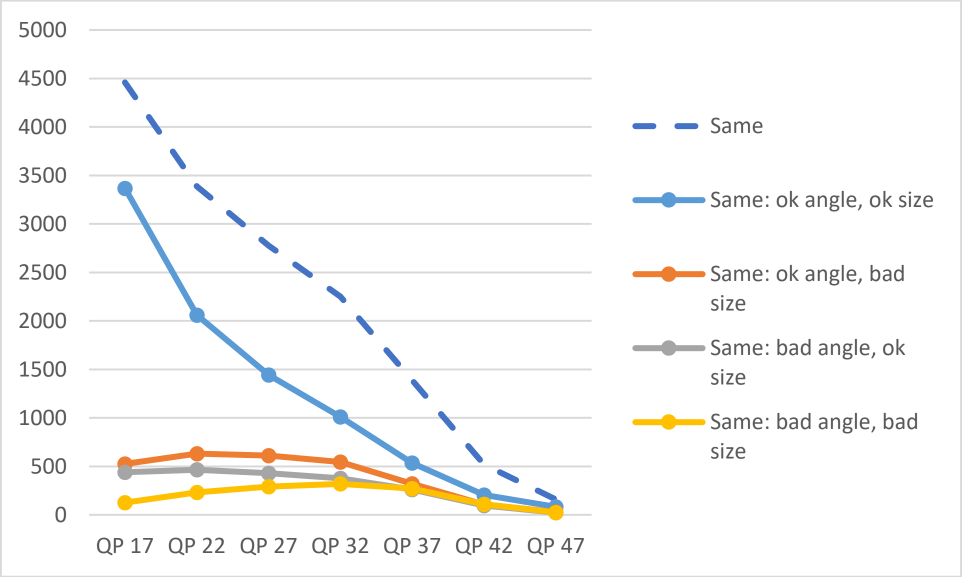

Fig. 3.9. Counts of *Same* keypoints for *Poznan Carpark*, HEVC encoder.

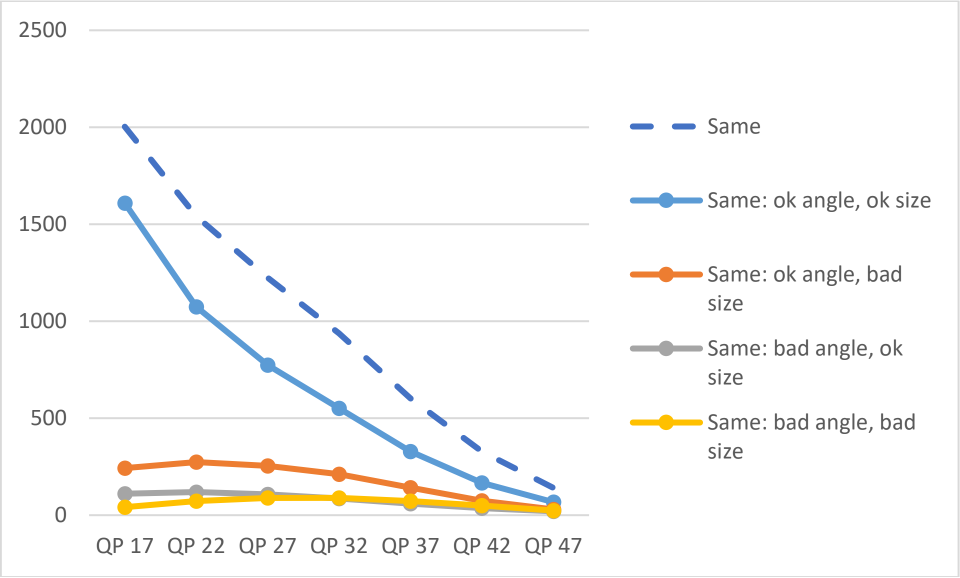

Fig. 3.10. Counts of *Same* keypoints for *Poznan Street*, HEVC encoder.

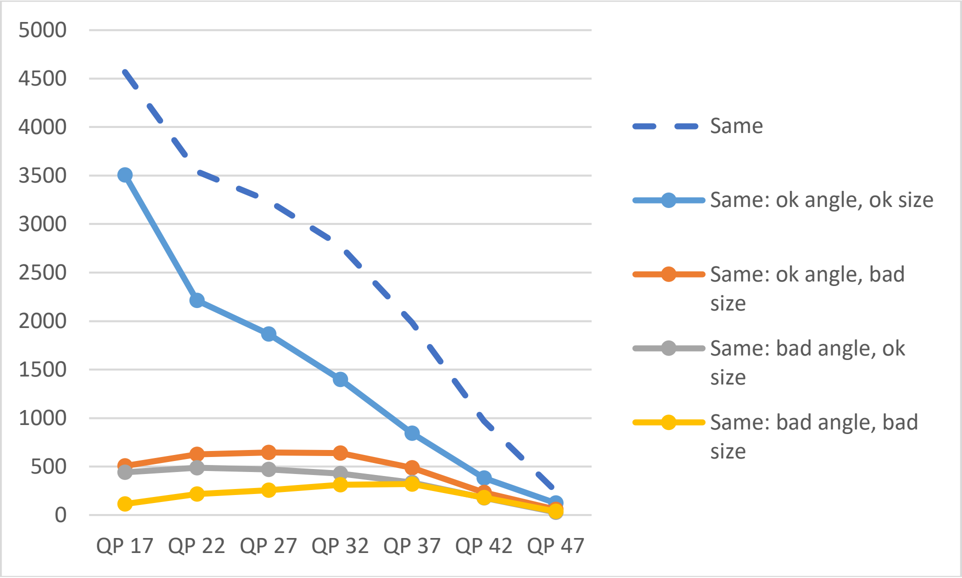

Fig. 3.11. Counts of *Same* keypoints for *Poznan Carpark*, VVC encoder.

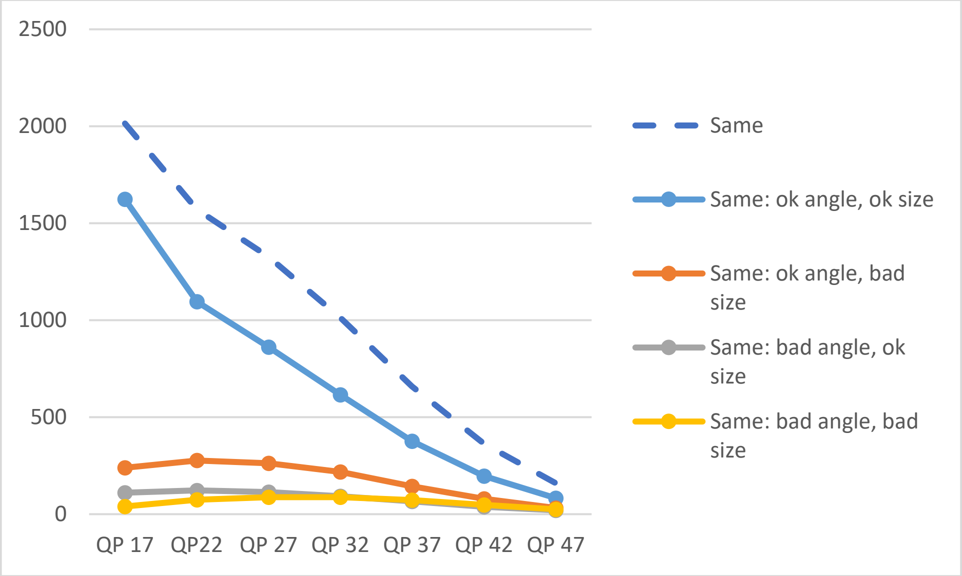

Fig. 3.12. Counts of *Same* keypoints for *Poznan Street*, VVC encoder.

As can be seen, for each QP most of the keypoints in *Same* category has undistorted both *orientation* and *size* parameters. Most of the rest require correction for only one parameter. Only small number of *Same* keypoints require corrections for both parameters at the same time.

Figures 3.13-3.16 show results of similar analysis for *Moved* category of the extracted keypoints.

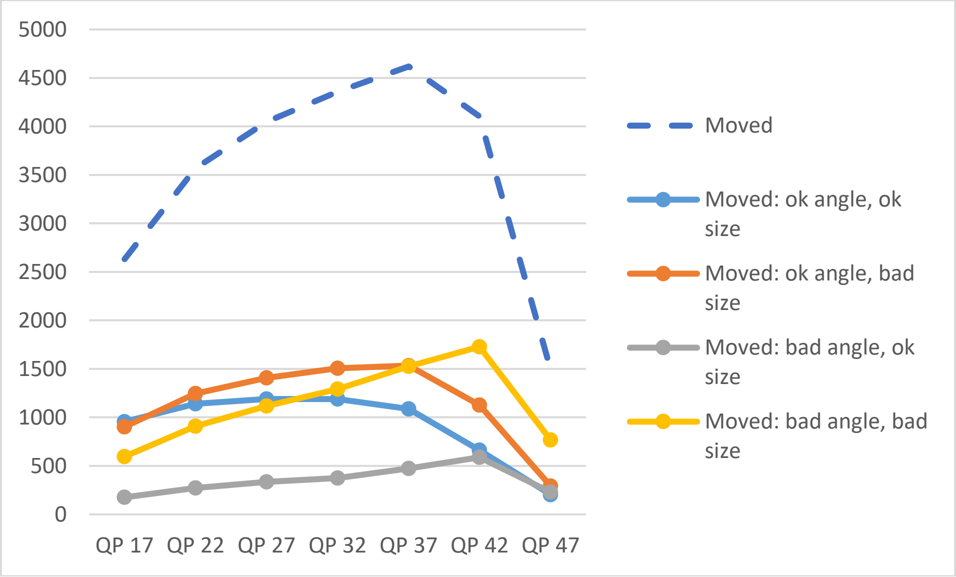

Fig. 3.13. Counts of *Moved* keypoints for *Poznan Carpark*, HEVC encoder.

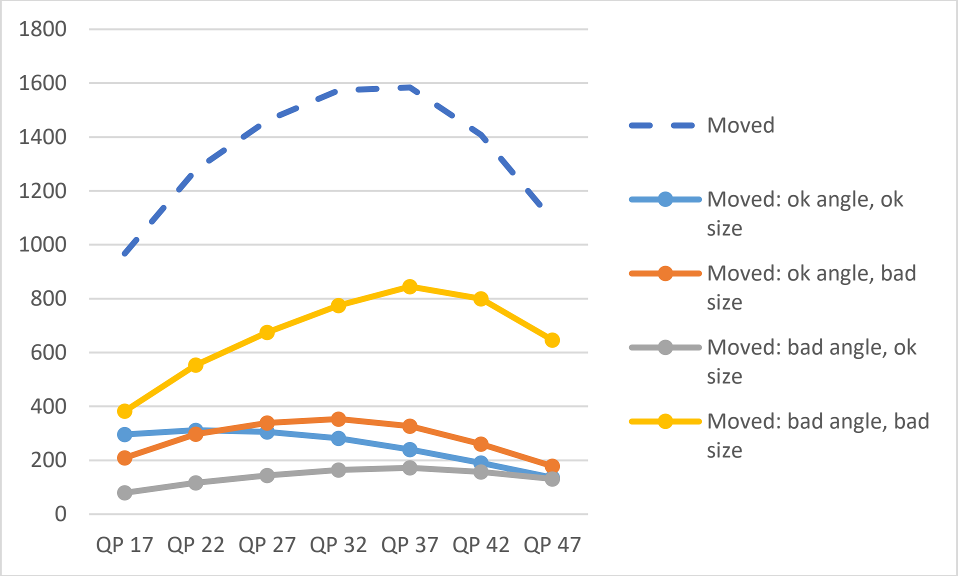

Fig. 3.14. Counts of *Moved* keypoints for *Poznan Street*, HEVC encoder.

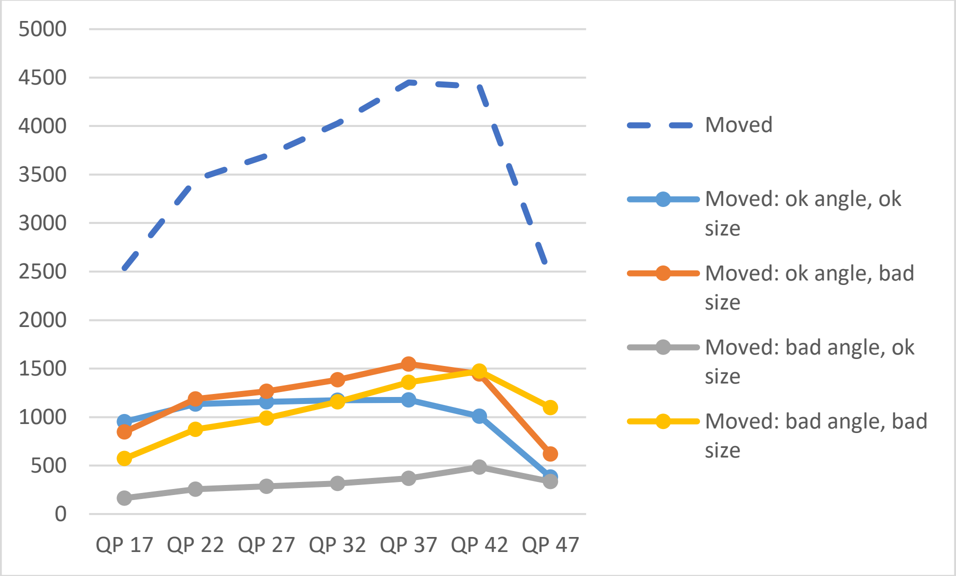

Fig. 3.15. Counts of *Moved* keypoints for *Poznan Carpark*, VVC encoder.

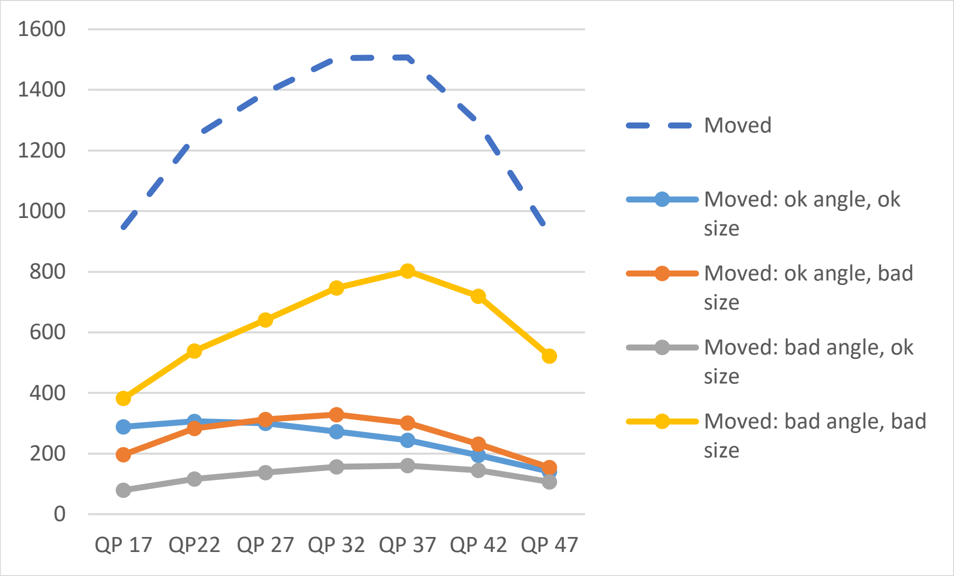

Fig. 3.16. Counts of *Moved* keypoints for *Poznan Street*, VVC encoder.

The analysis shows that in the *Moved* category the *orientation* or *size* parameters are distorted more frequently. In general, in this category, at least one parameter require correction.

## 3.4. Estimated amount of data required for differential transmission of features

The Table 3.1 presents assumed amount of data per category, needed for refinement features extracted from decompressed video.

Table. 3.1. Numbers of bits required for proposed transmission of parameters/corrections for each keypoint category.

| | *Same* | | | | *Moved* | | | | *Missed* | *New* |
|---|---|---|---|---|---|---|---|---|---|---|
| | *ok angle, ok size* | *ok angle, bad size* | *bad angle, ok size* | *bad angle, bad size* | *ok angle, ok size* | *ok angle, bad size* | *bad angle, ok size* | *bad angle, bad size* | | |
| *position x* | *0* | | | | *3* | | | | *11* | *0* |
| *position y* | | | | | | | | | *10* | *0* |
| *angle* | *0* | *0* | *6* | *6* | *0* | *0* | *6* | *6* | *6* | *0* |
| *size* | *0* | *6* | *0* | *6* | *0* | *6* | *0* | *6* | *8* | *0* |
| *id* | *14* | | | | *14* | | | | *0* | *0* |

The *Same* keypoints do not need any position correction. All of the positions (excluding the central) inside 3×3 window for *Moved* keypoints (8 possible locations) can be enumerated and encoded using 3 bits.

The previous assumption regarding permissible *orientation* error was ±5 degrees. It implies that the 360 degrees interval is divided into 36 bins. So coding of correction for quantized *orientation* parameter requires 6 bits..

The *Same* and *Moved* class contain only keypoints matched within certain octave, so the possible error of *size* is also limited. For the sake of simplicity we assume that the correction can be stored using 6 bits.

Some additional data (*id*) is required for *Same* and *Moved* keypoints identification in the decoder. If the total number of keypoints in decoded frame does not exceed 16 384, the identification can be done using 14 bits.

After correction of all *Same* and *Moved* keypoints, the remaining ones from decoded frame are assumed as *New*. They should be rejected without any correction or identification data.

Finally, the *Missed* keypoints have to be transmitted from the encoder with all of the parameters. Here, for HD frames we require 11 bits for horizontal position and about 10 bits for vertical position. *Orientation* parameter is represented by 6 bits, and the *size* parameter by 8 bits. Fortunately, the *Missed* keypoints do not need to be labeled with *id*.

Above analysis is very coarse, without exploiting statistics of parameters and prediction between neighboring keypoints.

To assess proposed method of differential transmission of keypoints, it must be compared with full transmission of original keypoints from uncompressed sequence. That without compression, would be done as transmission of *Missed* keypoints, so it would consume 35 bits for each original keypoint. The figure 3.17 shows the total data (in bits) required for partial transmission of SIFT keypoints for different sequences as a function of QP parameter. The particular quantities of data were calculated as weighted sum of numbers of keypoints in each category, and numbers of bits required for correction in the category. Horizontal, dashed lines refer to full transmission for original keypoints.

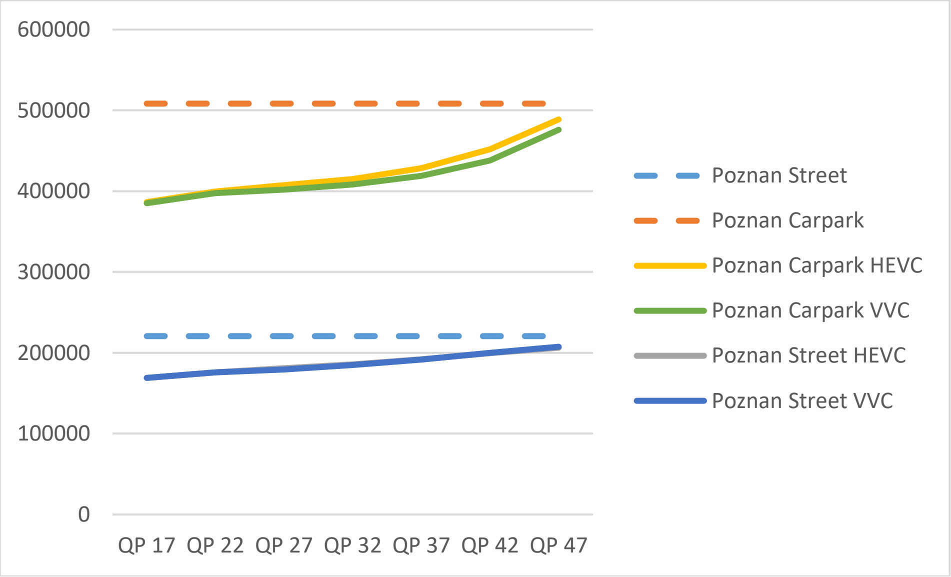

Fig. 3.17. Total data required for partial transmission of SIFT keypoints.

Figure 3.18 illustrates the percentage relation between data required for full transmission vs. partial transmission. As can be seen the partial transmission always reduces the size of features stream.

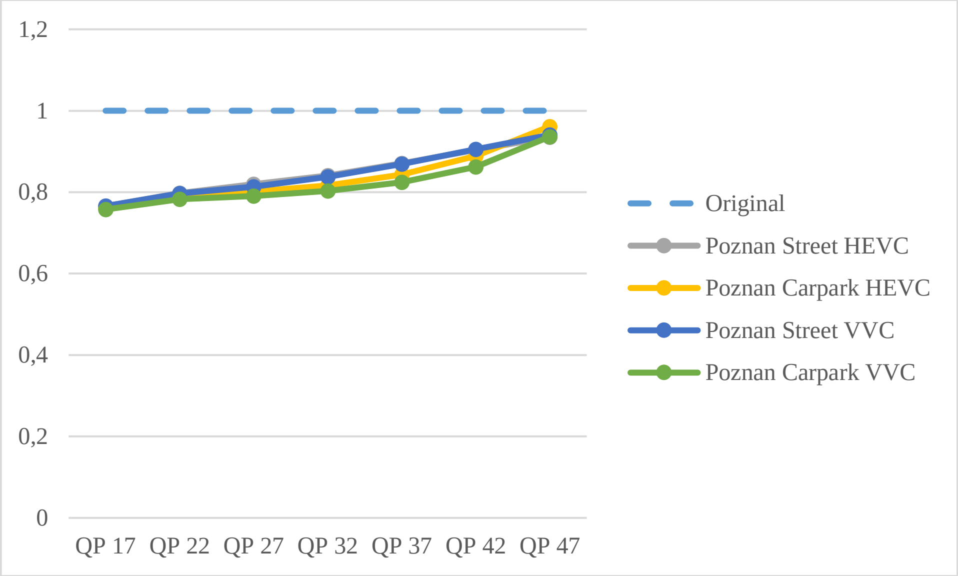

Fig. 3.18. Percentage relation of data between full and partial transmission.

## 4. Conclusions

The experiments show that even for simple differential transmission of SIFT keypoints, without using sophisticated mechanisms of entropy coding, the quantity of data to be transmitted can be reduced by up to 20%. Biggest savings can be achieved when the quantization parameter for video encoding is relatively low.

The presented method can be developed in the future, e.g. the entropy coding of parameters or the prediction using neighboring keypoints can be applied. Additional research should be also done for keypoints extracted from different types of frames (I, B, P). The method with some modifications should be also tested for another types of features (eg. SURF, HoG).

## 5. Acknowledgment

This work was supported by Project 0314/050.

## 6. References

[1] “Use cases and requirements for Video Coding for Machines,” Doc. ISO/IEC JTC1/SC29/WG11 N19506, June 2020.

[2] S. Mackowiak, M. Domański, D. Cywiński, J. Szekiełda, „[VCM] Partial transmission of SIFT features with compressed video”, Doc. ISO/IEC JTC 1/SC 29/WG 2 M56679, April 2021.

[3] S. Mackowiak, M. Domański, J. Stankowski, D. Cywiński, J. Szekiełda, „[VCM] Influence of HEVC and VVC coding on the SIFT characteristic points extracted from the received video”, Doc. ISO/IEC JTC 1/SC 29/WG 2 M56679, April 2021.

[4] Rec. ITU-T H.265 | ISO/IEC 23008-2 High efficiency video coding.

[5] https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tree/HM-16.20.

[6] Rec. ITU-T H.266 | ISO/IEC 23090-3 Versatile video coding.

[7] https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-11.0.

[8] Lowe D. G., “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60(2), 2004, pp91-110.

[9] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, “Poznan multiview video test sequences and camera parameters”, ISO/IEC JTC1/SC29/WG11 MPEG Doc. M17050, Xian, China, Oct. 2009.

[10] M. Domański, T. Grajek, S. Mackowiak, D. Mieloch, S. Różek, O. Stankiewicz, J. Stankowski, „[VCM] Test material for stereoscopic and multiview video coding for machines”, Doc. ISO/IEC JTC 1/SC 29/WG 2 M55108, October 2020.