Single-Trial EEG Classification of Artifacts in Videos

MARYAM MUSTAFA, STEFAN GUTHE, and MARCUS MAGNOR, TU Braunschweig

In this article we use an ElectroEncephaloGraph (EEG) to explore the perception of artifacts that typically appear during rendering and determine the perceptual quality of a sequence of images. Although there is an emerging interest in using an EEG for image quality assessment, one of the main impediments to the use of an EEG is the very low Signal-to-Noise Ratio (SNR) which makes it exceedingly difficult to distinguish neural responses from noise. Traditionally, event-related potentials have been used for analysis of EEG data. However, they rely on averaging and so require a large number of participants and trials to get meaningful data. Also, due the the low SNR ERP's are not suited for single-trial classification.

We propose a novel wavelet-based approach for evaluating EEG signals which allows us to predict the perceived image quality from only a single trial. Our wavelet-based algorithm is able to filter the EEG data and remove noise, eliminating the need for many participants or many trials. With this approach it is possible to use data from only 10 electrode channels for single-trial classification and predict the presence of an artifact with an accuracy of 85%. We also show that it is possible to differentiate and classify a trial based on the exact type of artifact viewed. Our work is particularly useful for understanding how the human visual system responds to different types of degradations in images and videos. An understanding of the perception of typical image-based rendering artifacts forms the basis for the optimization of rendering and masking algorithms.

Categories and Subject Descriptors: I.3.6 [**Computer Graphics**]: Methodology and Techniques

General Terms: Human Factors

Additional Key Words and Phrases: Perception, rendering, perception of rendering artifacts, EEG, SVM, wavelets, human visual system

ACM Reference Format:

Mustafa, M., Guthe, S., and Magnor, M. 2012. Single-Trial EEG classification of artifacts in videos. ACM Trans. Appl. Percept. 9, 3, Article 12 (July 2012), 15 pages.

DOI = 10.1145/2325722.2325725 http://doi.acm.org/10.1145/2325722.2325725

1. INTRODUCTION

The exponential growth of user expectations for visually believable movies, games, and simulated environments is putting an increasing pressure on computing hardware and software to create highly complex but visually appealing and plausible imagery. However, this is becoming progressively difficult due to the evident limitations of computing hardware. Given the importance of visual fidelity in today's environment it is essential to understand and analyze how the Human Visual System (HVS) perceives rendering flows of complex, photo-realistic image sequences [Bartz et al. 2008]. This understanding will allow computer graphics practitioners to take advantage of the flexibility and robustness

This work was funded in part by ERC grant no. 256941 'Reality CG'.

Authors' addresses: M. Mustafa (corresponding author), S. Guthe, and M. Magnor, Institut für Computergraphik, TU Braunschweig, Mühlenpfordtstr. 23, D-38106 Braunschweig, Germany; email: mustafa@cg.cs.tu-bs.de.

DOI 10.1145/2325722.2325725 http://doi.acm.org/10.1145/2325722.2325725

associated with human vision. An image or video that is perceptually accurate is not necessarily also statistically accurate [McNamara et al. 2010]. The human eye tends to overlook many different types of artifacts and distortions within a visual stimulus. An understanding of the visual system's response to these distortions will allow an optimization of rendering systems and decrease the gap between hardware performance and required performance [O'Sullivan et al. 2004].

Despite a substantial amount of research in applying perception to graphics and IBR [Vangorp et al. 2011; Ferwerda et al. 1996; Anderson et al. 2011], only recently has there been a growing interest in using an EEG to analyze visual processing [Mustafa et al. 2012; Lindemann and Magnor 2011; Rossion and Caharel 2011; Agam and Sekuler 2007]. Also, the use of an EEG in computer graphics research is now much easier with the advent of systems like the Emotiv EPOC neuroheadset [Emotiv 2012]. Apart from being in expensive ($299) it requires little prior or specialized knowledge to use. This makes an EEG an excellent tool for the analysis of the perception of images and videos. However, because of the very low Signal-to-Noise Ratio (SNR) of EEG data the application of the EEG has been limited. Most of the related work in EEG use Event-Related Potential's (ERP's) [Agam and Sekuler 2007; Moulson et al. 2011] which, because they rely on averaging, require a large number of participants and a large number of trials per participant to get meaningful data. In contrast Lindemann et al. [2011] used Principal Component Analysis (PCA) to classify EEG data. However, this only partly solves the issues with the signal-to-noise ratio and is mainly used to speed up the classification by reducing the amount of data rather than improving the final classification result. We use a wavelet-based approach along with a standard support vector machine to achieve creditable single-trial classification results.

Our work focuses on the classification of artifacts that usually occur in IBR. According to Vangorp et al. [2011] the most common artifacts that occur are ghosting, blurring, and popping. In our test scenes we present videos containing these typical IBR artifacts to participants and then use the recorded EEG to determine the perceived quality. Usually the quality of a video or rendered output is determined either by user studies or the use of quality assessment algorithms. Typically the use of psychophysical experiments and user studies is limited because of the large number of participants required. Also, user studies can at best only measure the explicit output of the visual cognitive process [Korsar et al. 2003]. Quality ratings acquired through user studies are always filtered by some decision process which, in turn, may be influenced by the task and/or rating scale the participants are given [Ponomarenko et al. 2009]. Similarly the judgment of a user regarding a visual stimulus is often biased by external factors such as mood, expectation, or past experience. In contrast, our approach allows for the objective prediction of the perceptual quality of an image sequence from a single trial with a few participants and a few trials per participant.

In physically-based rendering, the quality can be defined as the difference of the output of a proposed algorithm against a ground-truth reference. However, as far as the perceived quality goes, things are not as straightforward. The two most important things are plausibility and absence of objectionable artifacts. In our earlier work [Mustafa et al. 2012] we looked into how the HVS reacts to artifacts in an image sequence but our analysis was limited only to ERP's which meant looking at averaged data only without attempting to classify single trials based on artifact presence. In this article, we use a wavelet-based algorithm for the classification of EEG signals for different artifacts versus ground-truth sequences and analyze the overall emotional response to the video. Our main contribution is a method for processing the EEG data using wavelets as proposed by Olkkonen et al. [2006] and then an SVM to classify a single trial based on the type of distortion. We show that for typical IBR artifacts in an image sequence it is possible to differentiate a single trial based on the type of the artifact viewed and determine the perceived quality of the video. We also show that it is possible to classify the trials based on the exact type of artifact viewed. Furthermore, we present evidence of a clear emotional response linked with each artifact.

2. RELATED WORK

IBR is a vast area and we will only be focusing on the relevant perceptually-based algorithms and related work in perceptual IBR. There has been recent interest in studying visual processing for image rendering and analysis techniques [Ferwerda et al. 1996; Vangorp et al. 2011]. However, most of the research is geared towards using perception-based algorithms to create rendered sequences or perceptual algorithms to determine the quality of the rendered output [Seshadrinathan and Bovik 2010]. Most of the current work with EEG has been in the area of Human Computer Interaction (HCI). Shenoy and Tan [2008] present the idea of Human-Aided Computing which uses an EEG to label images implicitly. They use brain processes to show that users can implicitly categorize pictures based on content. Their work, however, required users to memorize the images and to be attentive to the content viewed. The most relevant work is of Vangorp et al. [2011] who conduct psychophysical experiments to understand the perception of artifacts in rendering of facades. They looked at the user feedback from rendered sequences that moved over facades of buildings. We use their work to decide the kind of artifacts to look into. However, our work is focused on using an EEG to measure the actual perception of these artifacts in the primary visual cortex and to then use these measurements along with wavelets and an SVM to classify the videos based on the type of artifacts present.

2.1 Perception-Based Rendering Algorithms

In 2001, McNamara [2001] already looked into the idea of including a perceptual model into a rendering pipeline. That author employed a model based on aspects of the human visual system as this portion of the process of perception is well understood. However, the author mentions that perception overall is a much more complex process that requires more research in the future.

In the same year, Luebcke and Hallen [2001] used an approximation to an empirical perceptual model for real-time rendering. They used a point-based rendering system called QSplat [Rusinkiewicz and Levoy 2000] that constructed a point-cloud hierarchy over a given model. The point cloud was then used during rendering as a multiresolution approximation of the underlying geometry. The perceptual model was combined with gaze tracking to produce a detailed map that defined the required rendering precision.

As a conservative estimation of the perceptual quality, Farrugia and Péroche [2004] use information from the human visual system, to define when an approximated image is indistinguishable from its original. Even though their approximated images are perceptually of the same quality, they miss out on further optimizations due to the remaining perceptual processing.

As part of a 2010 Siggraph Course by Křivánek et al. [2010] on ray tracing solutions for film production rendering, Fajardo gave a very nice example of a case where all prior metrics would fail to some extent. In order to reduce the noise in indirect illumination, all specular lobes are widened in secondary bounces. This leads to a convincing looking image without any visual artifacts that accurately conveys the overall lighting situation. However, it is very well distinguishable compared against a ground-truth reference.

2.2 Perception and EEG

There is a growing interest in using EEG for the analysis of Human Visual Perception. Recently Moulson et al. [2011] analyzed the perception of faces using an EEG. They looked at the N170 component of an ERP using a traditional component analysis and single-trial classification. The authors use statistical classifiers to decide if the temporally distributed pattern of activity in reaction to faces was different from that elicited by non-faces on trial by trial bases and if these patterns of activity differed among non-faces that varied in how face-like they were. The results showed that both analysis showed strict preference for veridical face stimuli within the N170 time window.

Similarly, Rossion et al. [2011] used an EEG to look into how fast visual stimuli are classified as faces by the brain. They used ERP's to show a dissociation between the ERP component P1, which reflects low-level visual cues, as opposed to component N170, which is in response to the perception of a face regardless of low-level visual cues like color.

Recently Lindemann and Magnor [2011] and Lindemann et al. [2011] have been using an EEG to assess the quality of compressed images and video artifacts. They reported that when shown different images of decreasing quality the participants' EEG results showed corresponding changes in image quality. Their work showed that the brain response varied with the image compression value. However Lindemann looked into static images [Lindemann and Magnor 2011] and later [Lindemann et al. 2011] static images they zoomed into. We wanted to look into ways of predicting video quality of complicated realistic image sequences with motion and natural scenes.

The most relevant research has been done by Mustafa et al. [2012] who looked at artifacts in videos and the corresponding EEG results. They showed the obvious brain response to artifacts in videos in the form of Event-Related Potentials (ERP's). Our work is in part based on this paper, however, we concentrate on using this EEG to classify a single-trial EEG into different categories based on the level of distortion. There is little related research into using an EEG to determine quality of a rendered output and to categorize the visual stimulus based on the exact artifact present.

2.3 Wavelet-Based Classification

In recent years, wavelet-based, and especially shift-invariant, otherwise known as complex wavelet-based analysis, has become more popular in the context of EEG or brain wave data. However, most of the publications in this area are either focusing strong abnormalities [Subasi et al. 2005] or more invasive brain wave recording than EEG [Olkkonen et al. 2006].

Olkkonen et al. [2006] were the first to apply a complex wavelet transform for filtering EEG data. They used a separate Hilbert transform in Fourier space, therefore guaranteeing true shift-invariance. In order to avoid the required Fourier transformations the authors also proposed to use a discrete version of the Hilbert transform as defined by Oppenheim et al. [1999].

With the advent of lifting steps to create second-generation wavelets [Sweldens and Schroder 2000], a different construction of the complex wavelet transform became possible. Barria et al. [2012] show that the resulting dual-tree wavelet transformation almost forms a Hilbert pair. Since we would like to achieve the highest classification accuracy rather than optimal running time, we chose to stay with the separate Hilbert transform and the lifting algorithm for final transform only.

All final classifications require some kind of either Support Vector Machine (SVM) or neural network. As SVMs are very well established in this field, we chose a multiclass support vector machine (C-SVM) with Radial Basis Functions (RBF) as kernel function. The SVM we use throughout this article is freely available from the authors [Chang and Lin 2011].

3. ARTIFACT CLASSIFICATION

The most straightforward way to classify single-trial EEG data is using it in its raw form. While this is useful for ERP's where a large number of trials are averaged, that is, at least 10 participants with multiple trials each, the low signal-to-noise ratio and the overall amount of noise makes this approach less than ideal for the single-trial setting.

Traditionally, the variants of the discrete Fourier transform have been used as the brain activity is limited to certain frequency ranges and most of the frequencies outside of these ranges can be regarded as noise. However, since the Fourier transform loses all temporal information outside a single phase shift per frequency, it cannot be used directly. Instead, the windowed Fourier transform is used to

regain some of the temporal resolution. Still, a high-frequency and low temporal resolution causes issues if a captured brain wave rapidly changes its frequency even by small amounts.

In contrast to the discrete Fourier transform, the Discrete Wavelet Transform (DWT) has a much lower-frequency resolution but the temporal resolution adapts to the frequency, that is, the temporal resolution is directly proportional to the frequency. Since each frequency range we're interested in roughly covers one frequency band of the wavelet transform, it seems to be the ideal choice for us. However, a regular wavelet transform is not shift-invariant and will therefore have issues with phase shifts. A Complex Discrete Wavelet Transform (CDWT), on the other hand, can easily be made shift-invariant as we will see shortly.

When analyzing EEG data from face and object recognition, Rousselet et al. [2007] found that the 5Hz to 15Hz range produced the best results in their setting. However, we found that using the range from 2.5Hz to 20Hz increases the classification accuracy compared to 5Hz to 20Hz (we can't use 5Hz to 15Hz as we are limited to multiples of 2 because of the wavelet transform). This can be explained by the slightly less than perfect frequency cut-off of the wavelet filter functions (see Figure 2) and the fact that a lower-frequency phase shift shows up in higher-frequency bands.

3.1 Wavelet Transformation

Given a discrete input signal $f(t)$ and the wavelet filter pair consisting of a low-pass filter $g(t)$ and a high-pass filter $h(t)$, the general discrete wavelet transform is defined as follows.

$$\begin{aligned} s_0(t) &= f(t) \\ s_{n+1}(t) &= \sum_{k=-\infty}^{\infty} s_n(k)\, g(2t-k) \\ d_{n+1}(t) &= \sum_{k=-\infty}^{\infty} s_n(k)\, h(2t-k) \end{aligned}$$

In case of the Haar wavelet [Haar 1910], functions g and h form a orthonormal basis and are defined as follows.

$$\begin{aligned} g(t) &= \begin{cases} 1 & 0 \le t \le 2 \\ 0 & \text{otherwise} \end{cases} \\ h(t) &= \begin{cases} 1 & 0 \le t \le 1 \\ -1 & 1 \le t \le 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

In turn, this leads to the very simple definition of the Haar transform.

$$\begin{aligned} s_0(t) &= f(t) \\ d_{n+1}(t) &= s_n(2t+1) - s_n(2t) \\ s_{n+1}(t) &= \frac{1}{2} s_n(2t) + \frac{1}{2} s_n(2t+1) \end{aligned}$$

As can easily be seen, s_n is simply the average of two consecutive samples and d_n is the delta between these two. In order to easily construct higher-order wavelets, we use an approach called lifting [Sweldens and Schroder 2000] where the calculation of s_{n+1} is based on d_{n+1} as well. The Haar

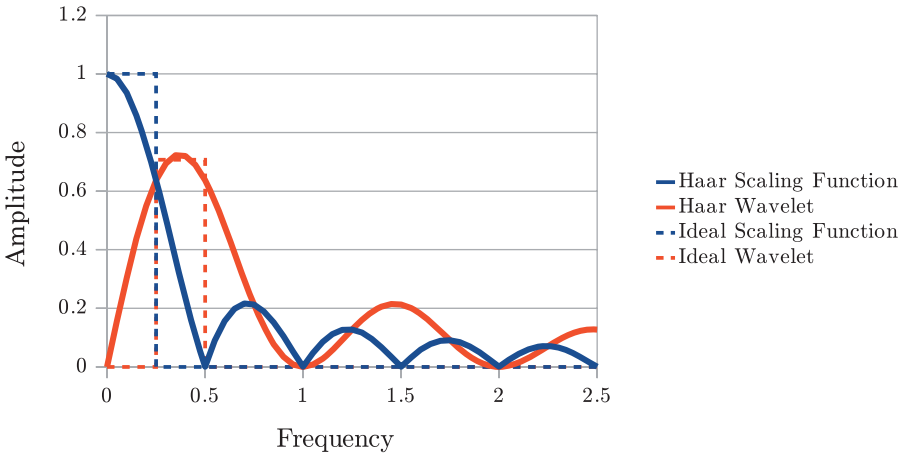

Fig. 1. Frequency response of Haar wavelet and scaling function compared against optimal frequency cut-off.

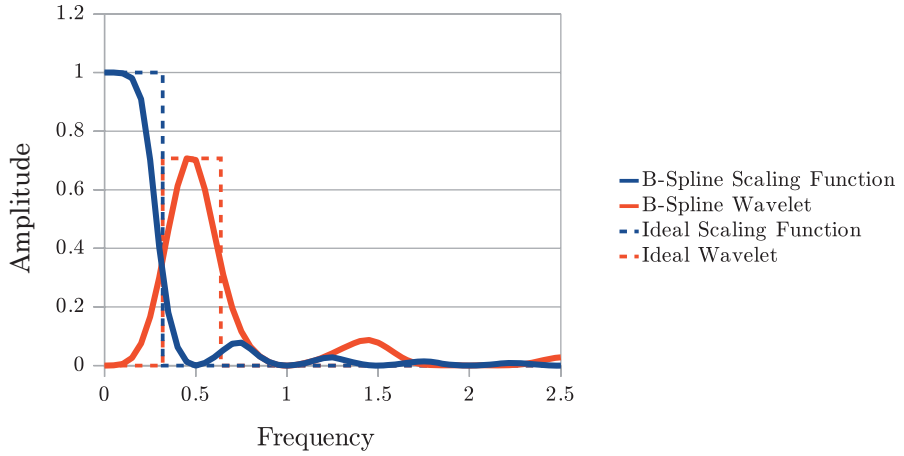

Fig. 2. Frequency response of cubic B-spline wavelet and scaling function compared against optimal frequency cut-off.

wavelet transform can then be written as follows.

$$d_{n+1}(t) = s_n(2t+1) - s_n(2t)$$
$$s_{n+1}(t) = s_n(2t) + \frac{1}{2}d_{n+1}(t)$$

Since we are interested in frequency ranges, however, we need to take a closer look at the frequency response of the Haar wavelet filter pair. As seen in Figure 1, for the Haar wavelet there is both a large frequency overlap between the wavelet and the scaling function. Furthermore, there are a significant amount of frequencies outside of the optimal ranges that will show up in the frequency bands.

In order to get both a better frequency cut-off and less overlap between wavelet and scaling function, higher-order wavelets, such as cubic B-spline wavelets (see Figure 2) can be used. Note that the cut-off frequencies can be different for each wavelets but they always follow the rule that the frequency doubles from one band to the next.

The lifting steps for the cubic B-spline wavelet transform are as follows.

$$\begin{aligned} d_{n+1}(t) &= s_n(2t+1) \\ &\quad -\frac{9}{16}s_n(2t+2) + \frac{1}{16}s_n(2t+4) \\ &\quad -\frac{9}{16}s_n(2t) + \frac{1}{16}s_n(2t-2) \\ s_{n+1}(t) &= s_n(2t) \\ &\quad +\frac{9}{32}d_{n+1}(t) - \frac{1}{32}d_{n+1}(t+1) \\ &\quad +\frac{9}{32}d_{n+1}(t-1) - \frac{1}{32}d_{n+1}(t-2) \end{aligned}$$

Unfortunately, the wavelet transform is not shift-invariant due to its down-sampling property, that is, the fact the each set of samples d_{n+1} contains only half the number of samples as d_n. However, the family of B-spline wavelets, as any discrete, symmetric filter, is linear in the phase of incoming frequencies which means that the filter has no phase distortion or constant group delay.

3.2 Shift-Invariant Transform

In order to create a shift-invariant wavelet transform, we either have to make our input signal or the actual transform shift-invariant in some sense. However, the easiest way to create a shift-invariant transform is making the input shift-invariant using an analytic function.

3.2.1 *Analytic Function.* The analytic function is defined as a complex function where the imaginary part is the same as the real, except that all frequencies have been shifted by 90 degree. Since the 90 degree shift is also linear in phase, the transformation from real to imaginary has constant group delay. The analytic function is defined using the Hilbert transform H as follows.

$$f_a(t) = f(t) + jH(f)(t)$$

If we assume that f^* consists of a single frequency, we get.

$$\begin{aligned} f^*(t) &= a\sin(\omega t + x) \\ f_a^*(t) &= a(\sin(\omega t + x) + j\cos(\omega t + x)) \end{aligned}$$

which leads us to the following for the absolute value of the analytic function of a single frequency.

$$\begin{aligned} \left\| f_a^*(t) \right\| &= \|a\| \sqrt{\sin^2(\omega t + x) + \cos^2(\omega t + x)} \\ &= \|a\| \end{aligned}$$

Since the wavelet transform has a linear phase, transforming the function f^* will only change its amplitude a, the frequency ω, and the phase x to a', ω' and x'. Furthermore, applying the wavelet transform on top of the Hilbert transform will produce the exact same result for a', ω' and x'. As the absolute value of the transformed function is now a' regardless of the initial phase, the whole transform is shift-invariant.

3.2.2 *Hilbert Transformation.* So far, we have treated the Hilbert transform as some kind of black box that causes a phase delay of 90 degrees. There are several ways to write down the Hilbert transform but one of it's continuous closed forms is as follows.

$$H(u)(t) = -\frac{1}{\pi} \lim_{\epsilon \downarrow 0} \int_{\epsilon}^{\infty} \frac{u(t+\tau) - u(t-\tau)}{\tau} \, \mathrm{d}\tau$$

In frequency space, the Hilbert transform is a phase shift by 90 (or $\frac{\pi}{2}$) degree. However, we seek to use a discrete Hilbert transform that does not require a Fourier transform of the EEG data. As the convolution is defined for continuous signals only, we first have to either reconstruct a continuous function from the EEG sample points or calculate discrete filter coefficients using some kind of weighted numeric integration. Since reconstructing a continous EEG signal might introduce unwanted frequencies, we use the discrete Hilbert transformation defined as follows.

$$H_{discrete}(u)(t) = -\frac{1}{2\pi} \sum_{\tau} \frac{u(t+\tau)}{\tau - \frac{1}{2}} + \frac{u(t+\tau)}{\tau + \frac{1}{2}}$$

Calculating a discrete Fourier transform on the preceding kernel shows that this is indeed the exact transform we require.

3.2.3 *Complex Wavelet Transformation.* Computing a separate Hilbert transformation prior to the actual wavelet transformation as in Olkkonen et al. [2006] allows us to use the same filter coefficients for both the real and the imaginary filters. At the same time, this approach fits our analysis framework best as it guarantees true shift-invariance (within the limits of the accuracy of the Hilbert transform).

In order to achieve a better frequency cut-off than Olkkonen et al. [2006], we use cubic interpolating spline wavelets with lifting [Sweldens and Schroder 2000] for both the real and imaginary portion of our analytic function as this produces the overall highest classification accuracy.

Assuming all input and output coefficients s_i, d_i to be complex numbers, the filter coefficients for the complex wavelet transform are equivalent to the filter coefficients of the regular wavelet transform. Thus, the lifting steps are the same as well.

3.3 Support Vector Machine Classification

Before applying the SVM, we remove any data outside the 2.5Hz–5Hz, 5Hz–10Hz, and 10Hz–20Hz frequency bands as additional frequency bands either contain noise only or mostly noise which leads the SVM astray.

We are using a standard support vector machine [Chang and Lin 2011] for all classification tasks. For the statistics, we performed a standard 5-fold cross-correlation test.

The data is split randomly into 5 groups of 288 trials. Using a C-SVM with a Radial Basis Function $e^{-g|x_i - x_j|^2}$ (RBF) classifier and a set of fixed parameters, the support vector machine is trained with data from 4 groups (976 trials) and tested against the trials in the remaining group (288 trials). This process is repeated until all trials have been classified. As proposed by Chang and Lin [2011], the process is repeated until the best set of parameters has been found.

4. EXPERIMENT

4.1 Participants

This experiment is based on our earlier work [Mustafa et al. 2012] and follows the same experimental setup. Eight (3 male, 5 female) healthy participants with an average age of 25 and with normal or

Fig. 3. Example of two artifacts shown in the videos, left: blurring on person, right: ghosting.

corrected-to-normal vision took part in the experiment. All participants had average experience with digital footage and no involvement in professional image/video rendering or editing.

4.2 Stimuli

The basic stimulus for the experiment was a 5.6 second video (resolution: 1440x1024, 30 fps) of a person walking along a park trail from left to right. The occurrence of the artifact was delayed by +/− 4 frames (+/− 132ms) to avoid locking the participants' attention to a fixed time. Five different kinds of artifacts were incorporated into the scene. These artifacts included both temporal and spatial aspects. The following 6 test cases were shown (Figure 3).

—Popping on Person (popP): a small rectangular area containing the walking person freezes for one frame.

—Popping: A static rectangular area of the image freezes for one frame.

—Blurring on person: a small rectangular area containing the walking person (left part of Figure 3) is blurred with a Gaussian kernel with a size of 15 pixels in 10 successive frames. The blurring area moves along with the motion of the person.

—Blurring: A static rectangular area in the center of the scene is blurred with a Gaussian kernel with a size of 15 pixels in 10 successive frames.

—Ghosting on Person: A partly transparent silhouette of the person stays behind for 10 frames, fading to invisibility in the last 5 frames (right part of Figure 3).

—Ground Truth (GT): No artifacts.

4.3 Procedure

One trial consisted of a ready screen followed by the video with artifacts which was instantly followed by the quality assessment screen. Participants were instructed to follow the moving person with their gaze and rate the quality of every test case on an integer 1 (worst) to 5 (best) Mean Opinion Score (MOS) scale [International Telecommunication Union 2006]. The participants were not informed about the presence of artifacts in the videos.

They were instructed orally and received a training in which each of the 6 videos was shown 3 times. This prepared them for the procedure and showed the whole range of available video qualities. During the main experiment all videos were shown 30 times resulting in 180 trials per participant. The videos were played in a blockwise randomized order and the same video was not shown twice in a row.

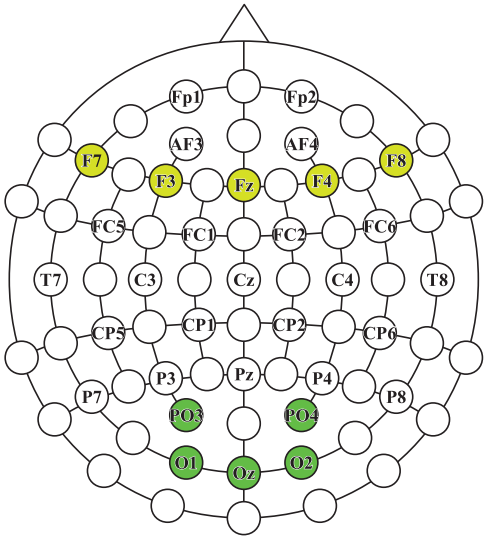

Fig. 4. EEG 32-electrode layout.

An EEG was recorded with the BioSemi Active Two system from 32 electrodes attached according to the international 10–20 system (Figure 4). Additionally a 4-channel EOG and mastoids were recorded which were used as a reference to remove data with accidental eye movements. The recorded data were referenced to the mastoids and filtered with a high-pass filter with a cutoff frequency of 0.1 Hz to remove DC-offset and drifts. Trials of a length of 1.2 seconds time locked to the appearance of the artifact occurrence were extracted from the continuous data. All trials with blinks, severe eye movements, and too many alpha waves were manually removed.

We assumed that the eye movements from watching the videos were the same for all participants given that there was only one moving object in the video which the participants were asked to follow.

5. RESULTS

Figure 5 shows the relative power increase over time for all artifacts averaged over all participants over all trials and over electrodes PO4, PO3,Oz, O1 and O2 (Figure 4) and as compared with ground truth (gt) with time 0 corresponding to the appearance of the artifact. We averaged over these electrodes as they correspond to the primary visual cortex in the brain and the areas that deal with motion. The EEG signal responding to the visual stimuli is strongest here.

Firstly, as can be seen clearly all artifacts were detected by the brain. The artifact which evoked the greatest response was “Popping on Person” (popP) which has the highest relative power and the least latency of response. This is followed closely by popping. Popping is a more obviously perceived artifact and evokes a quicker response and stronger response in comparison to popping not linked to motion. Ghosting as can be seen has the least response in terms of latency and relative power. It requires the brain to process the perceived distortion before a response occurs. This latency due to processing of the perceived stimuli is also seen with blurring, which is also a less obvious artifact. However, it is interesting to note that blurring linked to motion has a longer latency but a much higher response in terms of relative power increase. From the figures the difference in perception of artifacts related to motion as opposed to those independent of motion is clear. Both popping and blurring linked with the motion of the person produce a much larger response than popping and blurring not linked with the

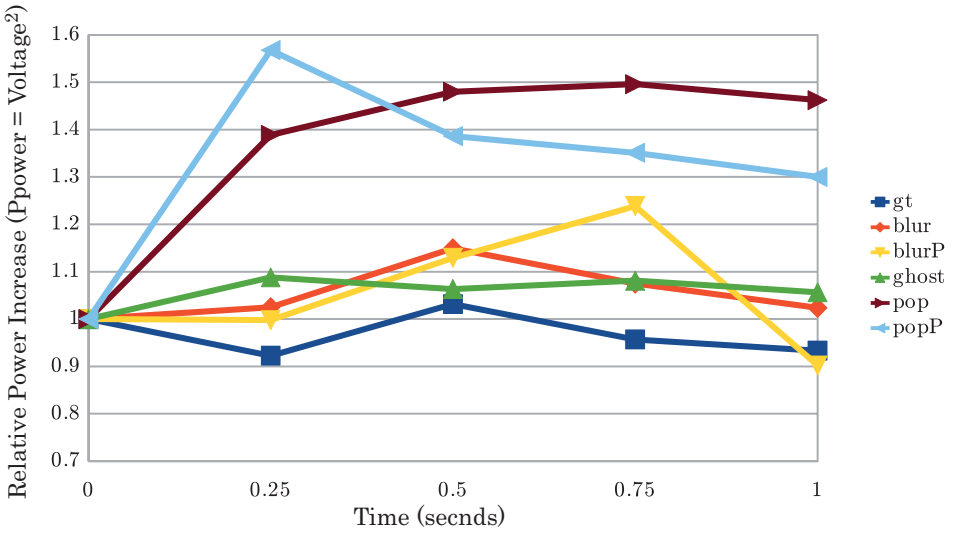

Fig. 5. Power increase for all artifacts in the 10Hz–20Hz range compared against ground truth(gt). The maximum neural response is for the artifacts popping and popping on person.

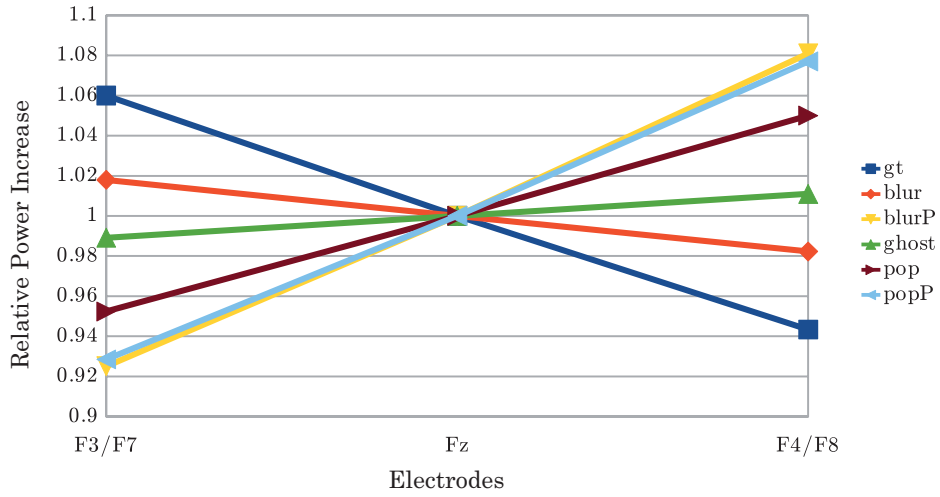

Fig. 6. Emotional response to artifacts in the 10Hz–20Hz range. F3/F7(averaged) are left frontal cotex and F4/F8(averaged) are right.

motion of the person. It is also clear that ghosting is the least perceived artifact evoking the smallest response.

5.1 Emotional Response

As can be seen from Figure 6 apart from a visual response there is a distinct emotional response to the artifacts as well. Previous data from EEG studies and emotion has provided evidence of lateralization of emotion in the frontal cortex [Korsar et al. 2003]. This theory predicts right hemisphere dominance for negative emotions. Figure 6 shows the response from electrodes F3/F7(averaged), Fz

Table I. Single-Trial Classification Accuracy Using Visual Processing and Emotional Data for Ground Truth with Correct Classifications on the Diagonal

	Trial Classified as	
Artifact	Ground Truth	Any Artifact
Ground Truth	**75%**	25%
Any Artifact	15%	**85%**

and F4/F8(averaged). As can be seen from Figure 4 these electrodes are located in the front of the head corresponding to the frontal cortex. F3/F7 are in the left frontal cortex and F4/F8 in the right. Fz is in the mid-line region. The results as can be seen from Figure 6 show an increased output in the right frontal cortex for test cases with more severe artifacts where the maximum output was for popping on person and blurring on person. This supports our conclusion that artifacts linked with motion not only evoke a larger visual response but are also emotionally more disturbing. This can theoretically be explained by the negative emotions elicited by bad video quality. It is also interesting to note as with the power response from the visual cortex ghosting is associated with a much smaller emotional response. Ground truth where there were no artifacts seems to evoke a positive emotional response as opposed to the negative responses from the videos with artifacts.

5.2 Wavelet-Based Classification

Given the statistical significance between ground truth and a given artifact in the EEG data [Mustafa et al. 2012], we were able to look at a total of three different classification tasks. The three classification categories we look at are as follows.

—First, we classify trials into one of two categories, trials with artifacts versus trials without artifacts.
—Second, we classify trials based on the severity of the artifact and look to only detect severe artifact, that is, popping and popping on person.
—Finally, we classify each trial based on the specific type of artifact. What artifact does a given trial contain?

We start by looking into only the response from the visual cortex and classifying trials based on that. However, it is interesting to note that as soon as we add the channels used for the emotional analysis, the accuracy of classification improves quite a bit.

5.2.1 *Ground-Truth Classification.* Just classifying the trials based on if there is an artifact present at all gives us an accuracy of 63% for just using the raw data. Using the wavelet transformed visual data increases the accuracy to 71%. Additionally using the emotional wavelet transformed data further increases the accuracy to 85% (see Table I). So for any given single trial from any given participant we can now determine whether an artifact was perceived or not. This allows us to determine the exact perceived quality of a visual stimulus. It is important to note that we train an SVM with just these two classes rather than using the same as for the per-artifact detection. As we use a differently trained SVM, we actually achieve a better accuracy for classifying ground-truth trials at the cost of classifying artifact trials.

5.2.2 *Severe Artifact Detection.* Instead of trying to classify for ground truth, we can also choose to find the most severe or objectionable artifacts. Starting with the raw data, we get an accuracy of already 75%. However, just using the wavelet transformed visual data increases the accuracy to 83%. Again, adding the wavelet transformed emotional data, we get a final accuracy of 94% (see Table II), leading us to the conclusion that severe artifacts can be reliably detected the easiest. Therefore with

Table II. Single-Trial Classification Accuracy Using Visual Processing and Emotional Data for Severe Artifacts with Correct Classifications on the Diagonal

	Trial Classified as	
Artifact	Ground Truth	Severe Artifact
Ground Truth	**95%**	5%
Severe Artifact	7%	**93%**

Table III. Single-Trial Classification Accuracy Using Visual Processing Data Only on a Per Artifact Basis with Correct Classifications on the Diagonal

	Trial Classified as					
Artifact	Ground Truth	Blurring	Blurring on P.	Ghosting	Popping	Popping on P.
Ground Truth	**50%**	15%	15%	12%	4%	4%
Blurring	10%	**54%**	15%	13%	3%	5%
Blurring on Person	8%	12%	**49%**	17%	5%	9%
Ghosting	17%	17%	18%	**35%**	6%	7%
Popping	7%	6%	4%	5%	**59%**	19%
Popping on Person	5%	5%	8%	4%	16%	**62%**

Table IV. Single-Trial Classification Accuracy Using Visual Processing and Emotional Data on a Per Artifact Basis with Correct Classifications on the Diagonal

	Trial Classified as					
Artifact	Ground Truth	Blurring	Blurring on P.	Ghosting	Popping	Popping on P.
Ground Truth	**63%**	10%	13%	11%	2%	1%
Blurring	11%	**68%**	13%	7%	0%	1%
Blurring on Person	9%	13%	**66%**	8%	1%	3%
Ghosting	17%	12%	18%	**49%**	2%	2%
Popping	3%	2%	3%	5%	**70%**	17%
Popping on Person	3%	2%	4%	2%	19%	**70%**

any single trial from any participant we can with an accuracy of 94% determine whether there was a severely perceived distortion in the rendered output. Note that this result was also gained by using an SVM that was trained for specifically recognizing severe artifacts rather than trying to distinguish between artifacts.

5.2.3 *Specific Artifact Classification.* We also looked into classifying trials based on the exact type of artifact appearing in the videos. Picking a random class for each trial would result in an expected accuracy of 16% so any resulting accuracy needs to be substantially better than this in order to claim a successful classification. Feeding all of the raw EEG curves into the SVM results in a classification accuracy of 39%. Using the wavelet transformed visual data only, we get a classification accuracy of 51% (see Table III). As can be seen from Table III, "Popping on Person" is the easiest artifact to classify and "Ghosting" the hardest. This is in keeping with the way these artifacts are percieved by the HVS as can also be seen from Figure 5.

Finally, using the wavelet transformed emotional data as well, we have a classification accuracy of 64% (see Table IV). As can be seen from Table IV we can now determine exactly which kind of artifact appeared in any given visual stimulus. So for any given single trial from any one participant we can determine the exact kind of artifact that was perceived by the viewer. As expected, classifying the ghosting artifacts is the worst scenario (still almost three times as good as random) whereas classifying the "Popping on Person" is the best one (with about 70% accuracy). This allows us to not only

determine the perceived quality of a rendered output but also determine the problems with the how it was perceived.

6. LIMITATIONS

While our current experimental setup provides new and relevant information it has some limitations. The main issue we see is the absence of eye movement information for more complicated test scenes. For our current video we could assume all participants were following the moving person since there was only one type of movement in the video. However, this becomes a problem with more complicated test scenes and for that we need to use an eye tracker. This would allow us to incorporate information regarding the exact viewing pattern of the participants during stimuli presentation. A more complete picture of participants' eye gaze pattern during stimuli presentation is essential for advances in realistic image and video synthesis. Also using sensors to capture physiological data would provide more concrete information regarding the participants' emotional state during trials.

7. CONCLUSION

Our work introduces a new method for the single-trial classification of typical IBR artifacts. We show that wavelets are an effective way to deal with the problem of low signal-to-noise ratios inherent in EEG signals. We also show that it is possible with a certain degree of accuracy to distinguish between different types of artifacts appearing in video stimuli. Our work analyzes the way the brain responds very differently to not only different types of artifacts but also to artifacts specifically linked with motion. Artifacts linked with motion evoke a much larger response in the brain. We also analyzed the effect that emotions play in the perception of distorted visual stimuli. Results of our work open up the possibility of shortening rendering times by eliminating computations that calculate image features which do not evoke a strong reaction in the brain as opposed to those which do. The brains response to artifacts is also essential for the modeling of masking algorithms for rendered image sequences.

REFERENCES

AGAM, Y. AND SEKULER, R. 2007. Interactions between working memory and visual perception: An ERP/EEG study. *NeuroImage 36,* 3, 933–942.

ANDERSON, E., POTTER, K., MATZEN, L., SHEPHERD, J., PRESTON, G., AND SILVA, C. 2011. A user study of visualization effectiveness using EEG and cognitive load. *Comput. Graph. Forum 30*, 791–800.

BARRIA, A., DOOMSA, A., AND SCHELKENSA, P. 2012. The near shift-invariance of the dual-tree complex wavelet transform revisited. *J. Math. Anal. Appl. 389*, 1303–1314.

BARTZ, D., CUNNINGHAM, D., FISCHER, J., AND WALLRAVEN, C. 2008. The role of perception for computer graphics. In *Eurographics State-of-the-Art-Reports*, 65–86.

CHANG, C.-C. AND LIN, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. 2*, 27:1–27:27.

EMOTIV. 2012. EPOC neuroheadset. http://www.emotiv.com/apps/epoc/297/.

FARRUGIA, J.-P. AND PÉROCHE, B. 2004. A progressive rendering algorithm using an adaptive perceptually based image metric. *Comput. Graph. Forum 23*, 605–614.

FERWERDA, A. J., SHILEY, P., PATTANAIK, N., AND GREENBERG, P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings of the ACM SIGGRAPH Conference*. 249–258.

HAAR, A. 1910. Zur Theorie der orthogonalen Funktionensysteme. *Math. Annal. 69,* 3, 331–371.

INTERNATIONAL TELECOMMUNICATION UNION. 2006. Mean opinion score (MOS) terminology. In ITU-T recommendation. P.800.1.

KORSAR, R., HEALEY, C. G., INTERRANTE, V., LAIDLAW, D. H., AND WARE, C. 2003. Thoughts on user studies: Why, how, and when. *Comput. Graph. Appl. 23,* 4, 20–25.

KŘIVÁNEK, J., FAJARDO, M., CHRISTENSEN, P. H., TABELLION, E., BUNNELL, M., LARSSON, D., AND KAPLANYAN, A. 2010. Global illumination across industries. In *ACM SIGGRAPH Courses*. ACM, New York.

LINDEMANN, L. AND MAGNOR., M. 2011. Assessing the quality of compressed images using EEG. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. 3170–3173.

LINDEMANN, L., WENGER, S., AND MANGNOR, M. 2011. Evaluation of video artifact perception using event-related potentials. In *Proceedings of the ACM Applied Perception in Computer Graphics and Visualization Conference (APGV)*. 1–5.

LUEBKE, D. AND HALLEN, B. 2001. Perceptually driven simplification for interactive rendering. http://www.cs.virginia.edu/~luebke/publications/pdf/egrw2001.pdf.

MCNAMARA, A. 2001. Visual perception in realistic image synthesis. *Comput. Graph. Forum 20,* 4, 211–224.

MCNAMARA, A., MANIA, K., BANKS, M., AND HEALEY, C. 2010. Perceptually-Motivated graphics, visualization and 3D displays. In *Proceedings of the ACM SIGGRAPH Conference*. 1–159.

MOULSON, M. C., BALAS, B., NELSON, C., AND SINHA, P. 2011. EEG correlates of categorical and graded face perception. *Neuropsychol. 49,* 14, 3847–3853.

MUSTAFA, M., LINDEMANN, L., AND MAGNOR, M. 2012. EEG analysis of implicit human visual perception. In *Proceedings of the ACM Human Factors in Computing Systems (CHI12)*.

OLKKONEN, H., PESOLA, P., OLKKONEN, J., AND ZHOU, H. 2006. Hilbert transform assisted complex wavelet transform for neuroelectric signal analysis. *J. Neurosci. Methods 151*, 106–113.

OPPENHEIM, A. V., SCHAFER, R. W., AND BUCK, J. R. 1999. *Discrete-Time Signal Processing* 2nd Ed. Prentice-Hall, Inc., Upper Saddle River, NJ.

O'SULLIVAN, C., HOWLETT, S., MORVAN, Y., MCDONNELL, R., AND O'CONOR, K. 2004. Perceptually adaptive graphics. In *Eurographics State-of-the-Art Reports*, 141–164.

PONOMARENKO, N., LUKIN, V., ZELENSKY, A., EGIAZARIAN, K., ASTOLA, J., CARLI, M., AND BATTISTI, F. 2009. A database for evaluation of full-reference visual quality assessment metrics. *Adv. Modern Radioelectron. 10,* 10, 30–45.

ROSSION, B. AND CAHAREL, S. 2011. ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Res 51, 12.*

ROUSSELET, G., HUSK, J., BENNETT, P., AND SEKULER, A. 2007. Single-trial EEG dynamics of object and face visual processing. *NeuroImage 36*, 843–862.

RUSINKIEWICZ, S. AND LEVOY, M. 2000. QSplat: A multiresolution point rendering system for large meshes. In *Proceedings of the ACM SIGGRAPH Conference*. 343–352.

SESHADRINATHAN, K. AND BOVIK, A. C. 2010. Motion tuned spatio-temporal quality assessment of natural videos. *Trans. Img. Proc. 19,* 2, 335–350.

SHENOY, P. AND TAN, D. 2008. Human-Aided computing: Utilizing implicit human processing to classify images. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. 845–854.

SUBASI, A., ALKAN, A., KOKLUKAYA, E., AND KIYMIK, M. K. 2005. Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing. *Neural Netw. 18,* 7, 985–997.

SWELDENS, W. AND SCHRODER, P. 2000. Building your own wavelets at home. *Comput.* 1995:5, 15–87.

VANGORP, P., CHAURASIA, G., LAFFONT, P.-Y., FLEMING, R. W., AND DRETTAKIS, G. 2011. Perception of visual artifacts in image-based rendering of facades. *Comput. Graphi. Forum 30*, 1241–1250.

Received May 2012; accepted June 2012