INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 04 MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 04 **m64708**

October 2023, Germany, Hannover

Title [MIV] New depth maps for selected CTC sequences

Source PUT, ETRI

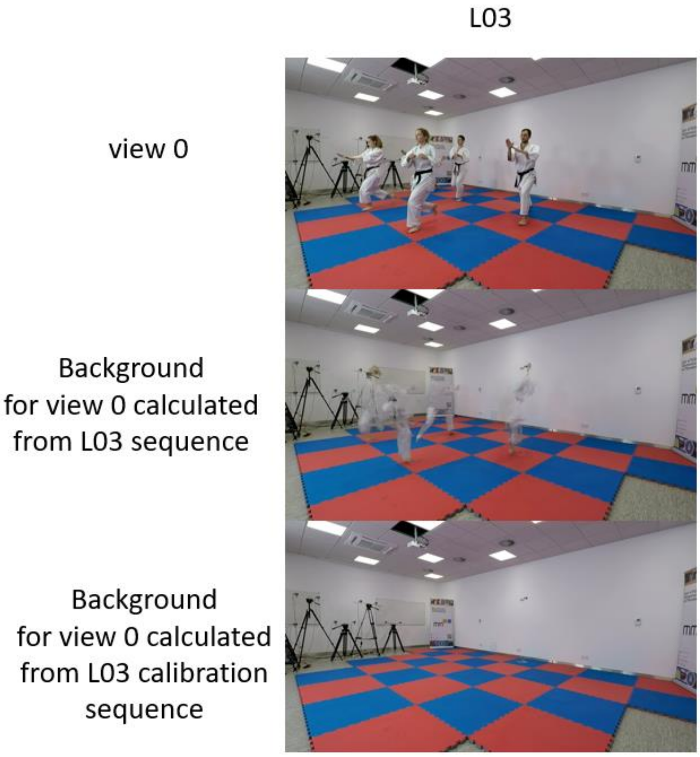
Authors Dominika Klóska, Dawid Mieloch, Jakub Kit, Adrian Śliwiński, Gwangsoon Lee

Abstract

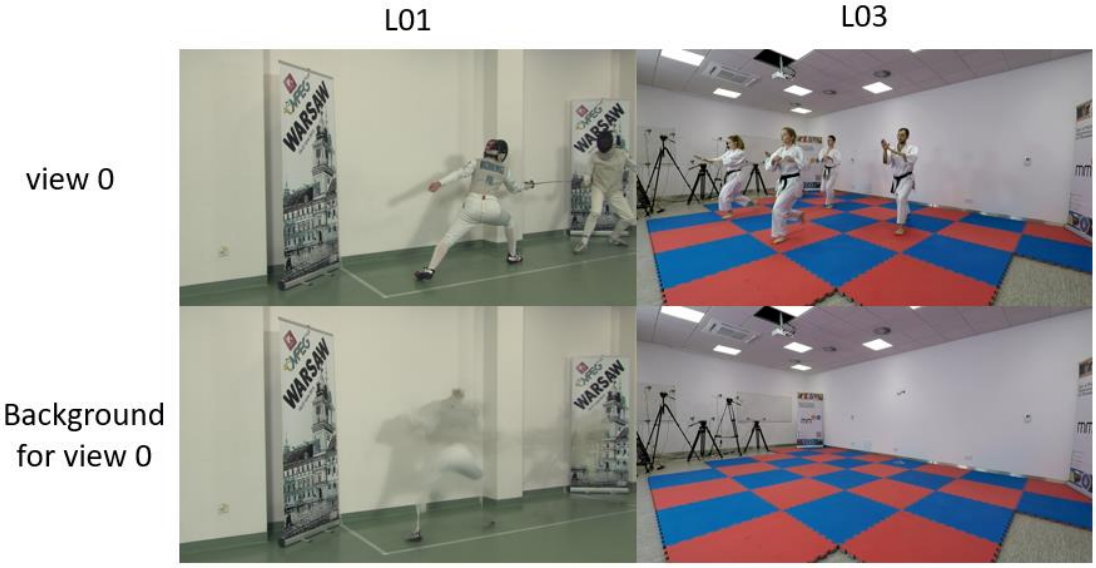
This document presents a proposal of enhancing depth maps by estimating them separately for the background and moving objects and then combining them into final temporally-stable depth map. Proposed approach is dedicated to be used with natural sequences and was used to enhance the quality of depth maps in a set of CTC sequences. The recommendation is to include proposed depth maps in the new CTC.

1 Proposal

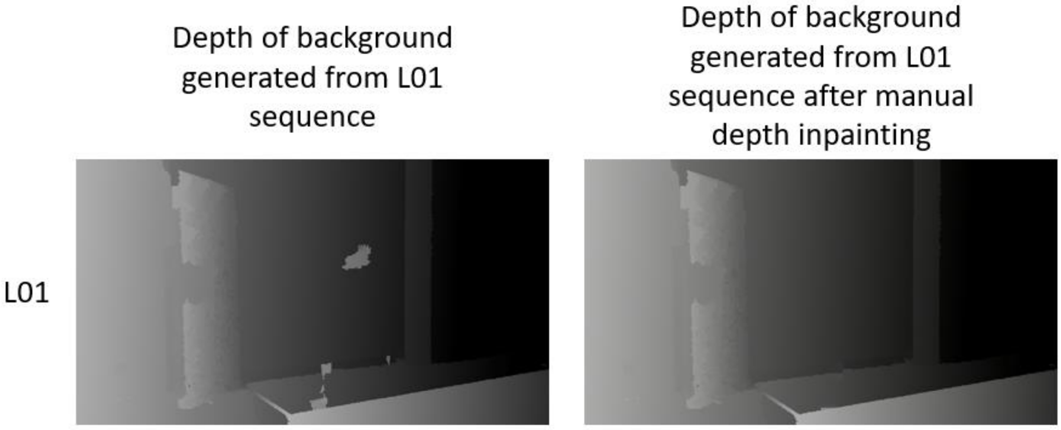
The first step in estimating a temporally-stable depth map for natural content is to create a still background. It is done by computing the median frame over time for the sequence. This process can be done for the actual sequence or the sequence used for calibration. Using the latter yields better results, for there usually is less movement in such sequence compared to the sequence created after the calibration. Such background is generated individually for each view of the sequence. After this process, estimated backgrounds are used for depth estimation.


L03
view 0
Background for view 0 calculated from L03 sequence
Background for view 0 calculated from L03 calibration sequence

Depth estimated from still background generated from the actual sequence may contain some artifacts in places where objects moved rapidly through time.

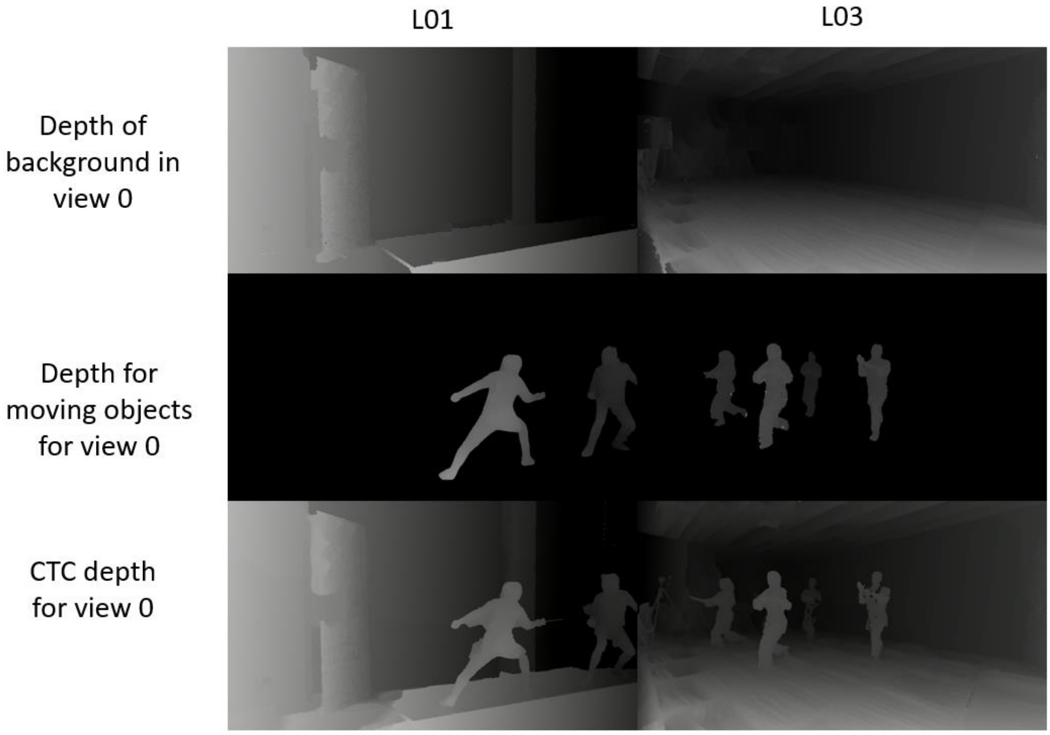


Because the background contains only one frame, it is possible to easily remove these artifacts manually, for instance, using depth inpainting software.


Depth of background generated from L01 sequence
Depth of background generated from L01 sequence after manual depth inpainting
L01

The next step is to detect and then cut out every moving object from the sequence. For that purpose, we used the Detectron2 library. As Detectron2 works for each frame independently, our algorithm has to merge similar objects from all frames into one list. After corresponding objects are identified (using their masks), their colors in neighboring frames are compared – if the average difference between frames for this object is high, then it is recognized as a moving one. This process generates a sequence containing only the moving objects and no background.

The depth for moving objects is estimated using IVDE 8.0, as it recognizes black areas as ones where depth should not be estimated.


L01
L03
Depth of background in view 0
Depth for moving objects for view 0
CTC depth for view 0

Last step is to combine depth map of the background with the depth map of moving objects – each views background with each views moving objects respectively.

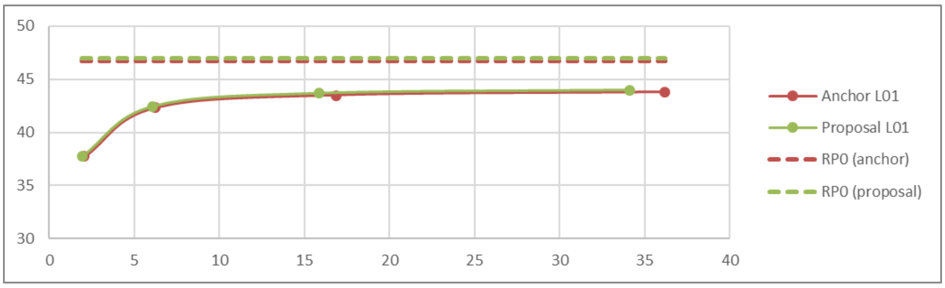
2 Results

L01 (Fencing):

L01								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L01	RP1	27.397	8.756	0.019	36.172	76%	24%	0%
L01	RP2	10.775	6.049	0.019	16.843	64%	36%	0%
L01	RP3	2.975	3.232	0.019	6.226	48%	52%	0%
L01	RP4	0.615	1.414	0.019	2.048	30%	69%	1%
L01	RP0							

L01								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L01	RP1	27.189	6.901	0.020	34.110	80%	20%	0%
L01	RP2	10.677	5.158	0.020	15.855	67%	33%	0%
L01	RP3	2.955	3.108	0.020	6.082	49%	51%	0%
L01	RP4	0.610	1.291	0.020	1.921	32%	67%	1%
L01	RP0							

	L01			
%	IV-PSNR			
	Anchor L01	Proposal L01	Delta	BD-rate
-5.7%	43.79	44.00	0.21	-10.4%
-5.9%	43.49	43.70	0.20	
-2.3%	42.31	42.47	0.15	BD-PSNR
-6.2%	37.72	37.75	0.03	0.5%
	46.76	46.96	0.20	

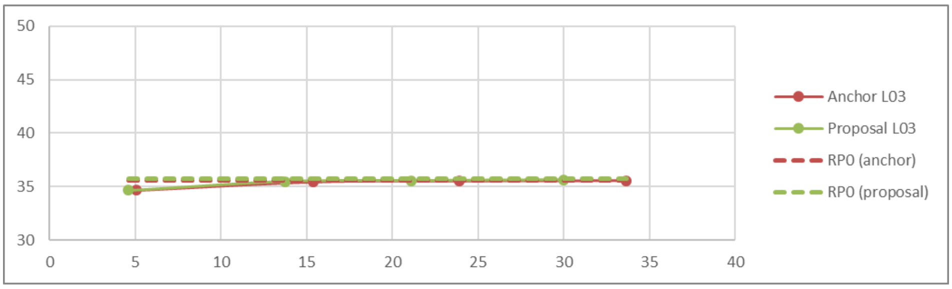


L03 (MartialArts):

L03								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L03	RP1	22.812	10.800	0.027	33.639	68%	32%	0%
L03	RP2	15.006	8.884	0.027	23.917	63%	37%	0%
L03	RP3	9.271	6.097	0.027	15.395	60%	40%	0%
L03	RP4	2.337	2.686	0.027	5.050	46%	53%	1%
L03	RP0							

L03								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L03	RP1	22.791	7.153	0.028	29.971	76%	24%	0%
L03	RP2	15.016	6.066	0.028	21.110	71%	29%	0%
L03	RP3	9.274	4.434	0.028	13.736	68%	32%	0%
L03	RP4	2.345	2.213	0.028	4.586	51%	48%	1%
L03	RP0							

	L03			
%	IV-PSNR			
	Anchor L03	Proposal L03	Delta	BD-rate
-10.9%	35.57	35.65	0.08	-15.6%
-11.7%	35.52	35.57	0.05	
-10.8%	35.43	35.48	0.05	BD-PSNR
-9.2%	34.67	34.65	-0.02	0.2%
	35.59	35.71	0.12	

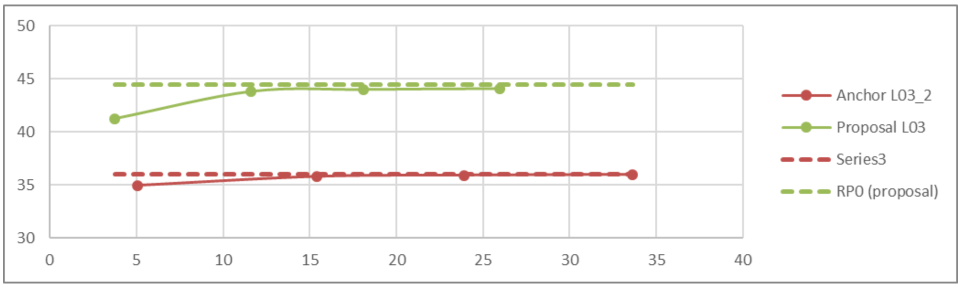


Posetraces for Fencing show improvement over CTC. Unfortunately, for MartialArts the depth estimated for left- and rightmost views (also in CTC) is too low to achieve high quality. When only 9 views are used, then the quality of posetraces noticeably improves, but this is done at the expense of available viewing space. Here are the results of coding for decreased number of views:

L03_2								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L03_2	RP1	22,812	10,800	0,027	33,639	68%	32%	0%
L03_2	RP2	15,006	8,884	0,027	23,917	63%	37%	0%
L03_2	RP3	9,271	6,097	0,027	15,395	60%	40%	0%
L03_2	RP4	2,337	2,686	0,027	5,050	46%	53%	1%
L03_2	RP0							

L03_2								
Test		Bitrate [Mbps]				Fraction [%]		
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*
L03_2	RP1	20,539	5,427	0,011	25,977	79%	21%	0%
L03_2	RP2	13,514	4,558	0,011	18,083	75%	25%	0%
L03_2	RP3	8,344	3,272	0,011	11,628	72%	28%	0%
L03_2	RP4	2,150	1,590	0,011	3,752	57%	42%	0%
L03_2	RP0							

	L03_2			
%	IV-PSNR			
	Anchor L03	Proposal L03	Delta	BD-rate
-22,8%	36,00	44,08	8,08	0,0%
-24,4%	35,93	43,99	8,06	
-24,5%	35,80	43,82	8,01	BD-PSNR
-25,7%	34,95	41,27	6,32	22,5%
	36,02	44,51	8,49	



3 Recommendation

We recommend including the proposed depth maps in CTC.

4 Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory).