Published in the Proceedings of the 13th ACM International conference on Multimedia ACM MM 05
Singapore, Nov 6-1, 2005, pp. 796-797

Perceptual Media Compression for Multiple Viewers with Feedback Delay

Oleg Komogortsev
Kent State University
Computer Science Department
okomogor@cs.kent.edu

Javed Khan
Kent State University
Computer Science Department
javed@kent.edu

ABSTRACT
Human eyes have limited perception capabilities; for example, only 2 degrees of our 140 degree vision field provide the highest quality of perception. Due to this fact the idea of perceptual focus emerged to allow a visual content to be changed in a way that only part of the visual field where a human gaze is directed is encoded with a high quality. The image quality in the periphery can be reduced without a viewer noticing it. This compression approach allows a significant decrease in the number of bits required for image encoding, and in the case of the 3D image rendering, it decreases the computational burden. A number of previous researchers have investigated the topic of perceptual focus but only for a single viewer. In our research we investigate a dynamically changing multi-viewer scenario. In this type of scenario a number of people are watching the same visual content at the same time. Each person has his/her own perceptual focus area which changes over time. The visual content is sent through a network with a fixed delay/lag which provides an additional challenge to the whole scheme. The goal of our work was to investigate and develop a method of multi-viewer perceptual focus zones adaptation for real-time media perceptual compression and transmission. In our research we also look into the impact that such a method can have on transmission bandwidth and computational burden reduction.

Keywords
Perceptual compression, media adaptation.

1. IMPACT OF THE FEEDBACK DELAY
Feedback delay is the period of time between the instance the eye position is detected by an eye tracker (the device which identifies the current viewer's eye position) and the moment when a perceptually encoded frame is displayed. A typical network delay ranges from 20ms to a few seconds. Due to the rapid movement nature of a human eye, current eye position might change significantly by the time that information reaches the content adaptation system. This concern is important because future eye movements should fall within the highest quality region of an image/video. Only then would a viewer not be able to detect the image spatial degradation used for perceptual coding. Our research focuses specifically on the issue of containing the targeted amount of the viewers' eye gazes in a high quality image area given a value of the feedback delay. The task proved to be challenging when multiple viewers are involved.

2. SACCADE WINDOWING
In our previous work we proposed the concept of a Saccade Window (SW). A Saccade window is named for a type of eye movements called saccades – "rapid eye movements used in repositioning the fovea to a new location in the visual environment" [1]. A human's eye perceives the highest quality picture during an eye movement called a fixation – "eye movement which stabilizes the retina over a stationary object of interest" [1]. The purpose of the saccade window is to contain eye fixations by estimating an eye speed due to the saccades. The saccade window is calculated based on a set of past eye position samples, the current value of the feedback delay T_d, and the amount of eye-gazes required to be contained inside of the SW. The detailed explanation of the theory of the SW is described in [2]. A saccade window conceptually represents a zone of the future perceptual visual attention for the viewer it is built for. In a multi-viewer scenario each viewer has his/her own saccade window.

3. PERCEPTUAL VISUAL FIELDS
In our perceptual visual field assimilation design, we break a visual plane into several perceptual visual fields (PVFs). PVFs are designed in such a way that they represent zones of perceptual attention of several viewers. Given that we have V viewers watching the visual data that is being perceptually adapted, we build a Saccade Window $SW_i(t)$ for each viewer "i" on the visual frame "t". We define perceptual visual field $PVF_V(t)$ as a zone created by the union of the intersections of exactly V saccade windows on a visual frame F(t), perceptual visual field $PVF_{V-1}(t)$ is defined by a zone created by the union of the intersections of exactly V-1 saccade windows, perceptual visual field $PVF_{v-2}(t)$ is defined by a zone created by the union of the intersection of exactly V-2 saccade windows, etc. That way each perceptual visual field presents a perceptual attention area for m viewers. m changes from 1 to V. $PVF_0(t)$ is represented by a part of the video frame which is not covered by any saccade window ($SW_i(t)$). From this it is possible to see that there may be up to V+1 perceptual visual fields on each visual frame F(t).

$$PVF_m(t) = \bigcup \bigcap_{i=1}^{m} SW_i(t) \quad (1)$$

$$PVF_0(t) = F(t) - \bigcup_{i=1}^{V} SW_i(t) \quad (2)$$

4. EVALUATION PARAMETERS

We selected two parameters – *average eye-gaze containment* and *average perceptual coverage* to evaluate each created perceptual visual field. Average eye-gaze containment is calculated through the percentage of the eye-gazes from all viewers contained inside a particular PVF over N visual frames. Intuitively eye-gaze containment is a way of measuring the amount of the viewers' attention captured by a particular PVF. Average perceptual coverage is a percentage of the video image covered by a certain PVF over N video frames. Intuitively frame coverage represents the size of the perceptual focus/attention area for one or more viewers within a visual frame. Perceptual coverage also defines the size of the visual frame which requires the highest quality coding.

5. PERFORMANCE RESULTS

To evaluate constructed perceptual visual fields we selected five student volunteers with normal vision and selected three MPEG-2 video clips with various visual content and resolution of 720x480 pixels. Each video had a duration of 1 minute and a frame rate of 30fps. The video clips are available at our website [3]. Three feedback delay scenarios were tested: 166msec, 500msec and 1 sec. In the case of the small feedback delay scenario of 166ms, the amount of visual attention from all viewers was divided almost evenly between all PVFs - each PVF contained around 10%-25% of total eye-gazes, depending on the video clip. Frame coverage was small for all PVFs from 0.2-13%, except "PVF 0" 77-89%. In the case of 500msec delay scenario, "high" PVFs (those PVFs that were created by the majority of viewers) captured significantly more visual attention then "low" ones (PVFs created by one or two viewers) - "PVF 5" and "PVF 4" combined contained around 60% of total eye-gazes, while remaining PVFs contained around 40%. Frame coverage was small for all PVFs 2-18%, except "PVF 0" 52-75%. In case of the large feedback delay of 1 sec. the amount of attention captured by "high" PVFs was even larger then in the previous case - "PVF 5" and "PVF 4" contained around 80% or more. Frame coverage was larger for all PVFs in general 6-21%, except "PVF 0" 27-54%.

As a part of our experiment we wanted to find an optimal set of PVFs (OPVF) for each feedback delay scenario which would ensure a substantial amount of viewers' attention captured while maintaining low perceptual coverage by the optimal PVF set. In our experiments we set targeted gaze containment (TGC) to 90%, meaning that the amount of viewers' visual attention captured by the OPVF set should not go below 90%. The term OPVF was first introduced in our previous work [2]. We compared the performance of the OPVF set to a case when perceptual adaptation of the visual content is performed for each viewer individually without considering the attention information received from other viewers. We call this scenario Saccade Windows Union (USW). In that scenario TGC was set to 90% as well. For the 166 msec. delay case, OPVF and USW methods performed with the same coverage and containment results. For the 500 msec. delay scenario, coverage by OPVF was 1.6 times smaller than the coverage by USW, in the case of 1 sec. delay the coverage difference between OPVF and USW reached 2 times. In all delay scenarios the amount of attention captured by OPVF was equal to or higher than that of USW, while providing equal or smaller perceptual coverage. Thus OPVF performed equally or better than USW in all cases.

6. CONCLUSION

Perceptual methods can provide additional means of compression and computation burden reduction. One of the big concerns of perceptual media adaptation and transmission is the issue of the feedback delay. We conducted a series of experiments in a scenario where a video should be transmitted through a network with a specific feedback delay/lag value. Our experiment assumed that multiple people were watching transmitted video data at the same time. We introduced a concept of perceptual visual fields that assimilates attention areas created by a number of viewers. The results of our experiments show that people tend to look at the same parts of the image. An important aspect of the proposed approach is that it is media independent. Many of the point-gaze based researchers deeply integrate perceptual attention schemes with the media. In contrast, we proposed perceptual visual fields as virtual areas superimposed on the rendering plane of any visual media. Once the size and the location of the area which requires the highest quality coding is obtained, then the actual fovea-matched encoding can be performed in numerous media specific ways with various computational-effort/quality/rate trade-off efficiencies. Mapping of eye sensitivity to bit-allocation is a separate problem by its own merit. The actual bit/computational savings will depend on the specific coding/rendering model. In our research we particularly concentrated on the issue of the reduction of the area which requires highest quality coding or perceptual coverage and not peripheral degradation. Our results show that the optimal perceptual visual field set selection method created in our research gives up to 2 times the reduction in the size of the highest quality coded area, while maintaining gaze containment of that zone above 90%. The proposed scheme provides the best results in the case of high (500 msec. and higher) delays/lags in the perceptual adaptation and transmission systems.

The work had being funded by DARPA Research Grant F30602-99-1-0515.

7. REFERENCES

[1] Duchowski, A. T. *Eye Tracking Methology: Theory and Practice.* Springer-Verlag, London, UK, 2003.

[2] Oleg Komogortsev, Javed I. Khan, "Predictive Perceptual Compression for Real Time Video Communication", In *Proceedings of the ACM Multimedia 2004*, New York, Oct., 2004. pp220-227.

[3] Komogortsev, O., Khan, J., Perceptual Visual Field Assimilation Tests. At *www.cs.kent.edu/~okomogor/ACM05VideoSet.htm.*