Friedrich Alexander University Erlangen-Nuremberg
Chair of Multimedia Communications and Signal Processing
Head: Prof. Dr.-Ing. André Kaup

Subjective Test 2004
Preliminary Report

Marcus Barkowsky
barkowsky@nt.e-technik.uni-erlangen.de
http://www.lnt.de/LMS
Cauerstrasse 7
91058 Erlangen
GERMANY

Contents

1 Introduction **5**

2 Setup **7**
2.1 Room and Surroundings . . . 7
2.2 Monitor . . . 8
2.3 Playout . . . 8
2.4 User Interface . . . 9
2.5 Test procedure . . . 9
2.6 Candidates . . . 10

3 Sequences **11**
3.1 Overview . . . 11
3.2 Classification . . . 11
3.3 Example pictures . . . 12

4 Hypothetical Reference Circuits **15**
4.1 Reference (HRC0) . . . 15
4.2 H.264 (HRC1-6) . . . 15
4.3 MPEG4 (HRC7-11) . . . 16
4.4 JPEG2000 (HRC12-13) . . . 23
4.5 Transmission distortion (HRC14-18) . . . 23
4.6 Bandwidth limitation (HRC19) . . . 29
4.7 Pictures of typical distortions . . . 29

5 Results **35**
5.1 Results for each SRC and HRC . . . 35

A Measurement **43**
A.1 PSNR . . . 43

B Observer validation **45**
B.1 Results for each Observer . . . 45

C H.264 **79**

D MPEG-4 **91**

E Codec comparison **101**

F Playback with mplayer **107**

Chapter 1

Introduction

The focus of this subjective videotest is on multimedia applications. In 2004 the video quality experts group (VQEG) prepared a test plan for a multimedia test which shall be conducted in 2005. The setup was chosen to conform to 'Draft version 1.4c', dated October 22nd, 2004, of this testplan and to the discussions on the VQEG mailing list and the VQEG multimedia forum. Additionally the well-known ITU standards (e.g. BT.500 and P.910) for subjective assessments were taken into account.
The subjective test was done at the Friedrich-Alexander University Erlangen-Nuremberg in the multimedia room of the Chair of Multimedia Communications and Signal Processing. It was prepared and supervised by the author of this document, Marcus Barkowsky (barkowsky@nt.e-technik.uni-erlangen.de). The selection of the SRCs and HRCs was discussed with Christian Schmidmer and Roland Bitto from OPTICOM.

Chapter 2

Setup

2.1 Room and Surroundings

The setup was chosen to provide a standard compatible, yet comfortable situation. The diagram in figure 2.1 shows the general setup of the viewing room and a zoom including the relevant distances.

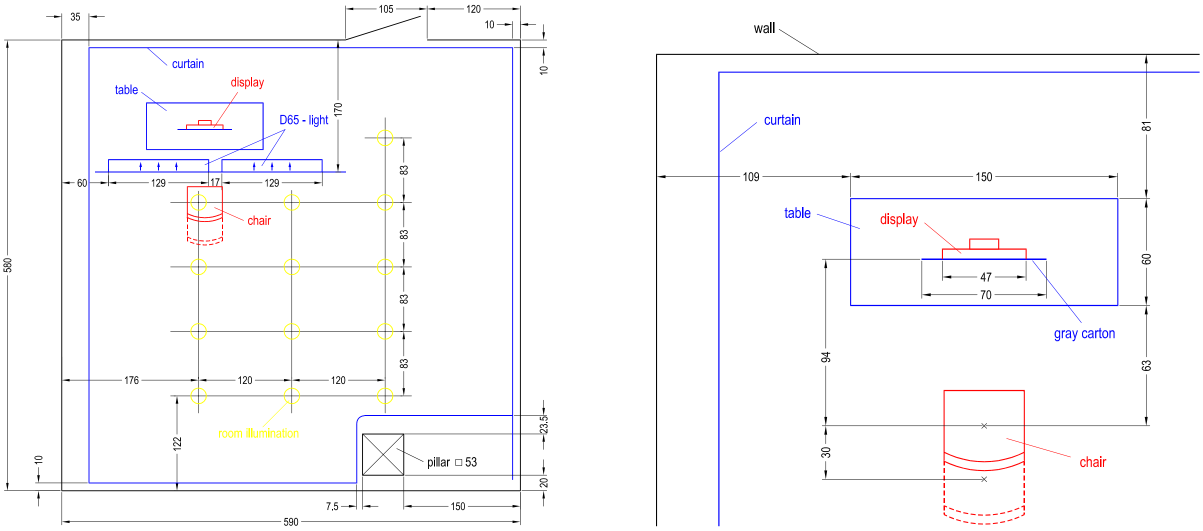


Figure 2.1: Setup of the viewing room and the observer's location

The background was a gray curtain which was illuminated by the D65 light. The following measures were done using a Barko TMF6 monitor measurement device. Unfortunately the device was last calibrated on 24/10/94.

Date	Angle from view center	Lux	cd/m^2	color temperature
Monday 11/29	0-45^o	3.2	3.3	
	90^o	2	2.6	
	135^o	0.28	0.36	
Tuesday 11/30	0-45^o	3.5	3.5	
	90^o	2.6	2.7	
	135^o	0.28	0.25	
Wednesday 12/1	0-45^o	3.2	3.6	
	90^o			
	135^o	0.34	0.4	
Thursday 12/2	0-45^o	3.4	3.35	6100
	90^o	2.5	3.0	6100
	135^o	0.33	0.3	5100
Friday 12/3	0-45^o	3.5	3.5	6000
	90^o	2.7	2.6	5700
	135^o	0.27	0.43	5100

The sound pressure level was measured using a 'Brüel & Kjær 2237 Controller' every morning. At the position of the observer it was always less than the minimum measurable level of 32dB.

 Draft 2005-04-21 Draft 2005-04-21

2.2 Monitor

The display device which was used is an EIZO CG21. This LCD was calibrated for correct brightness, color and gamma every morning. At the front of the monitor a gray paperboard (70x43.5cm) was attached to provide a gray surrounding, because the LCD was surrounded by a black cover. The paperboard contained a cut-out of 22 by 18 cm which was slightly larger than the active image area which was 19.1 by 15.6 cm. The illumination of the gray paper was also measured using the Barko device:

Weekday	Lux	cd/m^2	color temperature
Monday		1.8	5500
Tuesday		2.09	5500
Wednesday	1.8	1.85	5500
Thursday	1.7	1.7	5500
Friday	2.2	2.05	5400

The monitor was calibrated in the morning after a warm-up of half an hour. The following table shows the results of the measurement before and after the calibration using an Eye-One Device from Gretag-Macbeth, which was bought in August 2004:

Date	Measurement	Red	Green	Blue	White	Brightness
Monday	1st before calibration	(0.642,0.329)	(0.295,0.609)	(0.140,0.088)	(0.314,0.330)	125.4
11/29	2nd before calibration	(0.643,0.328)	(0.296,0.607)	(0.140,0.088)	(0.314,0.329)	124.9
	3rd before calibration	(0.644,0.327)	(0.296,0.607)	(0.140,0.087)	(0.314,0.328)	124.3
	after calibration	(0.642,0.329)	(0.296,0.607)	(0.140,0.088)	(0.313,0.330)	120.7
Tuesday	1st before calibration	(0.641,0.329)	(0.295,0.609)	(0.140,0.089)	(0.313,0.331)	125.3
11/30	2nd before calibration	(0.641,0.330)	(0.295,0.609)	(0.140,0.089)	(0.313,0.331)	125.1
	3rd before calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.088)	(0.314,0.330)	124.6
	after calibration	(0.641,0.329)	(0.295,0.609)	(0.140,0.089)	(0.313,0.329)	119.0
Wednesday	1st before calibration	(0.641,0.329)	(0.295,0.608)	(0.140,0.089)	(0.314,0.330)	121.4
12/1	2nd before calibration	(0.641,0.330)	(0.294,0.610)	(0.140,0.089)	(0.313,0.331)	121.3
	3rd before calibration	(0.642,0.329)	(0.295,0.608)	(0.140,0.088)	(0.313,0.329)	120.8
	after calibration	(0.641,0.329)	(0.296,0.608)	(0.140,0.089)	(0.313,0.329)	119.3
Thursday	1st before calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.088)	(0.314,0.329)	120.9
12/2	2nd before calibration	(0.642,0.328)	(0.296,0.608)	(0.140,0.088)	(0.313,0.329)	120.7
	3rd before calibration	(0.642,0.329)	(0.295,0.608)	(0.140,0.088)	(0.313,0.329)	120.4
	after calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.088)	(0.313,0.329)	120.6
Friday	1st before calibration	(0.641,0.330)	(0.295,0.609)	(0.140,0.089)	(0.313,0.330)	123.7
12/3	2nd before calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.089)	(0.314,0.329)	123.8
	3rd before calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.089)	(0.314,0.330)	123.8
	after calibration	(0.642,0.329)	(0.296,0.608)	(0.140,0.089)	(0.314,0.329)	119.9

The values which were used for calibration are specified in the VQEG multimedia test plan. The maximum brightness was chosen as the maximum supported value of 120cd/m^2. The whitepoint was 6500K, corresponding to x=0.313, y=0.329. The gamma factor was 2.2.
After the monitor calibration the gray image was measured using the Barko device. The monitor was displaying a gray background which was also used during the tests between the display of two test sequences. The image was generated using a Y-amplitude of 80. The following values were measured:

Weekday	Lux	cd/m^2	color temperature
Monday		8.3	6800
Tuesday		8.15	6850
Wednesday	8.2	8.25	6850
Thursday	8.4	8.3	
Friday	8.2	8.2	6850

2.3 Playout

The monitor was connected via a 5m DVI-cable to a Clipster station produced by DVS-Hannover. The Clipster station was playing back at 60Hz with a resolution of 1024x768 pixels (non-interpolated). Therefore the image content might have changed each 1/60s≈16.6ms. All processed video sequences (PVS) were concatenated. The Clipster device was controlled by a remote computer using the built-in telnet server. The remote computer sent the in- and outpoints and the playback command to the Clipster device. It was located in a neighbouring room. Due to the limited length of the DVI-cable (which is the longest available length conforming to the DVI specification) the Clipster device had to remain in the viewing room.

2.4 User Interface

The candidates used a custom device to perform the voting. It contained 5 buttons. Next to each button a string from the quality rating scale of the ACR test is printed. The english terms of the ACR test were substituted by their German correspondence:

English	German
Excellent	Sehr gut
Good	Gut
Fair	Befriedigend
Poor	Ausreichend
Bad	Mangelhaft

2.5 Test procedure

At the start of each test session, the candidate's name, gender and age was recorded. Then a random permutation of the 240 test sequences was generated using matlab. The permutation process was repeated if the difference of the PSNR values of two consecutive sequences exceeded 13. About 1000 permutations were necessary to meet this constraint. The training session contained 12 sequences which were manually selected to represent the complete quality range in the test.
Next the candidate was screened for normal visual acuity. There were 16 rows of 4 Landolt rings each on a white paper attached to a wall 4m from the candidates position. Both eyes were tested seperately. The candidate had to determine the position of the gap in the rings for one (randomly chosen) row. One example row is shown below. The outer diameter of the rings should be 6mm.

The color blindness was tested using Ishihara color plates under D65 light.
Then the following introduction was given: (Translated from German version)
For this video quality test we use 12 different video sequences. These sequences are distorted in 20 different ways. This results in a total of 240 sequences. The distortions are typical for multimedia applications. There are coding artifacts as well as transmission errors. Each sequence has a duration of 9 seconds. You will see each of these sequences exactly once. You can vary the viewing distance within the limits given by the white tape below the chair. The next step will be a training sequence. You will see a selection of the distorted sequences. They are representative of the sequences which you will see later. For each of these sequences you will have to give a quality rate later in the test. The voting will be done with this small device. You simply press the button which represents your opinion on the quality of the sequence. You can make your choice from "excellent", "good", "fair", "poor" and "bad". After you have pressed a button, the next sequence will be shown immediately. The votes from the trainings sequence will be dropped. The trainings session is finished, when the screen is showing "5 minutes break". Then we will start the test itself. Another hint concerning the voting: You are asked to give your opinion, that means you should neither classify the sequences in 5 classes, nor change your mind because you have seen the same sequence in a better or worse quality before.
Any questions?
(After the training session):
Now the video quality test starts. The procedure is identical to the training. You will see 60 sequences corresponding to approximately 10 minutes of video sequences. After that you are asked to have a break of 5 minutes. Please come to the neighbouring room then. There are four sessions of 10 minutes.

German version:
Für den Videobetrachtungstest verwenden wir 12 unterschiedliche Videosequenzen. Diese wurden auf 20 unterschiedliche Arten mit Störungen versehen. Daraus ergibt sich eine Gesamtzahl von 240 Sequenzen. Die Störungen sind typisch für Multimediaanwendungen und umfassen sowohl Codierartefakte, wie auch Übertragungsstörungen. Jede Sequenz ist etwa 9 Sekunden lang. Du wirst in dem Videobetrachtungstest jede dieser Sequenzen genau einmal sehen. Den Betrachtungsabstand kannst Du im Rahmen dieses weißen Streifens am Boden variieren.
Der nächste Abschnitt besteht aus einer Trainingssequenz. Du wirst eine Auswahl von gestörten Videosequenzen sehen, wie sie später im eigentlichen Test vorkommen werden. Für jede dieser Sequenzen sollst Du später eine aussagekräftige Bewertung abgeben. Die Bewertung erfolgt mit Hilfe dieses kleinen Kästchens, auf welchem Du nach

jeder Videosequenz bitte diejenige Taste drückst, die Deiner Meinung nach der Qualität der Videosequenz entspricht. Es gibt die Abstufungen „sehr gut“, „gut“, „befriedigend“, „ausreichend“ und „mangelhaft“. Nachdem Du eine Taste gedrückt hast, folgt sofort die nächste Videosequenz. Die Bewertungen während der Trainingssequenz werden verworfen. Wenn der Bildschirm „5 Minuten Pause“ anzeigt, ist der Trainingsabschnitt beendet und wir kommen zum eigentlichen Videobetrachtungstest. Noch eine Anmerkung zur Bewertung: Du sollst die Bewertung entsprechend Deiner eigenen Meinung abgeben, d.h. weder die Sequenzen in 5 Kategorien einteilen, noch Dich davon beeinflussen lassen, dass Du eine bestimmte Sequenz vorher schon einmal in schlechterer oder besserer Qualität gesehen hast.
Gibt es noch Fragen?
(Nach der Trainingssequenz):
Nun folgt der eigentliche Test. Der Ablauf ist identisch mit der Trainingssequenz. Du wirst nun viermal hintereinander ca. 10 Minuten lang Videosequenzen sehen. Nach jeweils 60 Sequenzen, entsprechend etwa 10 Minuten, wirst Du aufgefordert, Dich in einer fünfminütigen Pause zu erholen. Sage mir dann bitte kurz im Nachbarraum Bescheid.

The viewing distance could be chosen within the limits of 6-8 times the height of the image as specified in the VQEG multimedia test plan 1.4c.

2.6 Candidates

Thirty students participated in the subjective test. They were paid by OPTICOM. Six of them were women. The distribution of the age of all participants can be seen from the following table:

Number of participants	3	2	5	4	8	4	3	1
Age	20	21	22	23	24	25	26	27

Chapter 3

Sequences

3.1 Overview

For the test 12 sequences were chosen. Each sequence contains 220 frames. A frame rate of 25fps was assumed, which was the recording framerate for all sequences but sequence number 10.
There are four sources for the source sequences:

1. VQEG Phase I: The sequences 9-12 are from the Video Quality Experts Group Test Phase I. Sequences 9 is the well-known "Mobile and Calendar", which was SRC10 in the Phase I testset. Sequence 10 contains the "football" play which was SRC19. The SRC6 was used as sequence number 11 and contains a car "racing". The "fastfood" scene from SRC7 was used as sequence 12.

2. Virtual Highschool of Bavaria (VHB): Sequence number 5 shows Prof. Huber, head of the Chair of Information Transmission in Erlangen (Germany) in his lecture on information theory. He is explaining something to his students in front of the blackboard. This sequence was recorded using a consumer DV-Cam.

3. Discovery Channel Advertisement: The sequences 2, 3 and 4 contain different scenes from a digital betacam tape which was sent to Opticom and digitized at the Fraunhofer Institute for Integrated Circuits. Sequence number 2 shows an animation of the discovery channel logo, while sequences 3 and 4 are examples for high motion video content with lots of scene cuts.

4. ITU/IEC HDTV-Sequences: These sequences are donated by ABC, NASA and DemoGrafx for work in standardization related research. They were digitized at the RWTH-Aachen (Germany) and hosted at the university of hannover. The sequence 1 shows a flight over New York. People walking over a street at night is seen on sequence 6. A sherif's boat on its patrol can be seen in number 7 and the take-off of a space shuttle is the content of sequence 8.

The downsampling for the first three sources was done using the MPEG4-Filter which was proposed by Andreas Hutter. Some sequences were in high definition format and therefore the MPEG4 filter could not be applied. A spline filter of degree 3 was used instead.

3.2 Classification

The following table shows the characteristics of the different sequences. The TI and SI measure is described in ITU-T P.910.

SRC	Sequence	TI	SI	Characteristics
1	City	77	24	High detail, slow pan
2	Discovery animation	123	37	Synthetic, thin structures, medium motion, text
3	Discovery city	125	84	Many scene cuts, high detail, slow motion
4	Discovery orient	114	73	Many scene cuts, many faces, colorful
5	Huber1	126	26	High color contrast, critical finger movement, text on background
6	Night	108	26	Smoothly moving crowd, difficult faces, occlusion by car
7	Sheriff	108	23	Water movement, High detailed ships, slow pan
8	Shuttle	48	9	Noisy cloud, foggy background, vibrating motion
9	Mobile and Calendar	147	30	color, high detail, object and background motion
10	Football	124	40	High color contrast, fast blurred motion
11	Racing	145	46	Saturated color on gray background, color contrast text, noisy street
12	Fastfood	125	65	Noisy start, faces, fast panning and zoom

3.3 Example pictures

Sequence SRC 1

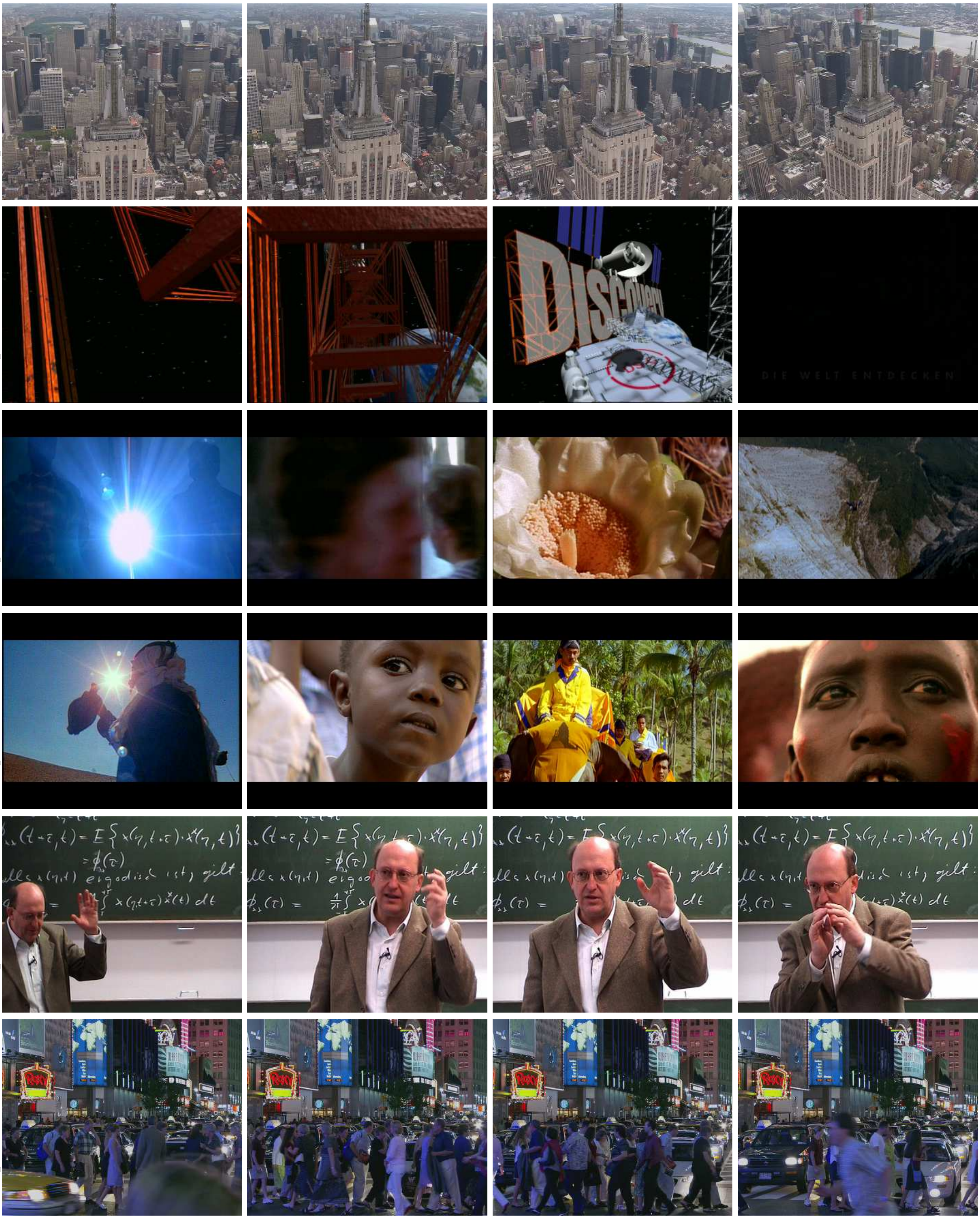
Sequence SRC 2

Sequence SRC 3

Sequence SRC 4

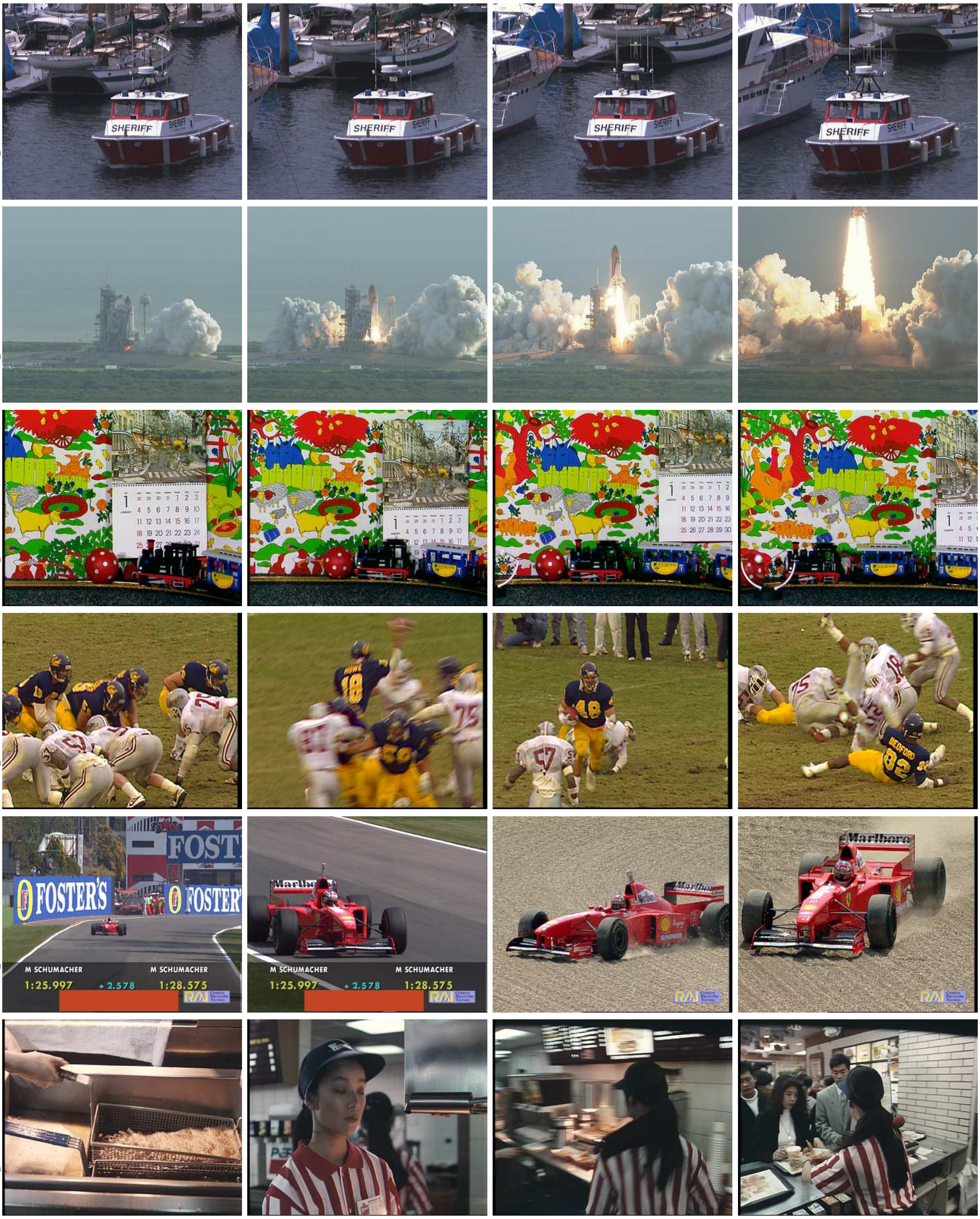
Sequence SRC 5

Sequence SRC 6



Frame 0 Frame 73 Frame 146 Frame 219

 Draft 2005-04-21 Draft 2005-04-21



Sequence SRC 7
Sequence SRC 8
Sequence SRC 9
Sequence SRC 10
Sequence SRC 11
Sequence SRC 12

Frame 0 | Frame 73 | Frame 146 | Frame 219

 Draft 2005-04-21 Draft 2005-04-21

Chapter 4

Hypothetical Reference Circuits

This chapter describes the generation of the Hypothetical Reference Circuits in detail.
The following table gives a quick overview of the HRCs used:

HRC	Codec	fps	Bitrate	Comments or Distortion
0		25		Reference
1	H.264	25	704	
2	H.264	25	320	
3	H.264	25	128	
4	H.264	12.5	128	
5	H.264	12.5	64	
6	H.264	8	64	
7	MPEG-4	25	704	
8	MPEG-4	12.5	704	
9	MPEG-4	12.5	320	
10	MPEG-4	8	320	
11	MPEG-4	8	192	
12	JPEG-2000	12.5	704	
13	JPEG-2000	25	704	
14	MPEG-4	12.5	320	Packetloss low, concealment
15	MPEG-4	12.5	320	Packetloss only once, concealment
16	MPEG-4	12.5	320	Packetloss twice, skipping concealed frames
17	MPEG-4	12.5	320	Packetloss twice, concealment
18	MPEG-4	12.5	320	Biterrors
19	H.264			Variable framerate and datarate

4.1 Reference (HRC0)

The HRC0 contains the reference sequences. Most of the sequences were generated from PAL (720x576 pixel) formats using the MPEG4-Filter from Andreas Hutter. This applies for the SRCs Nr. 2-5 and 9-12. The other sequences were reduced from HDTV format 1280x720 by first cropping the image with the center area to 880x720 pixel and then using a spline scale algorithm with splines of degree 3 to reduce the image by a factor of 0.4. No other filters were applied to the source material.

4.2 H.264 (HRC1-6)

The H.264 reference codec JM9.1 was used for the coding of the sequences. This version was published 19th of November 2004. The HRCs differ in the parameters bitrate and framerate.
For the coding no rate control was used but the overall quantisation parameter (QP) was adjusted to meet the bitrate limitation. This was done to avoid distortions due to changing quantizer step size across a picture or over the sequence.
For the temporal subsampled codings at frame rate 12.5 and 8.33 no temporal filtering was applied. At 12.5fps every other frame was eliminated from the sequence before, so the H.264 codec had to code each frame. Similar processing was applied to generate the 8.33fps coded sequences.
The decoded video was assumed to be displayed by an ideal player, which outputs every frame at exactly the correct temporal timestamp. Therefore the sequences are perfectly alligned to HRC0, except for the frame repetition and skipping due to a reduced framerate.

The table **??** displays the desired and reached bitrate and the PSNR as a measure for the distortion. In this PSNR measure only the reference frames which have a corresponding distorted frame are compared. For details see Appendix C:

		HRC1	HRC2	HRC3	HRC4	HRC5	HRC6
Max. Bitrate [kbit/s]		704	320	128	128	64	64
Frames per second [Hz]		25	25	25	12.5	12.5	8
SRC1	QP chosen	29	33	37	35	39	38
SRC1	Bitrate used	670	290	119	122	52	52
SRC1	PSNR reached	31.8	28.7	26.1	27.4	25.1	25.6
SRC2	QP chosen	29	35	41	38	44	42
SRC2	Bitrate used	659	298	127	125	55	56
SRC2	PSNR reached	36.6	32.4	28.5	30.5	26.7	28.0
SRC3	QP chosen	26	32	40	36	42	40
SRC3	Bitrate used	615	298	112	123	59	58
SRC3	PSNR reached	39.5	35.6	30.6	33.2	29.4	30.7
SRC4	QP chosen	24	30	36	34	39	38
SRC4	Bitrate used	703	287	119	117	59	56
SRC4	PSNR reached	39.5	35.5	31.6	32.9	29.8	30.3
SRC5	QP chosen	24	29	35	33	38	37
SRC5	Bitrate used	609	280	125	114	60	57
SRC5	PSNR reached	38.4	35.2	31.2	32.7	29.2	30.0
SRC6	QP chosen	30	36	41	39	43	42
SRC6	Bitrate used	703	275	116	117	60	55
SRC6	PSNR reached	32.3	27.9	24.8	26.1	23.9	24.4
SRC7	QP chosen	29	32	36	34	38	36
SRC7	Bitrate used	605	312	121	127	51	61
SRC7	PSNR reached	33.2	31.0	28.5	29.9	27.6	28.7
SRC8	QP chosen	22	26	32	30	34	33
SRC8	Bitrate used	575	310	109	115	58	57
SRC8	PSNR reached	41.0	38.3	34.6	35.9	33.7	34.3
SRC9	QP chosen	32	36	41	39	43	42
SRC9	Bitrate used	599	292	121	118	61	57
SRC9	PSNR reached	29.6	26.4	22.9	24.5	22.0	22.7
SRC10	QP chosen	34	39	46	43	47	46
SRC10	Bitrate used	627	317	115	121	68	59
SRC10	PSNR reached	31.6	28.1	24.4	26.1	24.0	24.5
SRC11	QP chosen	34	37	41	39	42	41
SRC11	Bitrate used	592	317	105	120	59	57
SRC11	PSNR reached	29.2	27.1	24.9	25.9	24.5	24.9
SRC12	QP chosen	29	35	42	40	45	44
SRC12	Bitrate used	675	310	119	114	60	55
SRC12	PSNR reached	35.4	31.4	27.0	28.3	25.2	25.8

4.3 MPEG4 (HRC7-11)

The MPEG-4 coding was performed using the Momusys MPEG4 reference software. The same method was applied as described for the H.264 coded sequences: The QP was chosen to match a specified bandwidth limit.
All MPEG-4 coded sequences were played back using the mplayer. The player was called with the correct number of frames per second given and it was forced to use SDL output. This SDL output was captured in direct conjunction with the timestamp, when the output appeared. After the player had finished, a 60fps sequence was generated using the captured frames and their corresponding timestamps. Using this method, every candidate watched exactly the same video frames as if he would have seen in that run of the mplayer command. Details can be found in annex F.
The following table shows the quantisation parameters used for the different sequences in order to get the correct bitrate:

		HRC7	HRC8	HRC9	HRC10	HRC11
Max. Bitrate [kbit/s]		704	704	320	320	192
Frames per second [Hz]		25	12.5	12.5	8	8
SRC1	QP chosen	10	7	12	10	15
SRC1	Bitrate used	649	702	318	317	185
SRC1	PSNR reached	28.9	31.4	28.0	29.0	27.0
SRC2	QP chosen	10	6	13	10	18
SRC2	Bitrate used	634	675	317	303	182
SRC2	PSNR reached	33.7	34.6	31.4	32.4	29.9
SRC3	QP chosen	8	5	12	9	18
SRC3	Bitrate used	639	640	303	306	182
SRC3	PSNR reached	36.3	37.9	33.5	35.7	32.2
SRC4	QP chosen	6	4	8	7	12
SRC4	Bitrate used	580	662	305	304	177
SRC4	PSNR reached	36.3	38.4	34.9	35.8	32.7
SRC5	QP chosen	6	4	8	6	10
SRC5	Bitrate used	529	617	284	311	187
SRC5	PSNR reached	35.6	37.5	34.1	35.5	32.9
SRC6	QP chosen	11	8	16	13	20
SRC6	Bitrate used	677	700	297	306	182
SRC6	PSNR reached	29.3	31.2	27.1	28.3	25.9
SRC7	QP chosen	8	6	11	9	13
SRC7	Bitrate used	689	667	288	292	176
SRC7	PSNR reached	31.6	33.3	30.2	31.3	29.5
SRC8	QP chosen	4	3	5	4	6
SRC8	Bitrate used	582	609	298	301	178
SRC8	PSNR reached	38.8	41.0	38.0	38.9	36.8
SRC9	QP chosen	12	9	18	15	22
SRC9	Bitrate used	660	694	288	299	188
SRC9	PSNR reached	27.6	29.8	25.4	26.5	24.3
SRC10	QP chosen	18	11	26	18	28
SRC10	Bitrate used	674	656	316	311	223
SRC10	PSNR reached	29.0	31.4	27.3	28.7	26.8
SRC11	QP chosen	14	10	18	15	22
SRC11	Bitrate used	641	684	317	305	187
SRC11	PSNR reached	28.2	30.0	26.8	27.7	25.9
SRC12	QP chosen	10	7	18	13	26
SRC12	Bitrate used	684	677	294	307	183
SRC12	PSNR reached	32.7	34.9	29.7	31.3	27.8

In contrast to the HRCs for H.264 and JPEG-2000 the real-world playback of the sequences using mplayer also introduced the possibility of an offset or even drift between the source sequence and the PVS. Therefore the corresponding original frame for every frame in the processed sequence was searched using the MSE criterion. This results in a delay value per frame which is depicted in figures 4.1-4.5. The problem that at the end of a sequence one frame was repeated is usually not noticable by the user, because the display switches to gray background immediately afterwards. If the frame freeze lasts only for a few frames, then the sudden switch to the gray picture masks the frame freeze and even if it is half a second, the viewer recognizes it as end-of-sequence artifact without further influence on the subjective score. The 8 frames per second HRCs 10 and 11 show a ramp which may be due to slower playback than should have been, e.g. 8fps instead of 8.33fps.

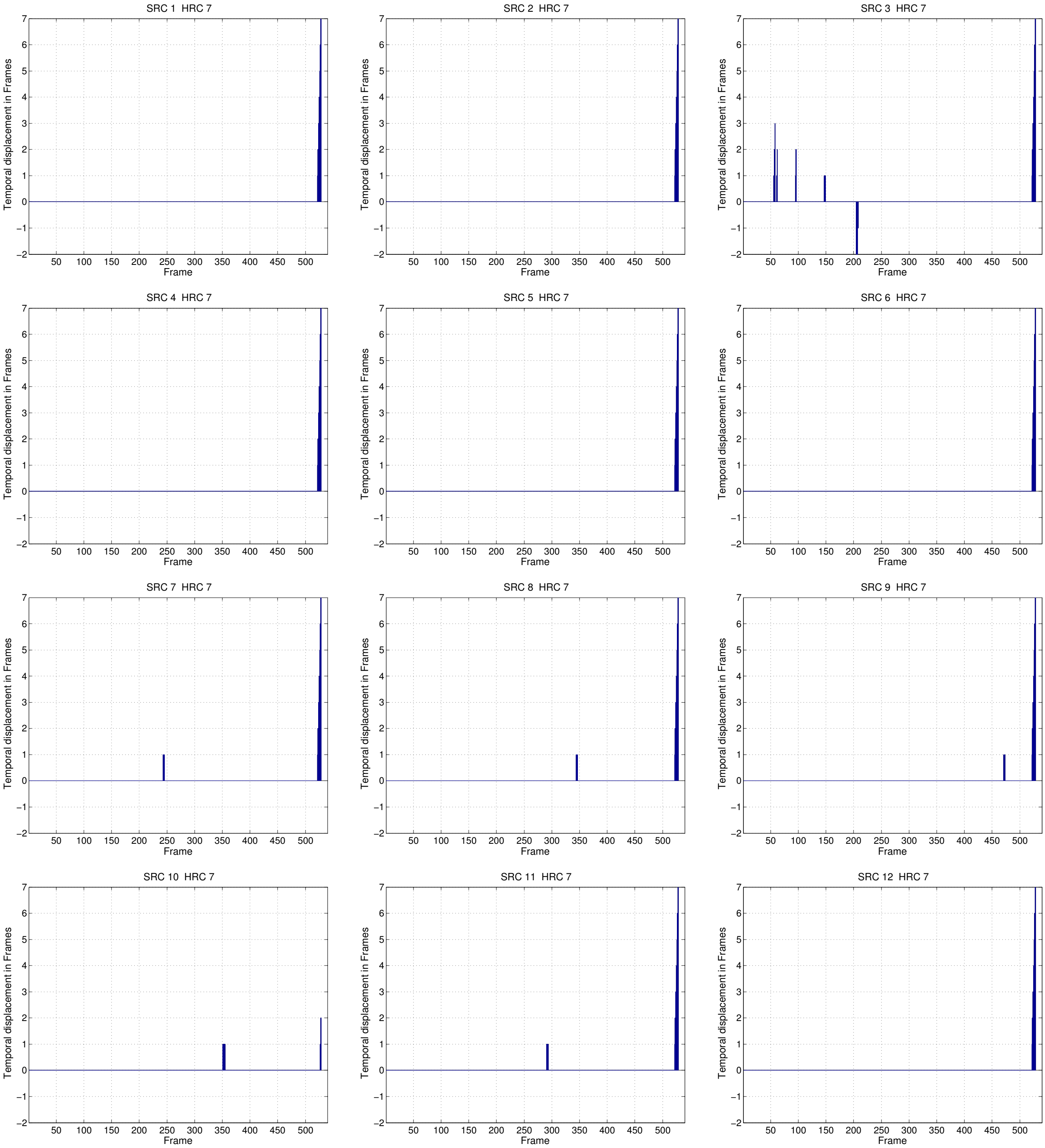


Figure 4.1: Temporal delay between original sequence and HRC7

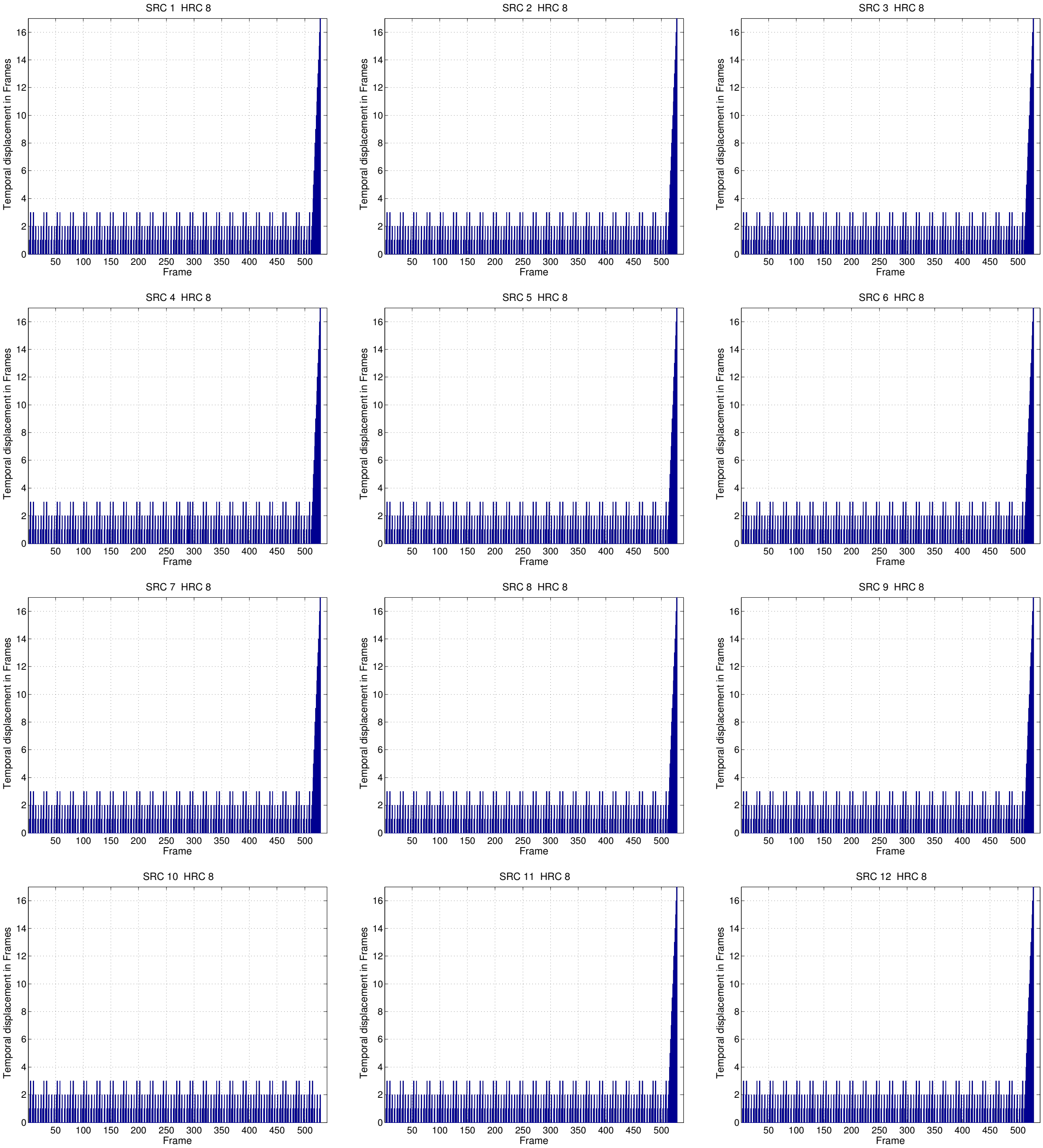


Figure 4.2: Temporal delay between original sequence and HRC8

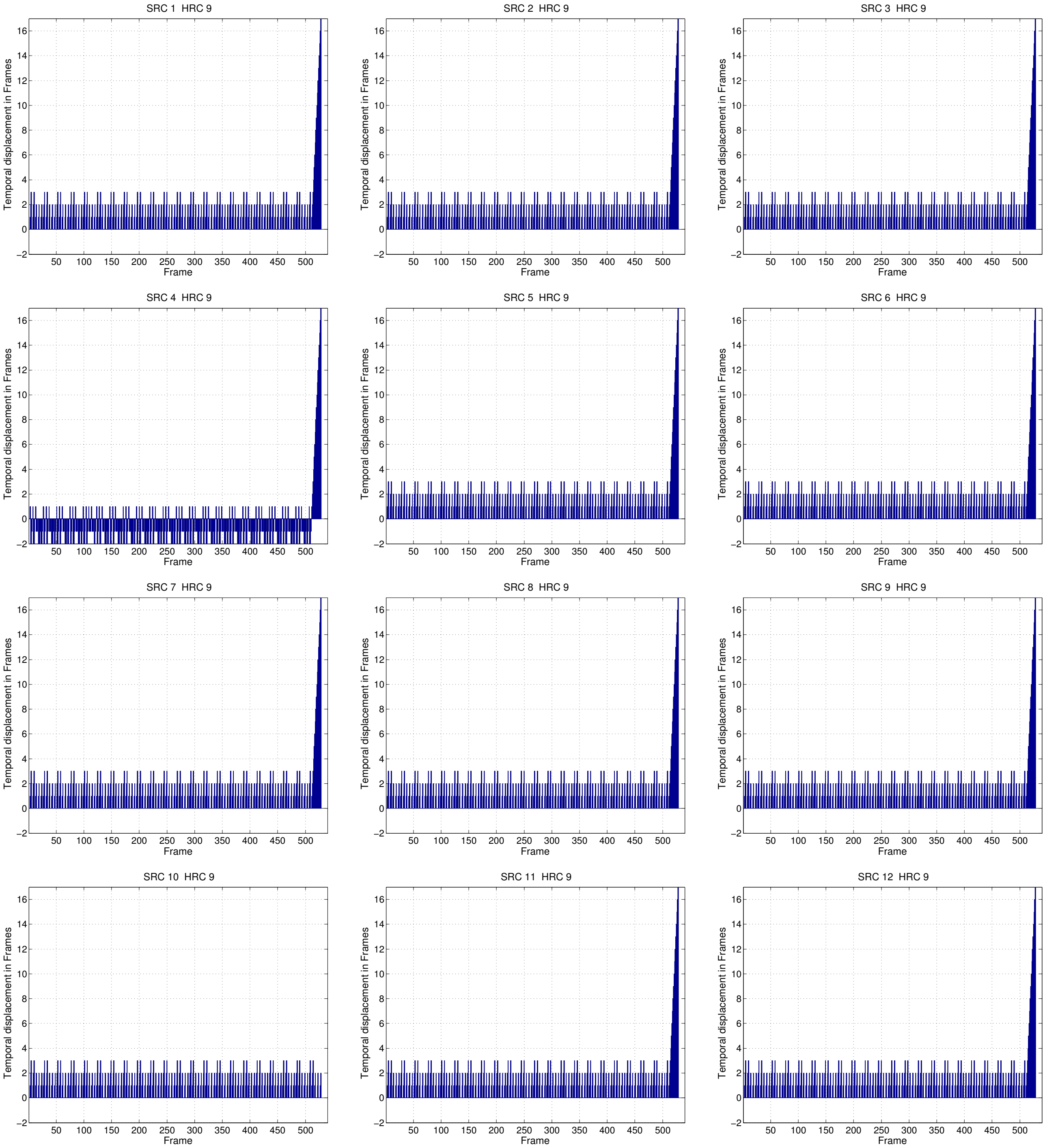


Figure 4.3: Temporal delay between original sequence and HRC9

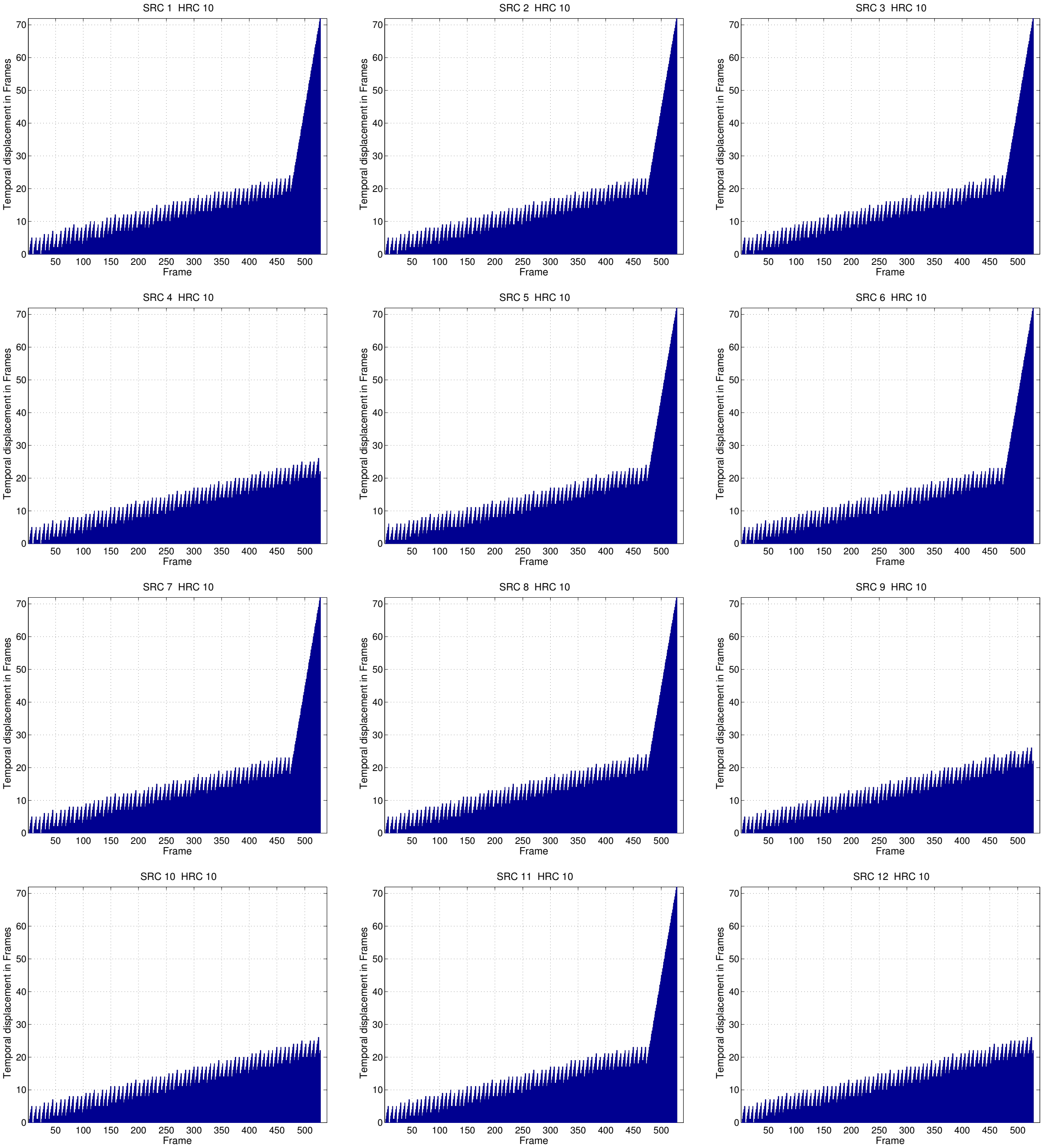


Figure 4.4: Temporal delay between original sequence and HRC10

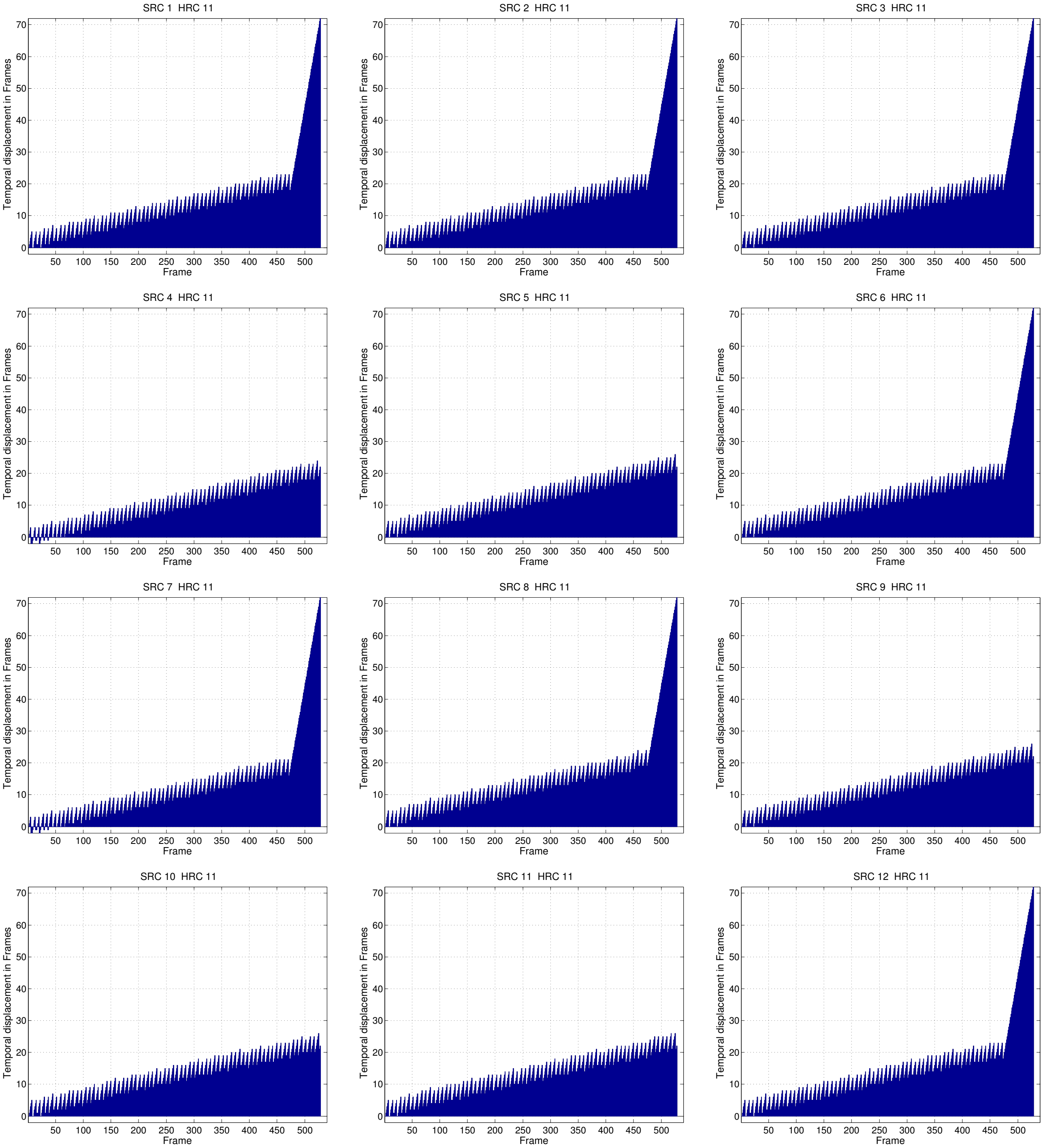


Figure 4.5: Temporal delay between original sequence and HRC11

4.4 JPEG2000 (HRC12-13)

MPEG-4 and H.264 are both blockbased video encoding standards. Both use a DCT to transform the image content. Quite a different approach uses wavelets. The artifacts that occur in wavelet coding are very different from those in DCT based encoding. Usually the image is smoothed.
Although there has been some work on video coding using wavelet based coders lately, for this test only JPEG-2000 could be used, because the other codecs were not available.
The JPEG-2000 standard features the inherent scalability feature that a JPEG-2000 image may be cut-off at an arbitrary point in the bitstream and still remains decodable. This property was used in the preparation.
The image was coded to achieve maximum quality and then the decoder only used as many bits of the bitstream as would have been available at the specified datarate.
The implementation of the codec that was used is the Java based JJ-2000 codec. The encoder parameter file is shown below:

```
# Maximum Datarate
rate = -1
# 5 DWT-levels
Wlev = 5
# 0th (most important) Layer optimized for 0.015 bpp. Then 20 Layers
# for scalability (logarithmic distance) between 0.5bpp and 2.0 bpp
# Layer 21 optimized for 2.0 bpp, then 11 additional layers until rate exceeded:
Alayers = 0.015 +20 2.0 +10
# Mode of progression: LRCP (Layer, Resolution, Component, Precinct)
Aptype = layer
# EPH (End of Packet Header) Markersegments
Peph = off
# Use of SOP (Start of Packet) Markersegments
Psop = off
```

The following table shows the encoded framerate and the PSNR values for the two HRCs:

		HRC12	HRC13
Max. Bitrate [kbit/s]		704	704
Frames per second [Hz]		12.5	25
SRC1	PSNR reached	27.34	25.08
SRC2	PSNR reached	34.34	30.64
SRC3	PSNR reached	37.32	33.72
SRC4	PSNR reached	32.75	29.69
SRC5	PSNR reached	31.94	27.27
SRC6	PSNR reached	26.72	23.92
SRC7	PSNR reached	29.85	26.68
SRC8	PSNR reached	37.93	34.79
SRC9	PSNR reached	22.69	20.08
SRC10	PSNR reached	32.27	29.15
SRC11	PSNR reached	27.59	25.02
SRC12	PSNR reached	33.80	29.68

4.5 Transmission distortion (HRC14-18)

Two different types of distortions have been evaluated. Distortions either result from a block loss or from distorted bits. In typical video transmission systems block-loss occurs more often than biterrors. So only one HRC deals with biterrors (HRC18). The other HRCs (HRC14-17) deal with block loss.
All HRCs are generated from the MPEG-4 bitstream generated for HRC9. They are played back using mplayer which uses concealment techniques.
HRC14 introduces an artificial distortion of setting several times 1kbyte of data to 0. This distortion occurs in the bitstream at the positions 10%, 20%, ...
HRC15 contains only one large distortion. At 40% of the bitstream 10kbyte of data is set to 0.
HRC16 is generated from HRC17 by replacing each frame which contains error concealment artifacts by the previous correctly displayed frame. This results in long lasting frame freezes but no concealment artifacts.
HRC17 contains three distortions which are chosen at 25%, 58% and 91%. At these positions 10kbyte are set to 0. The

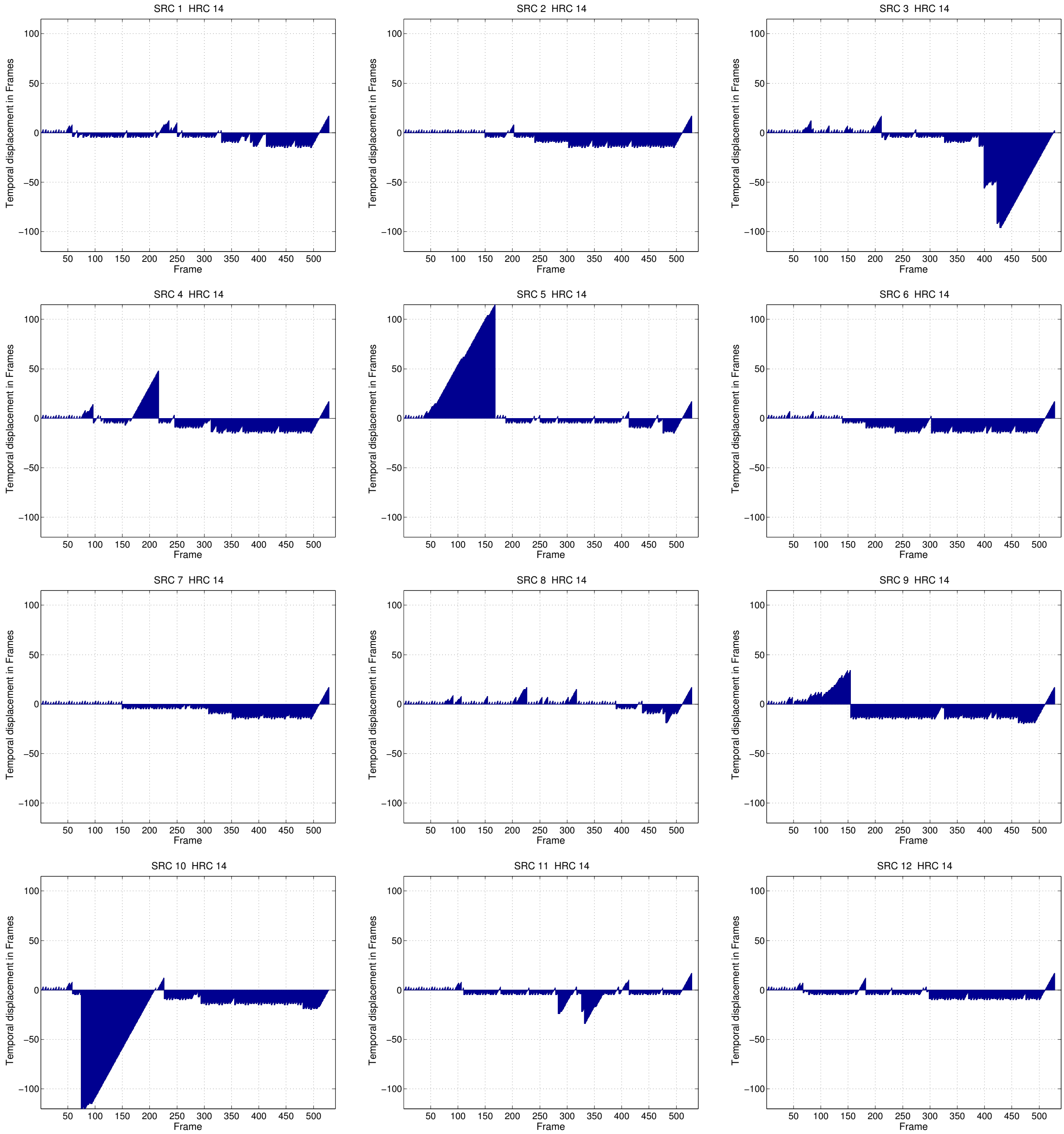


Figure 4.6: Temporal delay between original sequence and HRC14

positions are chosen to avoid collisions with the independently decodable intraframes that are present in the sequence. HRC18 was distorted by inverting on average every 100000th bit. A random factor was used to allow a variation of the error position.

All those distortions lead to frame freezes and frame skipping artifacts. The distortion was introduced, the sequence was played back with mplayer and resampled to 60fps for displaying. After that each frame of the displayed sequence was compared with the frames in the original (undistorted) sequence of hrc0. The best matching original frame was searched using the MSE criterion. Then the delay between the matching frames was plotted. Figures 4.6-4.10 show the resulting plots.

Draft 2005-04-21 Draft 2005-04-21

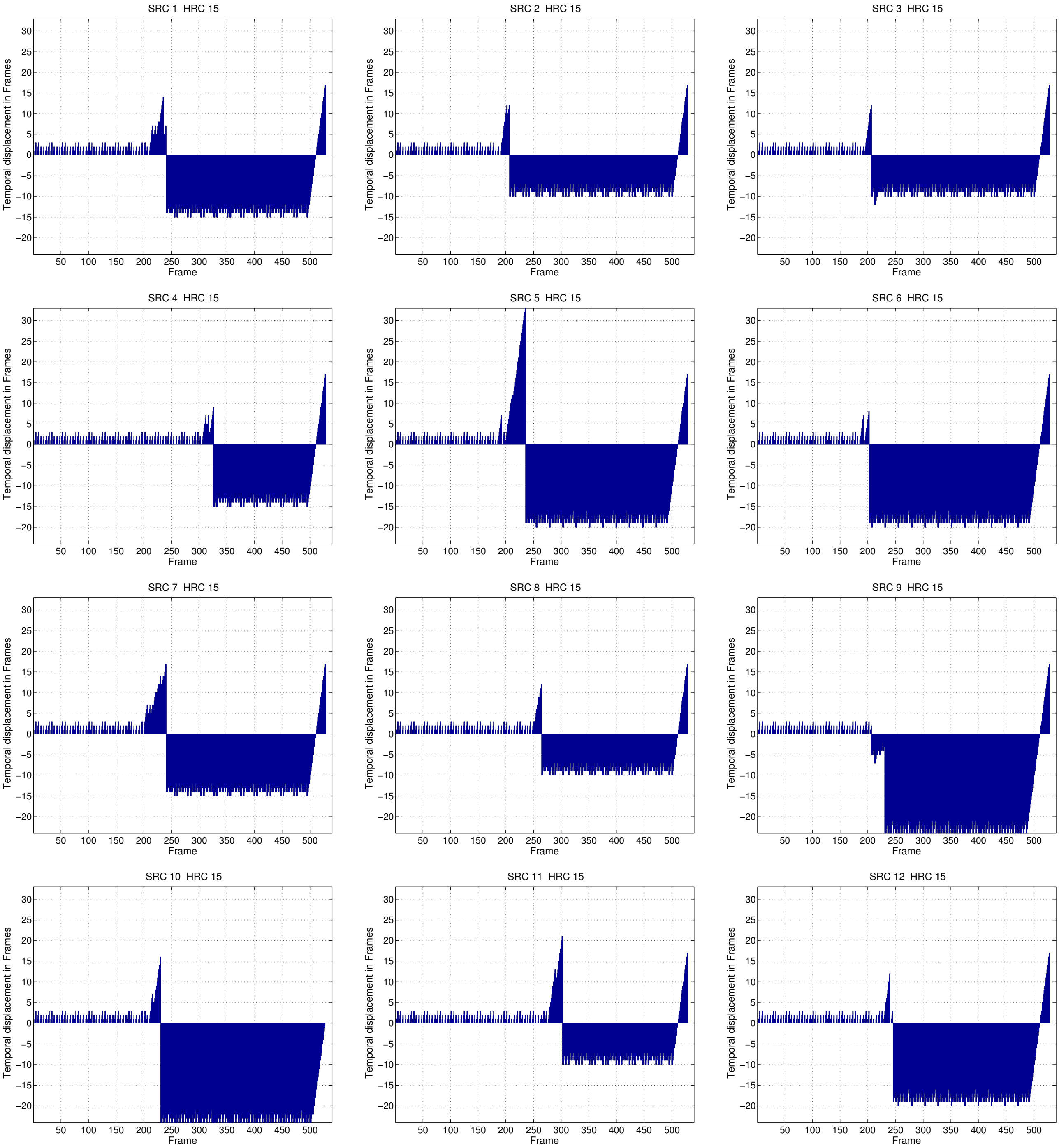


Figure 4.7: Temporal delay between original sequence and HRC15

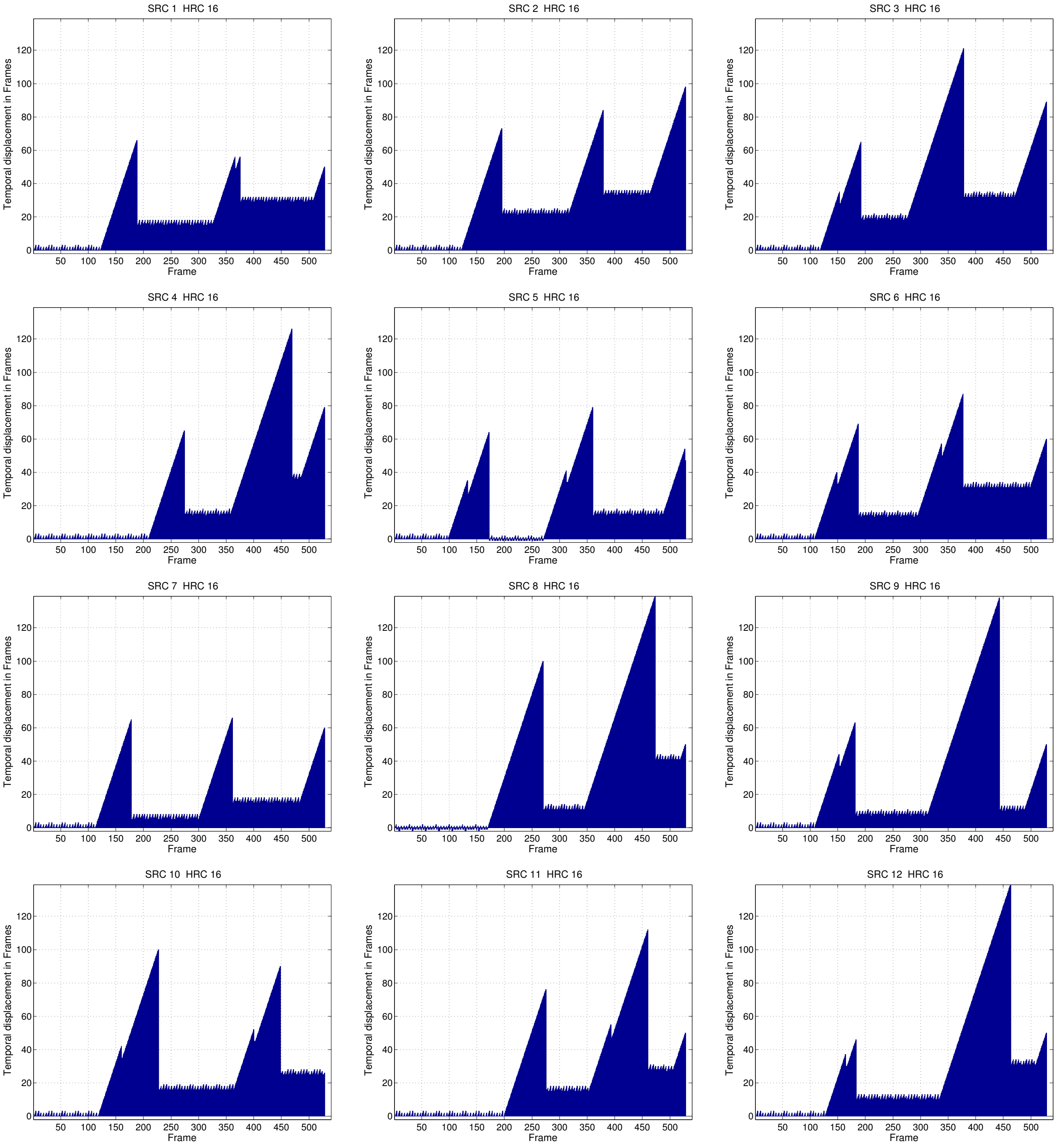


Figure 4.8: Temporal delay between original sequence and HRC16

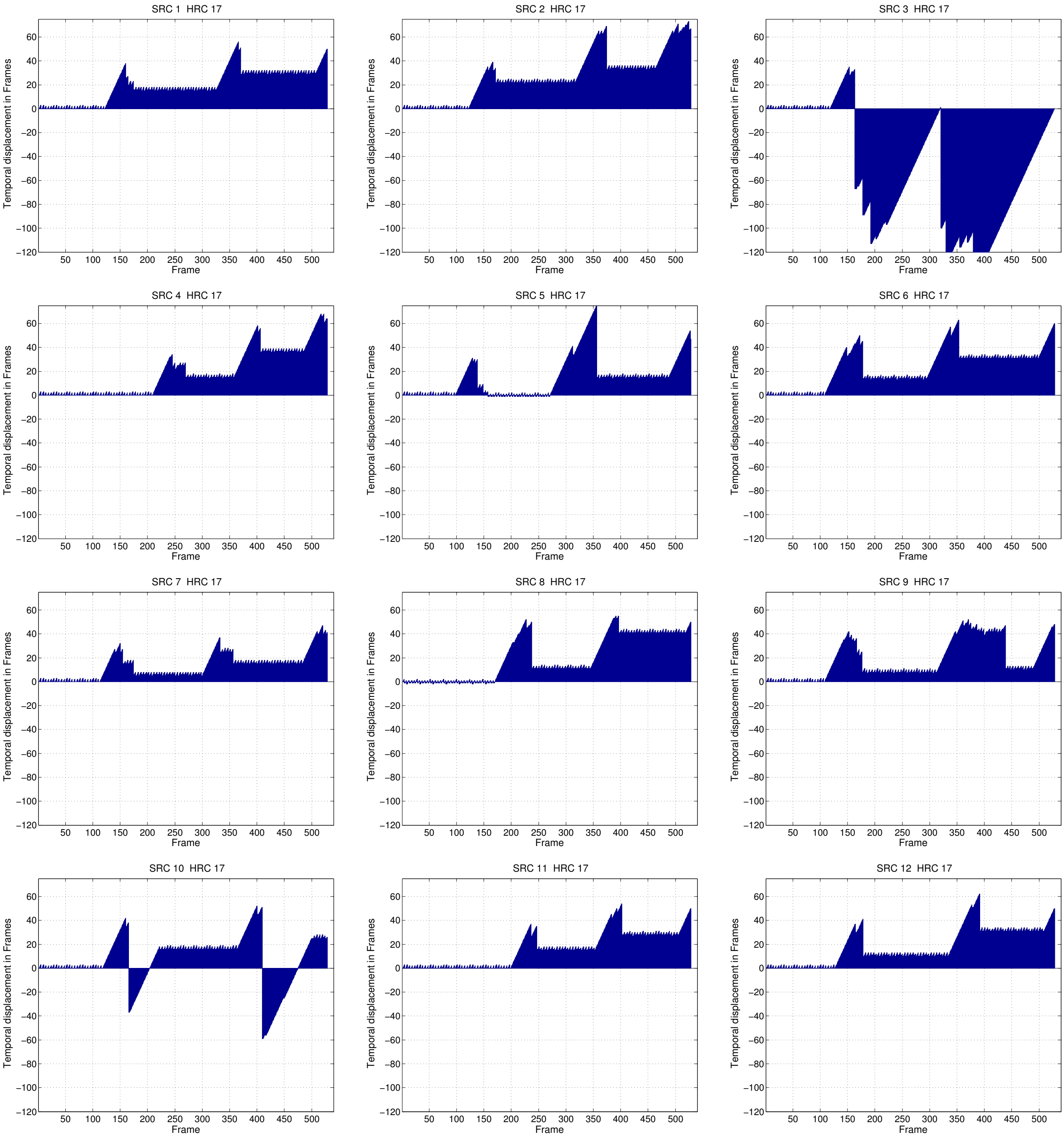


Figure 4.9: Temporal delay between original sequence and HRC17

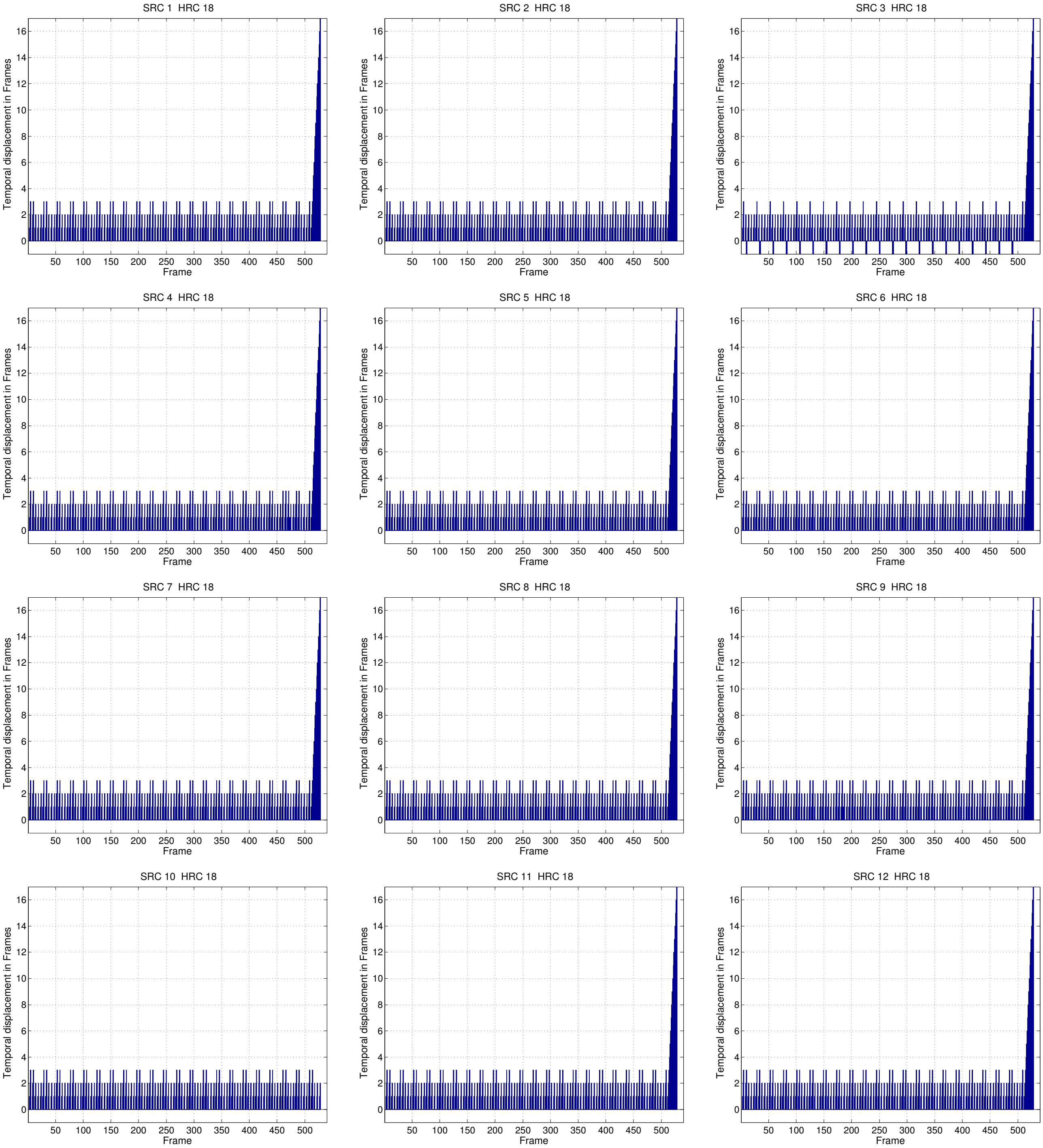


Figure 4.10: Temporal delay between original sequence and HRC18

4.6 Bandwidth limitation (HRC19)

The HRC19 is generated from HRCs 4,5 and 6. The following table specifies which frames are taken from which sequence and approximately how long this lasts in seconds. This HRC simulates a limitation of the bandwidth and a varying channel capacity, because in that case the encoder has to reduce or increase the output bitrate dynamically.

Seconds	FPS	Bitrate	HRC	Frames
0-2	12.5	128	4	0-179
3-4	12.5	64	5	180-299
5	8	64	6	300-359
6-7	12.5	64	5	360-479
8	8	64	6	480-527

4.7 Pictures of typical distortions

It is often difficult to imagine the distortions introduced by the codecs. Therefore figure 4.11 shows some frames that are encoded by H.264 and MPEG-4.
Also very interesting are the distortions that result from transmission errors. These are shown in figures 4.12-4.14.

Figure 4.11: Typical distortion of H.264 (HRC4) and MPEG-4 (HRC9)

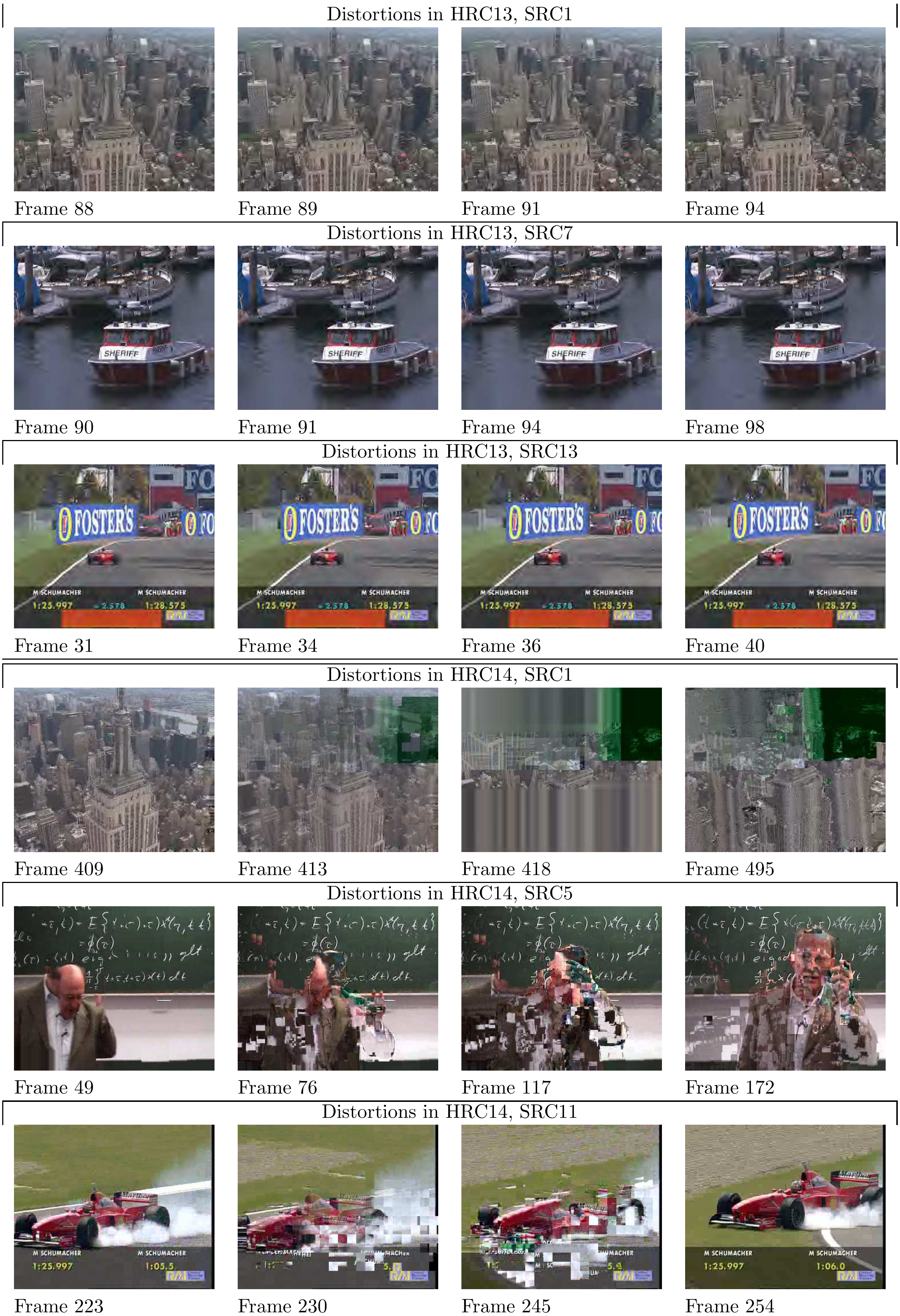


Figure 4.12: Typical distortion of HRC13 and HRC14

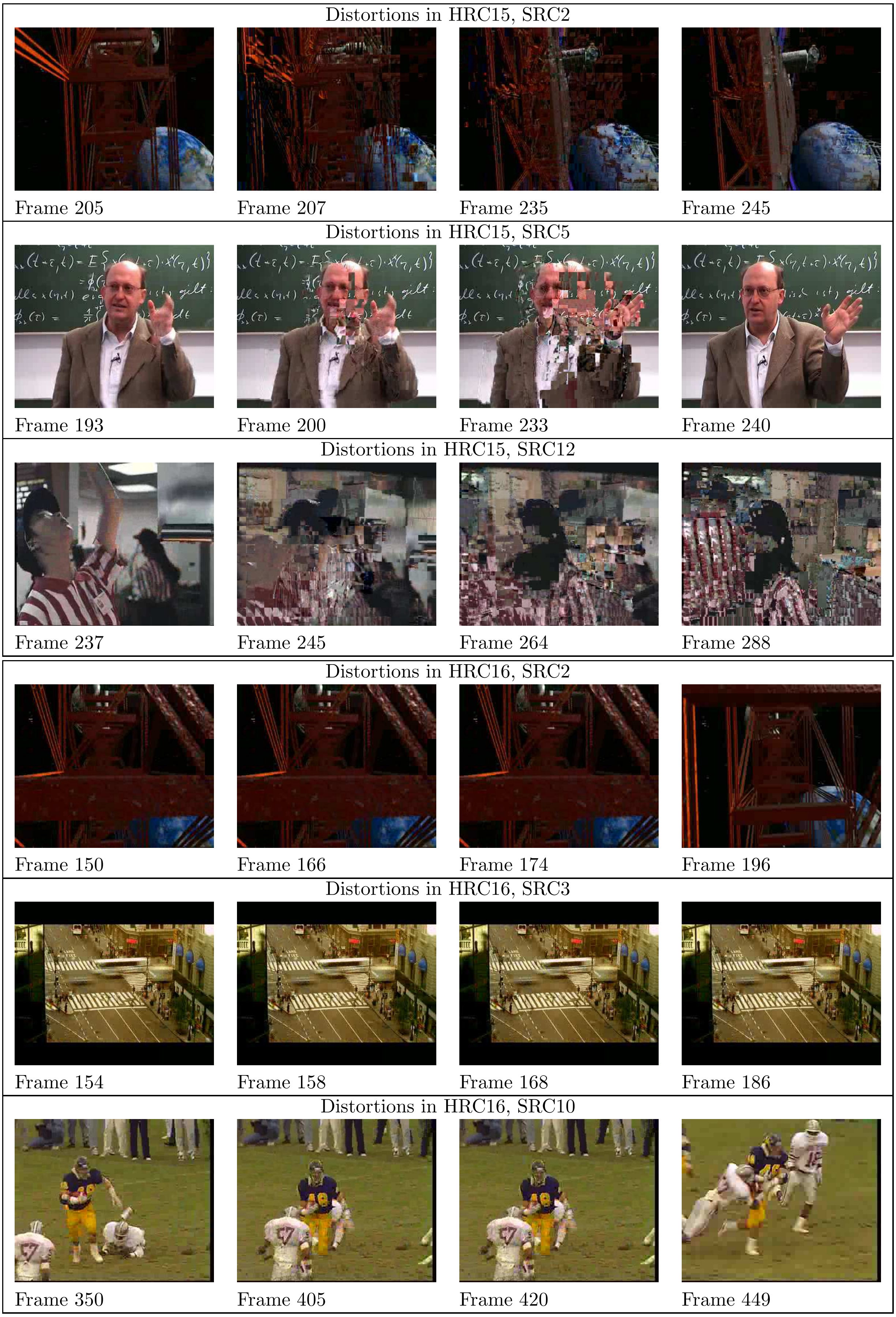


Figure 4.13: Typical distortion of HRC15 and HRC16

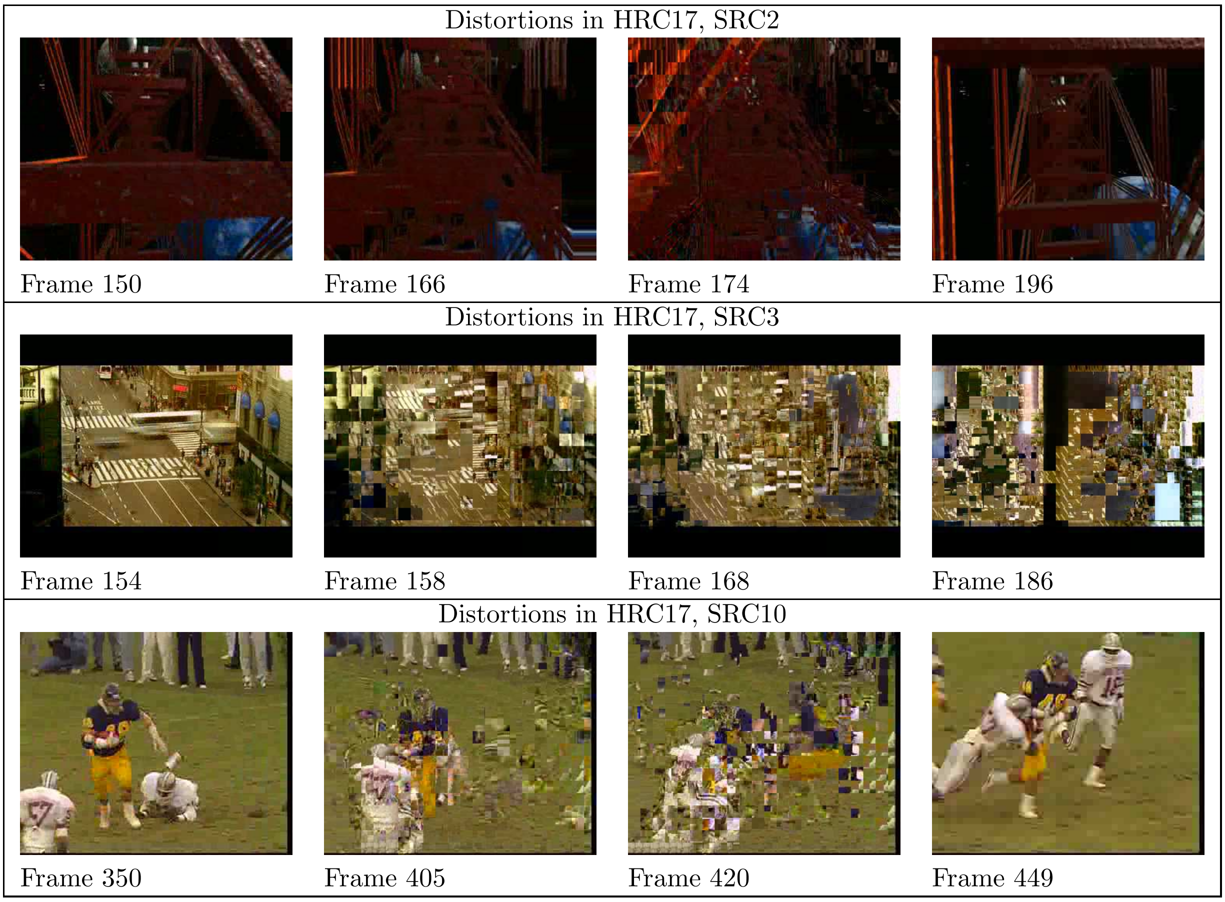


Figure 4.14: Typical distortion of HRC17