INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 4
MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 4 m62236
January 2023, Online

Title: [VCM] CE1.7 - Implementation of Poznan University of Technology tools from Proposal C in Video Coding for Machines Reference Software

Authors:
Marek Domański, Olgierd Stankiewicz, Sławomir Maćkowiak, Tomasz Grajek, Maciej Wawrzyniak, Jakub Stankowski, Sławomir Różek, Dominik Cywiński, Jakub Szekiełda, Jakub Siejak

Poznań University of Technology, Poznań, Poland

Abstract

1 Introduction

This document described RoI-based simplification and retargeting method implemented in VCM-RS software [1]. The method originally has been implemented in Poznan University of Technology responses (proposal C) [2,3,4] to "Call for Proposals for Video Coding for Machines" [CFP].

The important part of the contribution is the detailed description of the bitsream syntax.

2 General description

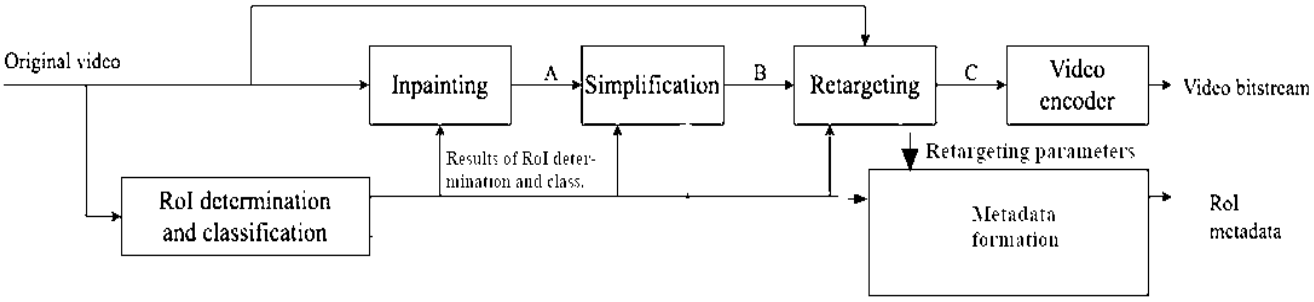

Fig. 1. The structure of the encoder

The block “Video encoder” is now called as inner encoder, here this is a VVC encoder.

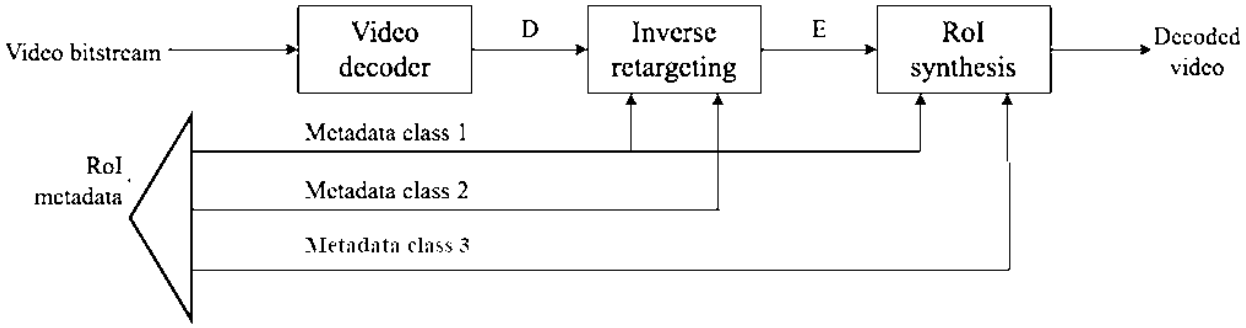

Fig. 2. The structure of the decoder.

Description of the building blocks in a VCM encoder:

a) ***RoI determination and classification***: In this block the original image/video is analyzed and potential regions of interest (RoI) are determined in the areas where interesting objects are located. This is done using a special deep neural network, specifically: detectron network. Then the RoIs are classified as corresponding to important objects, as corresponding to small objects and maybe into other classes.

b) ***Simplification***: The simplification is aimed at production of the content that may be represented with reduced number of bits after compression without any significant deterioration of the efficiency of the machine vision tasks executed on decoded (decompressed) image/video. In principle the simplification reduces details in the background whereas the objects and their neighborhoods remain untouched or only slightly

simplified. In particular, the results of RoI determination and classification are used to identifying objects, the bands around objects and the background. Here, we do consider the objects already replaced by simple content by the inpainting. The background is simplified, and possibly some bands around selected objects or even some objects (usually the objects of big size – relatively in an image or in a video frame). The goal of simplification is to produce such image/video that can be easily represented by a very small number of bits (in compressed representation) whereas the object still can be well recognized (detected, tracked etc.). This may be optimized in context of particular class of object within given RoI.

c) ***Retargeting***: The number of lines and columns of the input image or the video frame may be reduced according to the content identified in the process of RoI determination and classification. The lines and columns of samples are mostly removed when their do not comprise samples of objects. The lines and rows that comprise object samples may be reduced according to analysis of the objects. This reduction correspond to downsampling the object that must not deteriorate the ability to recognize the objects, therefore it is possible if the object is oversampled with respect of its texture and other features. The sample (line / column) removal must be preceded by the respective local lowpass (anti-aliasing) filtering. In classical retergatting methods energy function has to be formulated which controls which rows and columns are primarily removed. Such would require signalization of the energy function (image) to the decoder. In our proposal a simple yet efficient retargeting technique is used which omits implicit energy image formulation, and instead rectangle-shaped RoI are used. Therefore, only coordinates of RoIs are transmitted.

d) ***Object modeling using a dictionary*** :

Some of the objects detected (of size lesser that 200x200 pixels) are classified as dictionary objects. Those objects are inpainted and not sent through inner codec (VVC). Instead, those objects are signalled as indices pointing to patches from the dictionary (known to the decoder). In the decoder those objects are inserted to the decoded image.

3 Implementation details

The tool has been implemented as ROI component in VCM-RS [1] software.

- encoder_poznanroi.py – main encoder component file
- decoder_poznanroi.py – main decoder component file
- poznanroi_config_codec.py – hard-coded parameters of the codec e.g. bluring parameters.
- poznanroi_config_classes.py – classes of objects used in the encoder
- poznanroi_detectron.py – interface to detectron network
- poznanroi_blur.py – implementation of simplification (blurring) algorithm
- poznanroi_retarget.py – implementation of retargeting algorithm
- poznanroi_utils.py – utilities related to general processing.
- poznanroi_rois.py - utilities related to RoIs
- poznanroi_objs.py - utilities related to objects
- poznanroi_bitenc.py – bitstream encoding utilities
- poznanroi_bitdec.py – bitstream decoding utilities

The current implementation of the tools has additional requirements:

- Detectron framework
- Tracker library
- Bitstring python library – bitstream encoding of RoIs

It can be noted that the current implementation does NOT require GPU for operation (CPU only). It however employs GPU if available to speed-up processing. We have notices negligible differences in encoder performance when CPU is used vs GPU.

4 Syntax

Table 1. VCM RoI information syntax.

Syntax element	C	Descriptor
VCM_info(payloadSize) {		
bits_size_x	5	u
bits_size_y	5	u
output_image_size_x_minus_1	bits_size_x	u
output_image_size_y_minus_1	bits_size_y	u
If (codec_option_use_roi_retargetting) {		
roi_retargetting_info( )		
}		
if (codec_option_use_dictionary_objects) {		
dictionary_objects_info()		
}		

bits_size_x - number of bits on which output_image_size_x_minus_1 syntax element is signalled.

bits_size_y - number of bits on which output_image_size_y_minus_1 syntax element is signalled.

output_image_size_x_minus_1 - size of the original image to be outputted from the VCM decoder, minus 1.

output_image_size_y_minus_1 - size of the original image to be outputted from the VCM decoder, minus 1.

codec_option_use_roi_retargetting - flag active in all Poznan University of Technology responses to CfP: A, B and C

codec_option_use_dictionary_objects - flag active only in Poznan University of Technology response C to CfP

Table 2. RoI retargeting information syntax.

Syntax element	C	Descriptor
roi_retargetting_info() {		
hierarchy_power_base	8	u
if (hierarchy_power_base ! = 0) {		
bits_scale	3	u
bits_pos_x	5	u
bits_pos_y	5	u
bits_size_x	5	u
bits_size_y	5	u
bg_scale	bits_scale	u
bg_size_x_minus_1	bits_pos_x	u
bg_size_y_minus_1	bits_pos_y	u
bits_num_rois	5	5
num_rois	bits_num_rois	u
current_scale = 0		
for (i = 0; i < num_rois; i++) {		
if (current_scale!=0) {		
update_scale	1	u
if (update_scale):		
current_scale	bits_scale	u
}		
roi_scale[i] = current_scale		
roi_pos_x[i]	bits_pos_x	u
roi_pos_y[i]	bits_pos_y	u
roi_size_x_minus_1[i]	bits_size_x	u
roi_size_y_minus_1[i]	bits_size_y	u

}		
}		

hierarchy_power_base - base of power function basing on which retargetting scaling is calculated

bits_scale - number of bits on which bg_scale and roi_scale[i] syntax elements are signaled

bits_pos_x - number of bits on roi_pos_x[i] syntax element is signaled

bits_pos_y - number of bits on roi_pos_x[i] syntax element is signaled

bits_size_x - number of bits on roi_size_x_minus_1[i] syntax element is signaled

bits_size_y - number of bits on roi_size_y_minus_1[i] syntax element is signaled

bg_scale - scale level of the background (highest hierarchy RoI element)

bg_size_x_minus_1 - size (minus 1) of the background (highest hierarchy RoI element)

bg_size_y_minus_1 - size (minus 1) of the background (highest hierarchy RoI element)

bits_num_rois - number of bits on which num_rois syntax elements is signaled

num_rois - number of RoIs (excluding background)

current_scale - temporary decoding variale, used to update successive scales: roi_scale[i]

roi_scale[i] - scale of RoI index i

roi_pos_x[i] - position of RoI index i

roi_pos_y[i] - position of RoI index i

roi_size_x_minus_1[i] - width (minus 1) of RoI index i

roi_size_y_minus_1[i] - height(minus 1) of RoI index i

Table 3. Dictionary objects information syntax.

Syntax element	C	Descriptor
dictionary_objects_nfo() {		
bits_num_dictobjs	5	u
if (bits_num_dictobjs!=0) {		
num_dictobjs	bits_num_dictobj	u
if (num_dictobjs!=0) {		
bits_class_id	3	u
bits_patch_id	4	u
bits_start_x_min	5	u
bits_start_y_min	5	u
bits_start_width	5	u
bits_start_height	5	u
is_sequence	1	u
if (is_sequence) {		
bits_start_frame	4	u
bits_num_frames	4	u
bits_v_x_min	5	u
bits_v_y_min	5	u
bits_v_width	5	u
bits_v_height	5	u
}		
for (i = 0; i < num_dictobjs; i++) {		
dictobj_class_id[i]	bits_class_id	u
dictobj_patch_id[i]	bits_patch_id	u
dictobj_flipped[i]	1	
dictobj_start_x_min[i]	bits_start_x_min	u
dictobj_start_y_min[i]	bits_start_y_min	u
dictobj_start_width[i]	bits_start_width	u
dictobj_start_height[i]	bits_start_height	u
if is_sequence:		
dictobj_start_frame[i]	bits_start_frame	u
dictobj_num_frames[i]	bits_num_frames	u
dictobj_v_x_min[i]	bits_v_x_min	s
dictobj_v_y_min[i]	bits_v_y_min	s
dictobj_v_width[i]	bits_v_width	s
dictobj_v_height[i]	bits_v_height	s
}		
}		
}		

bits_num_dictobjs - number of bits on which num_dictobjs syntax element is signaled

num_dictobjs - number of dictionary objects

bits_class_id - number of bits on dictobj_class_id[i] syntax element is signaled

bits_patch_id - number of bits on dictobj_patch_id[i] syntax element is signaled

bits_start_x_min - number of bits on dictobj_start_x_min[i] syntax element is signaled

bits_start_y_min - number of bits on dictobj_start_y_min[i] syntax element is signaled

bits_start_width - number of bits on dictobj_start_width[i] syntax element is signaled

bits_start_height - number of bits on dictobj_start_height[i] syntax element is signaled

is_sequence - flag specyfying whether velocity parameters of dictionary objects are transmitted (or whether they are static across GOP).

bits_start_frame - number of bits ondictobj_start_frame[i] syntax element is signaled

bits_num_frames - number of bits on dictobj_num_frames[i] syntax element is signaled

bits_v_x_min - number of bits on dictobj_v_x_min[i] syntax element is signaled

bits_v_y_min - number of bits on dictobj_v_y_min[i] syntax element is signaled

bits_v_width - number of bits on dictobj_v_width[i] syntax element is signaled

bits_v_height - number of bits on dictobj_v_height[i] syntax element is signaled

dictobj_class_id[i] - index of class from dictionary for object i

dictobj_patch_id[i] - index of texture patch from dictinaary for class dictobj_class_id[i], for object i

dictobj_flipped[i] – value 1 means object is horizontally flipped

dictobj_start_x_min[i] - starting position of object i

dictobj_start_y_min[i] - starting position of object i

dictobj_start_width[i] - starting width object i

dictobj_start_height[i] - starting height object i

dictobj_start_frame[i] - index of first frame in GOP, when object i appears

dictobj_num_frames[i] - number of frames in GOP, in which object i appears

dictobj_v_x_min[i] - velocity (differential change) of starting position of object i

dictobj_v_y_min[i] - velocity (differential change) of starting position of object i

dictobj_v_width[i] - velocity (differential change) of width of object i

dictobj_v_height[i] - velocity (differential change) of height of object i

5 References

[1] Honglei Zhang , “Introduction to the VCM reference software (VCM-RS) “,ISO/IEC JTC 1/SC 29/WG 4 m62003, January 2023, Online

[2] Marek Domański, Olgierd Stankiewicz, Sławomir Maćkowiak, Sławomir Różek, Tomasz Grajek, Jakub Szekiełda, Dominik Cywiński, Jakub Siejak, "[VCM] Poznań University of Technology Proposal A in response to CfP on Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG 2 m60727, Online – October 2022.

[3] Marek Domański, Olgierd Stankiewicz, Sławomir Maćkowiak, Sławomir Różek, Tomasz Grajek, Jakub Szekiełda, Dominik Cywiński, Jakub Siejak, "[VCM] Poznań University of Technology Proposal B in response to CfP on Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG 2 m60728, Online – October 2022.

[4] Marek Domański, Olgierd Stankiewicz, Sławomir Maćkowiak, Sławomir Różek, Tomasz Grajek, Jakub Szekiełda, Dominik Cywiński, Jakub Siejak, "[VCM] Poznań University of Technology Proposal C in response to CfP on Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG 2 m60729, Online – October 2022.

[5] "Call for Proposals for Video Coding for Machines" wg2n00191

[6] ISO/IEC JTC1/SC29/WG5, "VVC Reference Model (VTM)"