



# LUND UNIVERSITY

## Effect of Compressed Offline Foveated Video on Viewing Behavior and Subjective Quality

Nyström, Marcus; Holmqvist, Kenneth

*Published in:*

ACM Transactions on Multimedia Computing, Communications, and Applications

*DOI:*

[10.1145/1671954.1671958](https://doi.org/10.1145/1671954.1671958)

2010

[Link to publication](#)

*Citation for published version (APA):*

Nyström, M., & Holmqvist, K. (2010). Effect of Compressed Offline Foveated Video on Viewing Behavior and Subjective Quality. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 6(1), 4:1-4:14. <https://doi.org/10.1145/1671954.1671958>

*Total number of authors:*

2

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00



# Effect of Compressed Offline Foveated Video on Viewing Behavior and Subjective Quality

MARCUS NYSTRÖM and KENNETH HOLMQVIST  
Lund University

Offline foveation is a technique to improve the compression efficiency of digitized video. The general idea behind offline foveation is to blur video regions where no or a small number of previewers look without decreasing the subjective quality for later viewers. It relies on the fact that peripheral vision is reduced compared to central vision, and the observation that during free-viewing humans' gaze positions generally coincide when watching video. In this article, we conduct two experiments to assess how offline foveation affects viewing behavior and subjective quality. In the first experiment, 15 subjects free-viewed six video clips before and after offline foveation whereas in the second experiment we had 17 subjects assessing the quality of these videos after one, two, and three consecutive viewings. Eye movements were measured during the experiments. Results showed that, although offline foveation prior to encoding with H.264 yielded data reductions up to 52% (20% average) on the tested videos, it had little or no effect on where people looked, their intersubject dispersion, fixation duration, saccade amplitude, or the experienced quality during first-time viewing. However, seeing the videos more than once increased the intersubject dispersion and decreased the subjective quality. In view of these results, we discuss the usage of offline foveated video in practical applications.

Categories and Subject Descriptors: E.4 [Data]: Coding and Information Theory—*Data compaction and compression*; H.1.2 [Information Systems]: User/Machine Systems—*Human information processing*

General Terms: Algorithms

Additional Key Words and Phrases: Eye-tracking, subjective quality, video compression, foveation

## ACM Reference Format:

Nyström, M. and Holmqvist, K. 2010. Effect of compressed offline foveated video on viewing behavior and subjective quality. *ACM Trans. Multimedia Comput. Commun. Appl.* 6, 1, Article 4 (February 2010), 14 pages.  
DOI = 10.1145/1671954.1671958 <http://doi.acm.org/10.1145/1671954.1671958>

## 1. INTRODUCTION

Current standards for video compression exploit the characteristics of the Human Visual System (HVS) rather sparingly. Lately, it has been suggested to improve compression efficiency by incorporating a previously unused feature of the HVS: the varying spatial resolution of the human retina. Humans have a high resolution only in the fovea, whereas peripheral regions of the retina sample the visual input very sparsely. As a consequence, peripheral vision is significantly reduced compared to foveal

Parts of the results of this article were presented at the Scandinavian Workshop on Applied Eye-Tracking (SWAET'07). Data recordings were performed at Lund Humanities Lab, Lund University.

Authors' addresses: M. Nyström (corresponding author), K. Holmqvist, Lund Humanities Lab, Lund University, Box 201, 22100, Lund, Sweden; email: [marcus@eit.lth.se](mailto:marcus@eit.lth.se).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2010 ACM 1551-6857/2010/02-ART4 \$10.00

DOI 10.1145/1671954.1671958 <http://doi.acm.org/10.1145/1671954.1671958>

ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 6, No. 1, Article 4, Publication date: February 2010.

vision, which spans only two degrees in the line of sight. This opens the possibility for substantial data compression.

The varying spatial resolution of the HVS has previously been used for improved data compression by removing redundant visual information in the periphery, or giving coding priority to foveal regions, given that the observer's (predicted) position of gaze is known and measured continuously. One of the early attempts was presented in Kortum and Geisler [1996] where image resolution was matched with the resolution degradation of the HVS, implemented by increasing the pixel size away from a person's online measured gaze position. They called these image manipulations *foveation*. Results reported of substantial bitrate reduction due to foveation with minimal perceptual loss in image quality. Foveated compression, mainly adapted for real-time usage, has been extended to video [Geisler and Perry 1998, 1999], and has been investigated in several other papers [Sheikh et al. 2001; Wang and Bovik 2001; Bergström 2003; Wang et al. 2003].

Maybe a less intuitive way to use the fact that vision is reduced in the periphery, called *offline foveation*, is to beforehand try to predict where viewers will look and keep a high display fidelity only in these regions. If the predicted regions are consistent with the gaze positions of actual viewers, then offline foveation should not reduce the subjective quality of the video, since viewers do not look at peripheral regions where quality has been reduced.

One of the main challenges in offline foveated video is how to accurately predict where future viewers will direct their gazes. Due to individual differences in where people look depending on task and personal interests, for example, it might at a first glance seem futile. Earlier works used gaze positions from previewers to produce offline foveated video [Stelmach and Tam 1994; Duchowski and McCormick 1998]. Unfortunately, their implementations decreased the subjective quality, and also seemed to change the viewing behavior of the tested subjects. These results were reevaluated in Nyström et al. [2004], reporting that offline foveation essentially did not change where subjects look during free-viewing. However, some of the subjects looking at the offline foveated videos had noticed the poor peripheral quality, hence decreasing their overall viewing experience. A refined implementation of offline foveation, presented in Nyström and Holmqvist [2007], showed to not decrease the subjective quality of the tested videos despite yielding substantial bitrate reductions. However, since only three short (eight second) video clips were used, only preliminary conclusions could be drawn of how offline foveation affects subjective quality in a more general sense. An example of this implementation is visualized in Figure 1. Figure 1(a) shows a frame from the standard test sequence *Football* where the gaze positions, one from each observer, are marked by crosshairs. Figure 1(b) depicts the same frame after offline foveation. Notice how unattended regions are blurred in the latter case.

In this article, we will measure how offline foveation, as implemented in Nyström and Holmqvist [2007], contributes to increased data reduction and, in particular, affects the perception of a video. Perceptual effects are quantified by gaze locations, fixation durations, saccade lengths, and subjective quality assessments. Two different experiments are conducted. In the first experiment, we perform eye-tracking on 15 subjects free-viewing six video clips before and after foveation, and analyze how offline foveation affects eye movement behavior. In the second experiment, 17 new subjects evaluate the quality of the same videos, again while their eye movements are recorded. We question whether traditional methods for subjective quality assessment, where videos typically are viewed twice or more before being judged, are suitable to evaluate offline foveated video. Therefore, we investigate the effect repeated viewings have on eye movement parameters and subjective quality by showing the videos three consecutive times to each viewer. While the quality assessments quantify the subjective changes in quality due to offline foveation, recorded gaze data will help us understand the nature of these changes. The main contributions of this article are (1) a new methodology to evaluate offline foveated video quality through tailored subjective tests and eye-tracking measurements, and (2) through this



Fig. 1. Example of offline foveation. The left picture shows where people look in the original frame and the right picture depicts the same frame after offline foveation. Each crosshair represents one viewer's position of gaze. Peripheral blurring due to offline foveation enables more efficient compression.

new methodology show that (and why) offline foveation indeed is successful in decreasing the bitrate without decreasing subjective quality during natural free-viewing.

Section 2 briefly covers the implementation of offline foveation and also presents how it influences compression efficiency. Then follow explanations of the experimental setup and methodology as well as presentation of the results of two experiments in Sections 3 and 4. Section 5 contains a general discussion of the results.

## 2. OFFLINE FOVEATION: IMPLEMENTATION AND COMPRESSION GAINS

Previous attempts to implement offline foveation based on gaze positions from several previewers can be found in Stelmach and Tam [1994], Duchowski and McCormick [1998], and Nyström et al. [2004]. As previously mentioned, these implementations largely failed to maintain a high subjective quality after offline foveation. The reason for this lack of success was, according to the authors, most likely caused by naive design criteria. In Stelmach and Tam [1994], for example, only one region was chosen as Region Of Interest (ROI) based on gaze coordinates, even though other regions also attracted several viewers' gazes. The approach in Duchowski and McCormick [1998] suffered abrupt, temporal onsets of new high-resolution regions, which were annoying for later viewers. In Nyström et al. [2004], the largest problem concerned small high-resolution regions originating from a single previewer's gaze, "floating around" outside the main ROIs.

The insights from these initial attempts motivated two new design criteria: (1) ROIs constructed from previewers should with a high probability confine gaze positions of new viewers watching the offline foveated video. (2) Offline foveation should not introduce sudden onsets or offsets of new, high-resolution regions. Based on the two previously described design criteria, we have earlier presented a method that transfers gaze positions recorded over a group of frames to Volumes Of Interest (VOIs). These VOIs are used to implement offline foveation in the wavelet domain by attenuating frequencies in regions unattended by the previewers [Nyström and Holmqvist 2007].

Initial subjective evaluations in Nyström and Holmqvist [2007] showed that this implementation of offline foveation did not decrease the subjective quality of the tested videos. Since we use this implementation of offline foveation herein, and since it has previously been described in detail, this section only outlines the main components of the implementation as visualized in Figure 2.

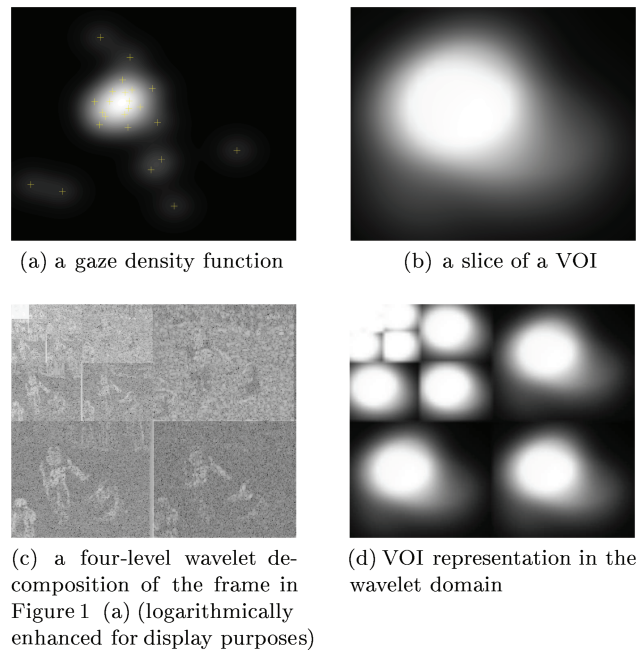


Fig. 2. Illustration of how offline foveation is implemented.

Figure 2(a) shows a Gaze Density Function (GDF) for one frame. It is generated by centering a Gaussian function at each previewer's gaze position, and then summing all such functions. A GDF estimates the probability of where later viewers will look; it is likely that new viewers will look at frame regions with high gaze density. VOIs are generated through a number of steps. First, a clustering algorithm is used to identify and remove spatial outliers from each frame. Then, remaining gaze positions are clustered in the temporal direction such that only gaze positions forming a stable cluster over a predefined amount of time are kept. To allow viewers of the foveated video to plan and execute a saccade ( $\sim 200$  ms), and dwell at this cluster location for a typical fixation duration ( $\sim 300$  ms), this predefined time is set to 500 ms. Gaze positions surviving both spatial and temporal clustering are used to create GDFs for each frame. A VOI is then generated by convolving a number of temporally adjacent GDFs with a one-dimensional Gaussian kernel operating in the temporal direction. Temporal filtering smoothes rapid, individual changes in gaze positions, which do not contribute to the collective modeling of visual interest. Figure 2(b) illustrates a slice of the VOI representing the same set of gaze positions as in Figure 2(a). Notice how the shape of the VOI is influenced only little by single, isolated gaze positions. Instead, a larger region is defined that with a high probability confine new viewers' gaze positions.

Several methods to reduce the display fidelity away from the position of gaze have been suggested. Early methods increased the pixel size in the periphery [Kortum and Geisler 1996]. Other methods used multiresolution, Gaussian pyramids, where visually attended regions were represented by layers from the bottom of the pyramid while higher layers comprised peripheral regions [Geisler and Perry 1998]. More recently, wavelets have successfully been used for this purpose [Duchowski 2000; Wang et al. 2003]. Wavelets offer a computationally efficient method for foveation that yields smooth peripheral degradation. We use the VOIs in the wavelet domain to reduce the display fidelity away from



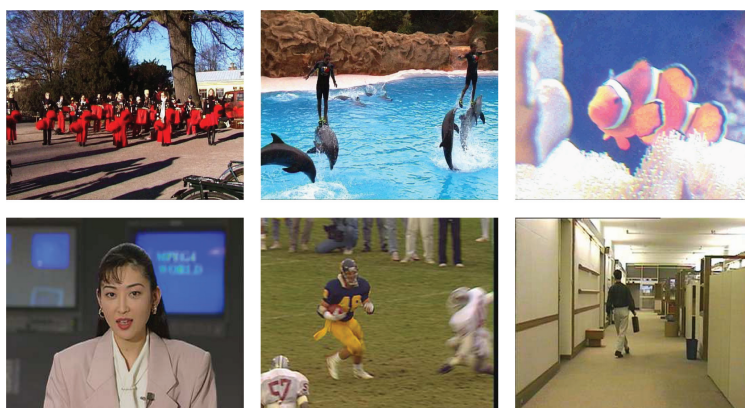


Fig. 3. Representative frame from each of the tested videos Alte, Dolphin, Fish, Aikyo, Football, and Hall.

regions with high gaze density. Figure 2(c) shows a four-level wavelet decomposition of the frame in Figure 1(a), and is generated by recursive lowpass and highpass filtering of the original frame followed by downsampling. Simply explained, the upper-left subband in the decomposition represents a lowpass filtered and downsampled version of the original frame whereas the other subbands represent additional information about the original frame, which is needed to synthesize the original frame through inverse wavelet filtering.

In practice, offline foveation is implemented by multiplying a wavelet decomposed frame with its corresponding VOI representation. Figure 2(d) shows how the VOI slice in Figure 2(b) is represented in the wavelet domain. The bright areas represent regions where wavelet coefficients are scaled by a value close to unity such that a high frame quality remains, whereas in the darker areas coefficients are downsampled by values between 0 and 1, consequently introducing different degrees of blurring after inverse wavelet transformation. The amount of peripheral blurring is decided by pilot testing such that it is not noticed by a new viewer looking at regions of high gaze density.

Since offline foveation introduces blur in unattended video regions and thus reduces entropy, it is expected to increase compression efficiency. We have estimated the effect offline foveation has on data reduction for the six different videos shown in Figure 3. For all videos used in this article, offline foveation was implemented as described in Nyström and Holmqvist [2007] with data recorded from Nyström et al. [2004] and Johannesson [2005]. Each of the videos was encoded with the Quicktime Professional H.264 codec before and after offline foveation for a set of encoding parameters. As can be seen from Table I, the average bitrate reduction due to offline foveation was on average 20%, and as much as 52% in one of the tested videos. H.264 is in no way optimized to encode offline foveated videos, and bitrate savings are therefore expected to increase if using a codec specifically tailored for this purpose.

### 3. EXPERIMENT I—FREE VIEWING

In order to investigate the perceptual effects offline foveation has on subjects during free-viewing, we measure in the first experiment how eye movements are influenced by offline video foveation in terms of spatial and temporal distribution, fixation time, saccade length, and repeated viewings. These measures will help us understand how offline foveated videos are perceived during task neutral, “normal” viewing conditions.

Table I. Additional Compression Due to Offline Foveation Before Video Encoding with H.264 for Different Quality Factors (as defined in Quicktime 7 Pro.) Results are presented for the six video clips in Figure 3

Video	Quality factor	Bitrate <sub>Unfov</sub> (kbps)	Bitrate <sub>Fov</sub> (kbps)	Gain
Alte	Lowest	475	413	0.13
	Low	1471	1050	0.29
	Medium	6734	3237	0.52
	High	16624	9161	0.45
	Highest	28948	19414	0.33
Dolphin	Lowest	812	658	0.19
	Low	2958	2151	0.27
	Medium	10017	6620	0.34
	High	23536	15372	0.35
	Highest	41282	29229	0.29
Fish	Lowest	567	490	0.14
	Low	1668	1276	0.24
	Medium	6037	4106	0.32
	High	14606	10431	0.29
	Highest	26573	20979	0.21
Aikyo	Lowest	78	77	0.02
	Low	120	118	0.01
	Medium	222	219	0.01
	High	568	534	0.06
	Highest	2268	1902	0.16
Football	Lowest	578	504	0.13
	Low	1301	1106	0.15
	Medium	2928	2461	0.16
	High	6466	5416	0.16
	Highest	15208	12269	0.19
Hall	Lowest	118	114	0.03
	Low	220	214	0.03
	Medium	726	680	0.06
	High	4214	3370	0.20
	Highest	12214	10066	0.18

### 3.1 Subjects

15 naive subjects (nine females) of ages  $30.2 \pm 16.1$  ( $M \pm SD$ ) years volunteered to take part in the experiment. They all had normal or corrected-to-normal vision.

### 3.2 Stimuli

Stimuli consisted of six original videos shown in Figure 3 and six offline foveated versions of these, thus 12 videos in total. The three videos in the upper row in Figure 3 were eight seconds long with resolution  $720 \times 576$  and those in the bottom row  $352 \times 288$  (CIF format) pixels and of durations five, three, and four seconds, counting from the left. All videos were displayed in color at 25 fps and encoded with H.264





Fig. 4. Experimental setup.

(in Quicktime 7 Pro.) at high bitrates (quality factor high). No sound was used. Offline foveation was implemented as described in Nyström and Holmqvist [2007].

### 3.3 Viewer Instructions

To mimic a typical viewing scenario, subjects were given no explicit task and were asked to view the stimuli as they normally would. To prevent subjects from trying to guess the purpose of the experiment, they were told that the study would investigate mental workload by measuring the pupil size. This way, focus was drawn away from the fact that gaze positions were recorded. Subjects were further informed that the same video clip could occur more than once during one presentation.

### 3.4 Experimental Setup and Procedure

Figure 4 illustrates the setup. Each subject was placed at a viewing distance of 76.5 cm in front of a 19-inch computer screen with resolution  $1280 \times 1024$  and update rate 75 Hz. The active screen area subtended a visual angle of 27.7 degrees horizontally and 22.5 degrees vertically. A chin rest was used to restrict head movements.

Prior to each recording, a 13-point spatial calibration was performed. During data recording, all 12 videos were presented one after the other on the screen, separated in time by a mid-gray image displayed for one second. They were displayed in full screen while maintaining their aspect ratio. No prefixation cross was used to restrict subjects' initial gaze position. The order was randomized with the restriction that two versions (unfoveated and offline foveated) of the same video could not be displayed directly after each other. To see how repeated viewing affects eye movement behavior, all 12 videos were presented twice more in the same manner. In total, each video was viewed three times by each subject.

Eye movements were recorded monocularly with an SMI iView X Hi-Speed eye-tracker, sampling gaze positions at 240 Hz with gaze positions accuracy  $0.2^\circ$ . On average, 9.6 gaze coordinates were recorded for each displayed frame. Offline analysis of gaze positions was performed to classify eye movements into fixations and saccades using IDFconvert, a software package provided by SMI. A Matlab script was developed to collect data about the subjects, communicate with the eye-tracker, display the videos in the Quicktime player, and control the accuracy in timing during the experiments.

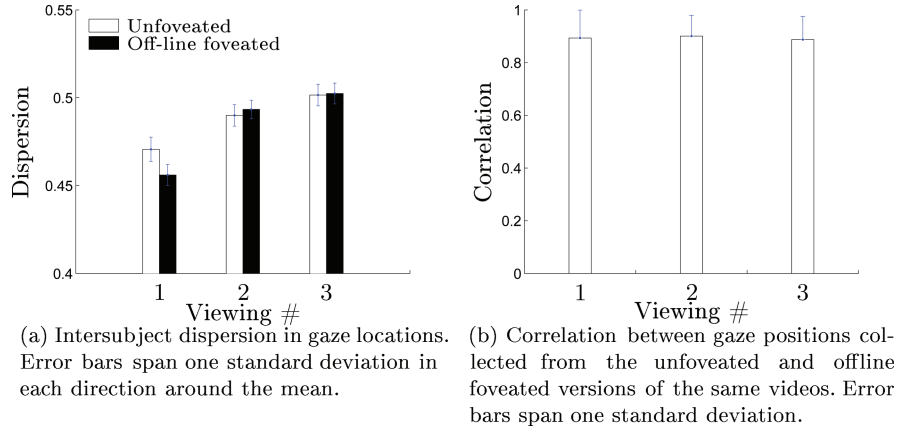


Fig. 5. Eye movement behavior during free-viewing before and after offline foveation.

### 3.5 Data Analysis and Results

The perceptual effects offline foveation has on eye movements are assessed by comparing the gazing behavior of viewers watching the tested videos before and after offline foveation. More precisely, we measure how offline foveation influences intersubject dispersion, where people look, and how fixation duration and saccade length are affected.

The intersubject dispersion  $D_t$  at time  $t$  is calculated as

$$D_t = \frac{1}{N} \sum_{i=1,2,\dots,N} \frac{g_{t,max}^{i'} - g_t^{i'}(x_i, y_i)}{g_{t,max}^{i'} - g_{t,avg}^{i'}} \quad (1)$$

$x \in \{1, 2, \dots, m\}$  and  $y \in \{1, 2, \dots, n\}$ , where  $m \times n$  represents the frame dimensions.  $g_t(x, y)$  represents a Gaze Density Function (GDF)

$$g_t(x, y) = \frac{1}{N} \sum_{i=1,2,\dots,N} \varphi_i(x, y) \quad (2)$$

generated by replacing  $N$  gaze positions  $(x_i, y_i)$  recorded at time  $t$  by a Gaussian function

$$\varphi_i(x, y) = e^{-\left(\frac{(x-x_i)^2}{2\sigma^2} + \frac{(y-y_i)^2}{2\sigma^2}\right)} \quad (3)$$

and then superimposing all Gaussian functions.  $g_{t,max}(x, y)$  and  $g_{t,avg}(x, y)$  denote the maximum and average of  $g_t(x, y)$  while  $i'$  denotes that the  $i$ th gaze position was excluded when generating  $g_t^{i'}(x, y)$ . In order to get an intuitive measure of how disperse gaze positions are relative each other,  $\sigma = 0.10m$  is chosen such that the “full width at half maximum height” of a Gaussian function approximately spans the parafoveal region ( $5^\circ$ ) of an observer in our experimental setup (refer to, e.g. Rajashekar et al. [2004] who used a similar strategy to motivate  $\sigma$ ).

Figure 5(a) illustrates the intersubject dispersion after one, two, and three viewings of the unfoveated (white bars) and offline foveated (black bars) videos. It can be seen that offline foveation has no or little effect on the intersubject dispersion, that is, the degree to which different subjects' gaze positions coincide. However, during first-time viewing, there is a significant ( $p < 0.05$  in a two-sample Kolmogorov-Smirnov test) effect showing that the intersubject dispersion decreases due to offline foveation. There are a number of explanations to why this effect is present. One is offline foveation

reduces stimulus-dependent cues in regions unattended by most previewers, and therefore prevents new viewers to saccade toward these regions. Another explanation is that offline foveation redistributes the semantic content of the video toward regions where the majority of previewers looked. It is currently debated how stimulus-dependent, bottom-up factors, and higher cognitive, top-down factors interplay to guide fixations in static and dynamic scenes (see, e.g., Parkhurst et al. [2002], Henderson [2003], Le Meur et al. [2007], and Nyström and Holmqvist [2008]).

Another clear effect is that the intersubject dispersion increases significantly after repeated viewings, both for unfoveated and offline foveated videos. This type of behavior is likely to occur since additional viewings encourage more individual viewing strategies deriving from, for example, memory-induced top-down effect (refer to Tatler et al. [2005]).

To estimate the similarity between two sets of gaze positions  $\mathcal{A}$  and  $\mathcal{B}$ , we compute the correlation coefficient

$$\rho = \frac{\sum_{x,y} (g_t^{\mathcal{A}}(x, y) - g_{t,avg}^{\mathcal{A}})(g_t^{\mathcal{B}}(x, y) - g_{t,avg}^{\mathcal{B}})}{\sqrt{(\sum_{x,y} (g_t^{\mathcal{A}}(x, y) - g_{t,avg}^{\mathcal{A}})^2)(\sum_{x,y} (g_t^{\mathcal{B}}(x, y) - g_{t,avg}^{\mathcal{B}})^2)}} \quad (4)$$

between the GDFs  $g_t^{\mathcal{A}}(x, y)$  and  $g_t^{\mathcal{B}}(x, y)$  generated from  $\mathcal{A}$  and  $\mathcal{B}$ , respectively. Figure 5(b) shows how gaze positions recorded from viewers watching the unfoveated video correlate with those watching the offline foveated video after the first, second, and third viewing. It can be seen that the correlation is high in all three cases, indicating that subjects' gaze positions are similarly distributed.

No significant differences in fixation duration or saccade length were found, neither between unfoveated and offline foveated videos nor between the first, second, or third time viewing.

To summarize the results in this section, offline foveation has no or little effect on gaze locations, fixation durations, and saccade lengths, given that the offline foveated videos are viewed for the first time with a neutral viewing instruction. However, after multiple viewing, subjects' gaze behavior changes and regions initially unattended become gazed at. This implies that some subjects look outside the regions of high quality in the offline foveated video, and thus are more likely to notice the introduced blur. We perform in the next section subjective evaluations to investigate whether offline foveation decreases the subjective quality after one, two, and three viewings.

#### 4. EXPERIMENT II—QUALITY ASSESSMENT

The stimuli and experimental setup were the same as in Experiment I. However, there were differences in both task instructions and procedure as described next.

17 naive subjects of ages  $23.8 \pm 4.2$  ( $M \pm SD$ ) years were asked to estimate the difference in quality between two versions, A and B, of the same video in an AB trial. They were told that the two versions resulted from different compression algorithms being applied to the original video. To encourage subjects to do their best and maintain focus during the evaluation, they were told that quality assessment is a difficult task and the differences in quality may sometimes be hard to notice. As in Experiment I, subjects were told that the study would investigate mental workload during quality assessment by measuring the pupil size.

The videos were assessed as follows. Each AB trial started by displaying a uniform mid-gray frame with a large, centered black capital A, followed by version A of the stimulus. Directly after A had been shown followed the same procedure for version B, whereafter a pop-up window containing a slider bar and a button appeared on the screen (see Figure 6). On the slider bar, three different levels of quality were given: *A better than B*, *A equal to B*, *B better than A*. Subjects could then adjust the slider to any (intermediate) level reflecting their experienced quality. A integer value between  $-5$  and  $+5$  was collected after each displayed AB pair to represent the difference in quality between A and B. When the

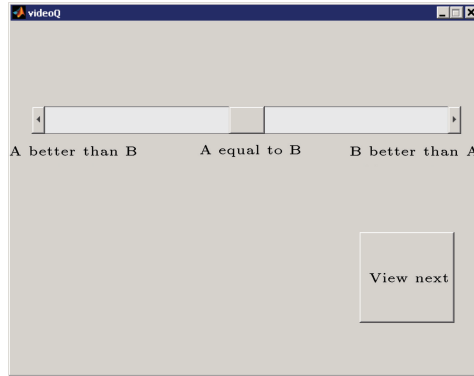


Fig. 6. Pop-up window for quality assessment.

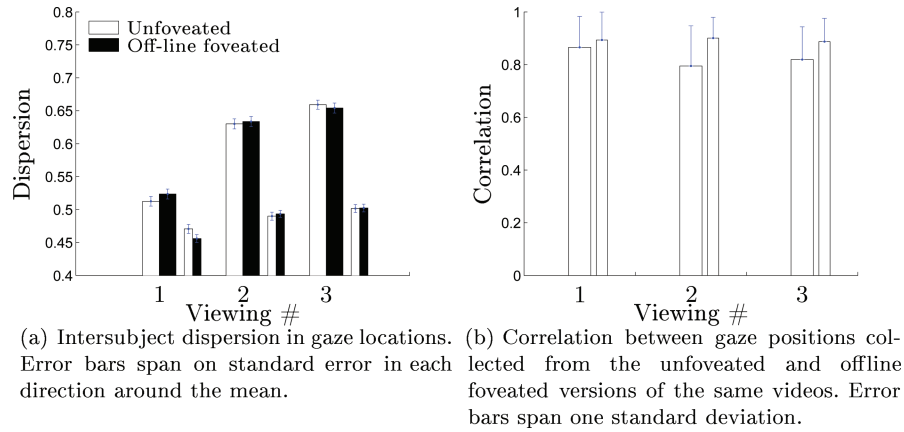


Fig. 7. Eye movement behavior during quality assessment before and after offline foveation. For comparison, results from Experiment I are included in the figure, depicted by the thinner bars.

quality vote was taken, a button press would continue to trial. The presentation order of the video (AB) pairs was randomized. As with version A, version B could denote either the unfoveated or the offline foveated version. After the data collection had completed, the quality votes were sorted such that a positive quality vote indicated that the unfoveated version was judged as being of better quality, and vice versa.

In standardized methods for quality evaluation, subjects are usually allowed to view the videos to be assessed several times before giving the actual judgment. Therefore, to see the effect a quality evaluation task has on repeated viewings, the aforesaid video pairs were shown another two times whereafter a second quality vote was given. Subjects did not know in advance that further chances to evaluate the quality would be given.

#### 4.1 Data Analysis and Results

Figure 7 compares the dispersion of and correlation between gaze positions collected before and after offline foveation during the first-, second-, and third-time viewing. For comparison, similar measures

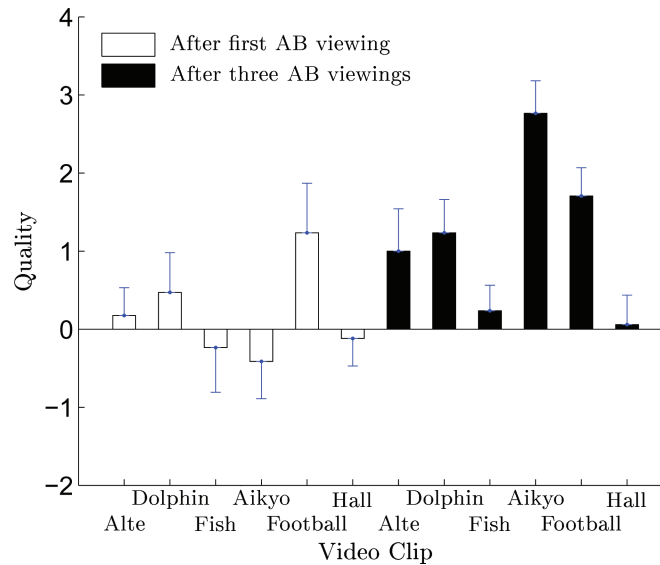


Fig. 8. Subjective votes reflecting the difference in quality between unfoveated and offline foveated versions of the same video clip. A value below zero indicates that the quality of the offline foveated video was judged as higher whereas the opposite is true for values larger than zero. Error bars span one standard error.

from the first experiment are given as bars with smaller widths.<sup>1</sup> As can be seen from the figure, the results are similar to those in the first experiment with the difference that the dispersions are significantly larger during quality evaluation. As was shown in Buswell [1935] and Yarbus [1967], certain task instructions can heavily influence where people look. One plausible explanation to the increase in dispersion is the active quality evaluation task given to the viewers. Such a task can make viewers look at regions where they would not look during free-viewing.

During first-time viewing, the dispersion during quality assessment is rather close to the baseline value (first-time free-viewing), and it can be assumed that subjects look at regions that remain in high quality after offline foveation. To address this assumption, Figure 8 compares the perceived quality of the six tested videos before and after offline foveation in addition to how the quality is affected by repeated viewings.

The white bars in Figure 8 show the average subjective quality of the videos after the first viewing. Error bars extend one standard error. A value larger than zero indicates that subjects preferred the quality of the unfoveated video. The opposite is true for quality values below the zero level. Offline foveation resulted in decreased quality in one of the tested videos, Football. The reason for this is most likely that eye-data used to implement offline foveation was slightly inaccurate temporally, such that foveation was performed with a slight lag in time. It is therefore no surprise that it is the video containing the fastest movements that gets a penalty in terms of subjective quality. The rest of the offline foveated videos were essentially indistinguishable from the unfoveated videos in terms of subjective quality. However, as a result of repeated viewings, subjects changed their viewing pattern and gazed directly at regions reduced in quality by offline foveation. The consequence of repeated viewings

<sup>1</sup>Using a centered letter (A or B) prior to video display showed to bias gaze positions toward the center of the display compared to Experiment I, where no such letter was used. However, this bias is present only until the initial saccade is launched and does not significantly influence the overall results.

in terms of subjective quality is illustrated by the black bars in Figure 8, where subjects strongly prefer the quality of the unfoveated versions. An interesting observation is the large change in subjective quality between the first and later viewings of Aikyo. Most likely, the facial region is such a strong visual attractor that it is initially hard to not gaze at. However, when looking outside the facial region, which is the case after repeated viewings, it is rather simple to see the introduced blurring effects.

## 5. DISCUSSION

Offline foveation removes high-frequency information from video regions where the majority of viewers do not look. Due to significant reductions of visual acuity in the peripheral vision, the removed frequencies are not necessary to get the impression of a high, uniform display quality. Consequently, offline foveation has the potential to significantly increase compression efficiency since smooth regions in general require fewer bits to represent digitally. We report on average compression gains of 20%, and individual gains of as much as 52% due to offline foveated video processing before H.264 encoding.

The 20% average compression gain should be regarded a lower bound of the expected increase in compression efficiency due to the following reasons. First, H.264 is in no way optimized to encode offline foveated video representations. Second, the degree of peripheral blurring is experimentally tuned, and it is therefore not clear how much additional blurring could be introduced without affecting the subjective quality negatively. Finally, the tested videos were assessed in a lab environment and presented without sound. It is likely that a setup allowing for a more engaging viewing experience would yield an even more coherent viewing pattern across subjects. Moreover, since offline foveation reduces the entropy of a video, and since it is independent of the video coder, offline foveation has the potential to improve the compression efficiency using other codecs than H.264.

With the substantial bitrate reduction due to offline foveation in mind, the main focus of this article was to investigate the perceptual effects offline foveation has on viewers during free-viewing and quality evaluation. The effects were quantified by measuring peoples' eye movement behavior and experienced quality before and after offline foveation of six videos in terms of gaze positions, fixation duration, saccade amplitude, repeated viewings, and subjective quality. In favor of successful applicability of offline foveated compression, results showed that offline foveation had no or very slight effects on these perceptual measures during first-time viewing. However, both gaze locations and subjective quality were affected as a result of multiple viewings.

The results showed that traditional methods for video quality assessment are not directly applicable to offline foveated video. In standards outlined in, for instance, VQEG [2003], it is advised to show the video to be assessed several times to the subject before the quality vote is given. However, as seen in this article, repeated viewings make subjects' viewing behavior deviate from normal, first-time free-viewing, thus shifting visual attention toward regions where viewers normally would not look. In view of this, traditional methods would all give very poor results in judged quality for offline foveated videos.

Today, there are some practical issues making it cumbersome to effectively utilize offline foveated systems for video compression. The one met with most skepticism is that eye-tracking recordings require expensive equipment and are time consuming, and therefore would be a bottleneck in a "real-world" application. We see two future solutions to this problem. First, it is by many envisioned that eye-trackers will be embedded in Web cameras, and that other low-cost, simple-to-use eye-tracking equipment will be available for practical use in a near future. Already today, such systems have been implemented [Pedersen and Spivey 2006]. This would make eye-tracking recordings, today most often taking place in laboratory environments, more autonomous and less time consuming, since individual



viewers themselves could download videos and record gaze positions through self-paced experiments. Second, there is no doubt that offline foveation would greatly benefit from algorithms that automatically and accurately predict where subjects will look, given only the raw video as input. Such algorithms would increase the practical usability of offline foveated compression since eye-tracking collections with human observers would be unnecessary. The development of such algorithms presents a great challenge for future research, which not only need to model the physical, bottom-up properties of the input video but also inherently complex top-down features that influence subjects' viewing strategies. To the authors' knowledge, no automatic algorithm has been used to implement offline video foveation and showed that it does not decrease viewers' subjective quality, although popular algorithms for gaze prediction exist [Itti 2004]. Based upon the results in this article, we welcome an intensified research on automatic gaze prediction for video compression.

We predict that offline foveation will be an interesting technology for future applications in video communications. In particular, it would be beneficial in bandwidth-constrained applications such as wireless communications in mobile devices, and for video streamed over Internet. For example, prioritizing regions with high gaze density can be useful to facilitate interpretation, recognition, and subjective quality of image and video data, especially at low bitrates.

#### ACKNOWLEDGMENTS

We would like to thank the eye-tracking seminar group at Lund Humanities Lab for valuable feedback on an earlier version of this manuscript. The test subjects are greatly acknowledged for participating in the data collections.

#### REFERENCES

- BERGSTRÖM, P. 2003. Doctoral dissertation. Ph.D. thesis, Linköping University, Sweden.
- BUSWELL, G. 1935. *How People Look at Pictures*. University of Chicago Press, Chicago, IL.
- DUCHOWSKI, A. 2000. Acuity matching resolution degradation through wavelet coefficient scaling. *IEEE Trans. Image Process.* 9, 8, 1437–1440.
- DUCHOWSKI, A. AND MCCORMICK, B. 1998. Gaze-Contingent video resolution degradation. In *Human Vision and Electronic Imaging*, B. Rogowitz and T. Pappas, Eds. Vol. 3299. 318–329.
- GEISLER, W. AND PERRY, J. 1998. A real-time foveated multi-resolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging*, B. Rogowitz and T. Pappas, Eds. Vol. 3299. 294–305.
- GEISLER, W. AND PERRY, J. 1999. Variable-Resolution displays for visual communication and simulation. *Soc. Inf. Display* 30, 1, 420–423.
- HENDERSON, J. 2003. Human gaze control during real-world scene perception. *Trends Cogn. Sci.* 7, 11, 498–504.
- ITTI, L. 2004. Automatic foveation for video compression using a neurobiological model for visual attention. *IEEE Trans. Image Process.* 13, 10, 1304–1318.
- JOHANNESSON, E. 2005. Master's thesis, Lund University.
- KORTUM, P. AND GEISLER, W. 1996. Implementation of a foveated image coding system for image bandwidth reduction. In *Human Vision and Electronic Imaging*, B. Rogowitz and J. Allebach, Eds. Vol. 2657. SPIE, San Jose, CA, 350–360.
- LE MEUR, O., LE CALLET, P., AND BARBA, D. 2007. Predicting visual fixations on video based on low-level visual features. *Vis. Res.* 46, 2483–2498.
- NYSTRÖM, M. AND HOLMQVIST, K. 2007. Deriving and evaluating eye-tracking controlled volumes of interest for variable-resolution video compression. *J. Electron. Imag.* 16, 1, 013006.
- NYSTRÖM, M. AND HOLMQVIST, K. 2008. Semantic override of low-level features in image viewing—Both initially and overall. *J. Eye Movem. Res.* 2, 2, 2:1–2:11.
- NYSTRÖM, M., NOVAK, M., AND HOLMQVIST, K. 2004. A novel approach to image coding using off-line foveation controlled by multiple eye-tracking measurements. In *Proceedings of the Picture Coding Symposium*.
- PARKHURST, D., LAW, K., AND NIEBUR, E. 2002. Modeling the role of salience in the allocation of overt visual attention. *Vis. Res.* 42, 107–123.

- PEDERSEN, B. AND SPIVEY, M. 2006. Offline tracking of eyes and more with a simple Webcam. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*.
- RAJASHEKAR, U., CORMACK, L., AND BOVIK, A. 2004. Point of gaze analysis reveals visual search strategies. In *Human Vision and Electronic Imaging*. Vol. 5292. 296–306.
- SHEIKH, H. R., LIU, S., EVANS, B. L., AND BOVIK, A. C. 2001. Real-Time foveation techniques for h.263 video encoding in software. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. 3. 1781–1784.
- STELMACH, L. B. AND TAM, W. 1994. Processing image sequences based on eye movements. In *Human Vision, Visual Processing and Digital Display*. Vol. 2179. 90–98.
- TATLER, B., BADDELEY, R., AND GILCHRIST, I. 2005. Visual correlates of fixation selection: Effects of scale and time. *Vis. Res.* 45, 5, 643–659.
- VQEG. 2003. Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II. [http://www.its.bldrdoc.gov/vqeg/projects/frtv\\_phaseII/downloads/VQEGII\\_Final\\_Report.pdf](http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII/downloads/VQEGII_Final_Report.pdf).
- WANG, Z. AND BOVIK, A. 2001. Embedded foveation image coding. *IEEE Trans. Image Process.* 10, 10, 1397–1410.
- WANG, Z., LU, L., AND BOVIK, A. 2003. Foveation scalable video coding with automatic fixation selection. *IEEE Trans. Image Process.* 12, 2, 243–254.
- YARBUS, A. 1967. *Eye Movements and Vision*. Plenum Press, New York.

Received March 2008; revised September 2008; accepted October 2008