Perceptual Attention Focus Prediction for Multiple Viewers in Case of Multimedia Perceptual Compression with Feedback Delay

Oleg Komogortsev
Kent State University
okomogor@cs.kent.edu

Javed Khan
Kent State University
okomogor@cs.kent.edu

Abstract

Human eyes have limited perception capabilities. Only 2 degrees of our 180 degree vision field provide the highest quality of perception. Due to this fact the idea of perceptual attention focus emerged to allow a visual content to be changed in a way that only part of the visual field where a human attention is directed to 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 bit-rate for a video stream, and in the case of the 3D stream rendering, it decreases the computational burden. A number of previous researchers have investigated the topic of real-time perceptual attention focus but only for a single viewer. In this paper 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 is using eye-tracking equipment. The visual content (video, 3D stream) is sent through a network with a large transmission delay. The area of the perceptual attention focus is predicted for the viewers to compensate for the delay value and identify the area of the image which requires highest quality coding.

CR Categories: I.6.4 [Simulation and Modeling]: Model Validation and Analysis; J.7 [Computers in Other Systems]: Process control, Real time.

Keywords: Perceptual attention prediction, compression, media adaptation.

1 Introduction

Human vision offers tremendous potential for perceptual data reduction. The diameter of the eye's highest acuity - the fovea extends only to 2 degrees. The parafovea (the next highest acuity zone) extends to about 4 to 5 degrees, and acuity drops off sharply beyond [Irwin 1992] that point. This idea is used in perceptual adaptation schemes to achieve an additional reduction in bit-rate and computational burden [Komogortsev and Khan 2004; Murphy and Duchowski 2001; Kortum and Geisler 1996; Lee et al. 2001]. Such adaptation methods increase the resolution and the image quality around an eye fixation point while reducing the image quality in the periphery field in accordance with the eye acuity function. A well-chosen acuity degradation function allows for the reduction of image quality on the periphery without a viewer noticing the degradative effect. Many researchers have investigated the eye acuity degradation from a fixation point [Kuyel 1998; Loschky and McConkie 2000], however in a situation where visual data is transmitted through a network with a high control loop delay or transmission lag, the acuity degradation becomes less important. In our previous work we developed a scheme for perceptual compression of the visual content for a single viewer for a case of a video transcoding system with a high delay [Komogortsev and Khan 2004]. The purpose of this paper is to investigate a case of perceptual video adaptation when multiple viewers are involved in the perception process. We designed our multi-viewer perceptual adaptation system for the situation where a visual media is perceptually transformed in real time and transmitted through a network with a large delay/lag. This delay gives an additional challenge to the perceptual adaptation system. It brings uncertainty into the system, due to the fact that the area of perceptual attention focus can change during transmission delay. In such a case the duration of the delay (particularly rendering and network delay) plays a critical role in affecting the size of the perceptual focus area, thus increasing the part of the image which requires highest quality coding. We tested our system with a range of delay values to better understand the impact these values have on the perceptual attention focus prediction and media compression. Practical implementation of this work can be applied to flight simulators, environment teleportation, virtual reality, telemedicine, remote vehicle operation, teleconferencing and any other scenario when multiple viewers are involved.

1.1 Previous Work

Many researchers have investigated perceptual video compression for a single viewer [Lee 2001; Geisler and Perry 1998; Wang et al. 2001; Westen et al. 1997]. Some scientists [Murphy and Duchowski 2001] have investigated the computational burden reduction that perceptual compression has to offer for 3-D model rendering. Several researchers performed a few multi-viewer experiments. Stelmach et al., [1991] performed an investigation in which twenty-four observers viewed 15, forty-five-second video clips while their direction of gaze was monitored. Video frames were divided on clusters and each of them was 6% of the total video frame. Dominant clusters contained between 78% of subjects' eye-gazes for fast motion videos and around 43% for slow motion videos. In the study

[Stelmach and Tam 1994] the test video sequences were perceptually pre-encoded. That scheme did not perform as well as was originally expected. Subjects were able to notice blurred areas on the images. The authors pointed out that it might be due to the fact that subjects subsequently looked at the different areas than the first time. Duchowski and McCormick [1998] looked at different Region of Interest placement strategies and perceptual degradation methods evaluated by number of viewers, for perceptually video encoding.

2   Human Visual system

Scientists have identified several intricate types of eye movements such as drift, saccade, fixation, smooth pursuit eye-movements, involuntary saccades. Among them, the following two play an important role in the design of the proposed system: (i) <u>Fixations:</u> - "eye movement which stabilizes the retina over a stationary object of interest" and (ii) <u>Saccades:</u> "rapid eye movements used in repositioning the fovea to a new location in the visual environment" [Duchowski 2003].

A visual sensitivity function which allows us to perform perceptual adaptation in a form of image degradation from a fixation point to the pheriphery is also called an eye contrast sensitivity function (CSF). Contrast sensitivity function can be described by the equation 2.1.1 (we adopted this function from the work of Daly and Ribas-Corbera [2001]). This function addresses the issue of entropy losses in the visual system. Also, this function accounts for the issues of cones, rods, and ganglion cells distributions.

$$S(x, y) = \frac{1}{1 + k_{ECC} \cdot \theta_E(x, y)} \qquad (2.1.1)$$

Here S is the visual sensitivity as function of image position (x,y), k_{ECC} is a constant (in this model k_{ECC}=0.24), and $\theta_E(x, y)$ is the eccentricity in the visual angle. Within any lossy compression scheme, an eye sensitivity function has to be mapped to the spatial degradation functions of a given encoding scheme.

3   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. This delay should be taken into consideration while using real-time perceptual media adaptation. This concern is important because future perceptual attention focus 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. It is noteworthy that the properties of visual media transmission might change over time thus increasing or decreasing feedback delay length. Typical network delays range from 20ms to a few seconds. Saccades can move the eye more than 10-100 degrees during that time, which has the potential of reducing the advantage of designing an accurate high acuity zone within the 2 degrees of an eye fixation point. Our research focuses specifically on the issue of containing targeted amount of the viewers' eye-gazes (perceptual attention focus) in a high quality image area given a value of the feedback delay. The task proved to be challenging when multiple viewers are watching same media content at the same time.

4   Perceptual Attention Field Assimilation Approach

4.1   Overview

The novel approach addresses the issue of predicting perceptual attention focus in a perceptual adaptation system with feedback delay/lag and a multi-viewer scenario. We propose a perceptual adaptation scheme which is based on a concept of a *Saccade Window* (W^{SW}). Saccade window predicts future perceptual attention focus area for a single viewer, which we also call a *Perceptual Attention Window* (W^{PAW}), through past eye speed behavior analysis. Perceptual attention window is a general concept that we use to predict a perceptual attention focus area for a viewer/s in a perceptual adaptation scheme with feedback delay.

Using individual saccade windows a set of perceptual attention fields is built which evaluates perceptual attention focus of multiple viewers.

Our system is implemented with an eye-tracker integrated full-logic MPEG-2 high-resolution region-based motion-vector reprocessing transcoder (which is drift free) [Khan et al. 2002].

4.2   Saccade Windowing

Saccade window is named for a type of eye movements described in section 2. A human's eye perceives highest quality picture during an eye movement called a fixation. Fixation presents the point of viewer's overt perceptual attention. The goal of the saccade window is to contain eye fixations by estimating the possible eye speed due to the saccades. Through the design saccade window compensates for the value of the feedback delay
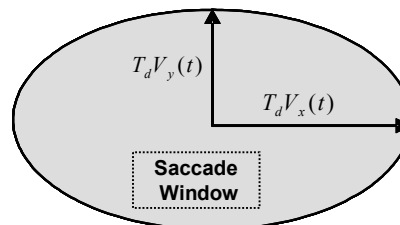

Figure 1. Saccade window diagram.

and predicts future perceptual attention focus for a viewer. The value of the feedback delay defines how far in the future such prediction will go. Mathematically the saccade window is calculated based on previous eye-speed behavior, current value of the feedback delay T_d and the amount of eye-gazes required to be contained inside of the window [Komogortsev and Khan 2004]. A SW is constructed as an ellipse created by the *gaze containment assured speeds* (CAS) $V_x(t)$ and $V_y(t)$ and the current feedback delay value T_d. The SW diagram is presented in Figure 1. Each CAS is calculated based on the previous eye-speed behavior. The history of previous eye-speed behavior is considered over "p" frames. In a multi-viewer scenario each viewer has his/her own saccade window. Every viewer's saccade window is updated for every frame through calculation of CAS for that particular viewer. For more information about SW construction and parameters consult [Komogortsev and Khan 2004].

4.3 Perceptual Attention Fields

The goal of *Perceptual Attention Fields* (PAF) design is to predict/identify the areas of the visual frame which will contain the most amount of attention when multiple viewers are involved. Given that we have V viewers watching the visual data that is being perceptually adapted, we build a Saccade Window $W^{SW_i}(t)$ for each viewer "i" on the visual frame "t". We define perceptual attention field $PAF_V(t)$ as a zone created by the intersection of exactly V saccade windows on a visual frame F(t), perceptual attention field $PAF_{V-1}(t)$ is defined by a zone(s) created by intersection of exactly V-1 saccade windows, perceptual attention field $PAF_{v-2}(t)$ is defined by a zone(s) created by the intersection of exactly V-2 saccade windows, etc. That way each perceptual attention field presents a perceptual attention area for m viewers. m changes from 1 to V. $PAF_0(t)$ is represented by a part of the video frame which is not covered by any saccade window ($W^{SW_i}(t)$). From this it is possible to see that there may be up to V+1 perceptual attention fields on each visual frame F(t).

$$PAF_m(t) = \bigcap_{i=1}^{m} W^{SW_i}(t) \qquad (4.3.1)$$

$$PAF_0(t) = F(t) - \bigcup_{i=1}^{V} W^{SW_i}(t) \qquad (4.3.2)$$

Non-zero perceptual attention fields are calculated by formula 4.3.1. PAF zero is calculated using 4.3.2. Perceptual attention fields are created for each frame and their construction is done in real-time.

The example diagram for the perceptual attention fields construction for five viewers is presented in Figure 2.

5 Experiment Setup

The test system was implemented using an integrated with MPEG-2 transcoder Applied Science Laboratories eye tracker model 504. The eye position camera captured eye coordinates at a rate of 60 samples per second. The ASL eye-tracker model compensates for slight head movements and has the following characteristics: accuracy - spatial error between true eye position and computed measurement is less than 1 degree, precision - better than 0.5 degree. We used 17 point calibration method to improve the accuracy of the experiment. After each run the accuracy of the calibration of each subject was checked and a viewer's calibration was corrected if needed.

As our test media we selected MPEG-2 encoded videos. Feedback delays of 166 msec., 500 msec., and 1 sec. were tested in our experiments. The resolution of the test videos was 704x480 pixels. The frame rate for each video was 30fps. The bit-rate was 10Mbs. Each video was displayed onto a projection screen in a darkened room. The projected physical dimensions of the image were the following: width 60 inches, height 50 inches. The distance between a viewer's eyes and the surface of the screen was about 100-120 inches. A group of five student volunteers was tested. All participants had normal or corrected to normal vision. A saccade window for each viewer was built using eye speed behavior analysis over all frames in a video sample. The *Target Gaze Containment* (TGC) parameter for each viewer was selected that on average saccade window contained 90% of that viewer's eye-gazes. Both RAS and TGC parameters were introduced in our previous work [Komogortsev and Khan 2004].

We selected three video clips with different content to evaluate our system's performance. Each video clip was 1 minute long. Bellow is a short description of each video clip.

Car: This video shows a moving car. It was taken from a security camera view point in the university's parking lot. The visible size of the car is approximately one fifth of the screen. The car moves slowly. Nothing in the background of this video distracts the subjects' attention. The snapshot is presented in Figure 3.

Shamu: This video captures a spotlighted, evening performance of Shamu at Sea World. This video consists of several moving objects: Shamu, the trainer, and the audience. Each of them is moving at different speeds during various periods of time. The snapshot is presented in Figure 2.

Airplanes: This video depicts a performance of the Blue Angels on Lake Erie. The flight formation of supersonic planes changes rapidly as does their flight speeds. The camera movements were rapid zoom and panning. The snapshot is presented in Figure 4.

The original video clips are available at our website [Komogortsev and Khan 2006].

6 Evaluation Parameters

We selected two parameters – *eye-gaze containmen*t and *perceptual coverage* to evaluate each created perceptual attention field.

6.1 Eye-gaze containment

We measure the amount of perceptual attention directed to a specific PAF or a particular Perceptual Attention Window by eye-gaze containment. Eye-gaze containment is represented by a percentage of the eye-gazes contained inside a specific area. To evaluate PAFs we defined quantity *average eye-gaze containment (AEGC)* as the fraction of gazes successfully contained within a specific PAF over N frames:

$$\xi_{PAF_i} = \frac{100}{N} \sum_{t=1}^{N} \frac{\left| E^{PAF_i(t)}(t) \right|}{|E(t)|} \qquad (6.1.1)$$

E(t) is the entire sample eye-gaze set from all viewers on the frame F(t) and $E^{PAF_i(t)}(t) \subseteq E(t)$ is the sample subset contained within perceptual attention field $PAF_i(t)$ for the frame F(t). N is the number of frames in a video sample.

6.2 Perceptual Coverage

We have defined a second performance quantity called *average perceptual coverage*. Average perceptual coverage is a percentage of the video image covered by a particular PAF. Average perceptual coverage is given by the equation (delta for area or volume):

$$\chi_{PAF_i} = \frac{100}{N} \sum_{t=1}^{N} \frac{\Delta(PAF_i(t) \cap F(t))}{\Delta F(t)} \qquad (6.1.2)$$

$\Delta F(t)$ is the area of the visual frame F(t). $\Delta PAF_i(t)$ is the area for perceptual attention field $PAF_i(t)$ for the frame F(t). N is the number of frames in a video sample.

7 PAF Performance results

7.1 Eye-gaze containment

One of the goals of our experiments was to measure the amount of perceptual attention directed towards different perceptual attention fields. Figures 5-7 present gaze containment performance results for various feedback delay values and test videos.

166 msec. delay: As we can see from Figure 5 the "PAF 5" created by intersection of all five viewers' Saccade Windows contained around 4%-11% of the total eye-gazes from all viewers. "PAF 4" contained 7%-25% of total eye-gazes. "PAF 3" contained 20%-27% of the eye-gazes. "PAF 2" contained 15%-44% of the eye-gazes. "PAF 1" contained 14%-21% of the eye-gazes. The remaining part of the video frame ("PAF 0") contained 7%-12% of viewers' eye-gazes.

500 msec. delay: Figure 6 shows the containment results. "PAF 5" contained 38%-42% of the eye-gazes. "PAF 4" contained 25%-29% of eye-gazes. "PAF 3" contained 16%-21% of the eye-gazes. "PAF 2" contained 7%-11% of the eye-gazes. "PAF 1" on average contained 3% of the eye-gazes. "PAF 0" contained 2%-4% of the eye-gazes.

Here we can see the tendency that the amount of contained eye-gazes is greatest for "high" perceptual attention fields and then the containment goes down. This tendency was true for all three video types.

1 sec. delay: Figure 7 shows the containment results. "PAF 5" contained 57%-73% of the eye-gazes. "PAF 4" contained 15%-26% of the eye-gazes. "PAF 3" contained 5%-9% of the eye-gazes. "PAF 2" contained 3%-4% of the eye-gazes. "PAF 1" on average contained around 1% of the eye-gazes. "PAF 0" contained 2%-3% of the eye-gazes.

The "highest" PAF contains a lot more gazes than sub-sequential ones.

7.2 Perceptual Coverage

Average perceptual coverage results for each perceptual attention field are presented in the Figures 8-10.

166 msec. delay: Figure 8. "PAF 5" covered less then 1% of the video frame. "PAF 4" covered 0.2%-2%. "PAF 3" covered around 1%-3%. "PAF 2" covered 3%-5%. "PAF 1" covered 7%-13%. "PAF 0" covered 77%-89%.

Frame coverage for "high" perceptual attention fields was quite small, with a slight decrease towards "PAF 1" and then with a big jump of the frame coverage for "PAF 0". All three test videos have performed in the same manner.

500 msec. delay: Figure 9. "PAF 5" covered 2%-4% of the video frame. "PAF 4" covered 2%-6%. "PAF 3" covered around 3%-10%. "PAF 2" covered 7%-11%. "PAF 1" covered 11%-18%. "PAF 0" covered 52%-75%.

Here, the frame containment graph was similar to the 166 msec. delay scenario, but the coverage in general was higher for all PAFs except "PAF 0".

1 sec. delay: Figure 10. "PAF 5" covered 12%-15% of the video frame. "PAF 4" covered 7%-10%. "PAF 3" covered around 6%-11%. "PAF 2" covered 6%-16%. "PAF 1" covered 11%-21%. "PAF 0" covered 27%-54%.

The "Car" and the "Airplanes" videos performed in a similar manner – large "PAF 0" coverage and much lower coverage values for other PAFs. This indicates that the majority of viewers concentrated on a specific area/object presented in these videos. The "Shamu" video coverage values distributed between PAFs more equally with "PAF 0" coverage being just slightly higher then the rest. This type of performance could have happened due to the fact that the "Shamu" video has more objects that viewers could have been interested in, and thus intersections between saccade windows created by the high delay covered larger areas.

8 Optimal PAF set

8.1 Construction

As a part of our experiment we wanted to find a specific set of PAFs for each feedback delay value which would provide targeted gaze containment (TGC) and predict a perceptual attention window for multiple viewers based on that. To calculate an optimal PAF (OPAF) set we selected TGC of 90%. This 90% value was chosen because we feel that it provides us with the best quality/compression ratio. Original saccade windows contained 90% of each viewer eye-gazes on average and thus the optimal PAF set should contain as much. The resulting PAF set we define as Perceptual Attention Window (W_{OPAF}^{PAW}) for multiple viewers.

To calculate the required number of PAFs to provide the needed TGC we used the following inequality:

$$TGC \le \sum_{i=V}^{k} \xi_{PVF_i} \qquad 8.1.1$$

"k" presents the highest sequential number of a PAF, which gaze containment brings summated average gaze containment to the targeted level. V represents total number of viewers. After "k" is found, the number of PAFs required to achieve the TGC, starting with the "highest" PAF, is calculated as N-k+1.

Formula for perceptual attention window construction using OPAF is:

$$W_{OPVF}^{PAW}(t) = \bigcup_{i=V}^{k} PVF_i(t) \qquad 8.1.2$$

The gaze containment formula for Perceptual Attention Window constructed by OPAF method is:

$$\xi_{OPVF} = \frac{100}{N} \sum_{t=1}^{N} \sum_{i=V}^{k} \xi_{PVF_i}(t) \qquad 8.1.3$$

The perceptual coverage formula for OPAF is:

$$\chi_{OPVF} = \frac{100}{N} \sum_{t=1}^{N} \frac{\Delta(\bigcup_{i=V}^{k} PVF_i(t))}{\Delta F(t)} \qquad 8.1.4$$

N in the last two formulas is the number of frames in a video sequence. V is the total number of viewers. k is calculated through solving 8.1.1.

8.2 Comparison

We decided to compare the gaze containment/coverage performance of the optimal PAF set to the the union of all viewers' saccade windows.

Saccade windows union (USW): The USW(t) calculated for each frame "t" is the union of all viewers' saccade windows for that frame. The performance under the USW scenario presents the case of when perceptual adaptation of the visual content is performed for each viewer individually without consideration of attention information from the other viewers.

Formula for perceptual attention window construction using OPAF is:

$$W_{USW}^{PAW}(t) = \bigcup_{i=1}^{V} W^{SW_i}(t) \qquad 8.2.1$$

where $W^{SW_i}(t)$ is the saccade window built for i-th viewer on the frame t.

The gaze containment formula for USW is:

$$\xi_{USW} = \frac{100}{N} \sum_{t=1}^{N} \sum_{i=1}^{V} \xi_{SW_i}(t) \qquad 8.2.2$$

where $\xi_{SW_i}(t)$ is the gaze containment for the saccade window $W^{SW_i}(t)$ built for i-th viewer on the frame t.

The perceptual coverage formula for USW is:

$$\chi_{USW} = \frac{100}{N} \sum_{t=1}^{N} \frac{\Delta(\bigcup_{i=1}^{V} W^{SW_i}(t))}{\Delta F(t)} \qquad 8.2.3$$

N in last two formulas is number of frames in a video sequence. V is the total number of viewers. Delta sign represents the area.

8.3 Performance results

Eye-gaze containment: Figure 11.

The gaze containment for the OPAF varied from around 90% to 95% depending on the delay. The gaze containment for the saccade windows union varied from around 90% to 98%, depending on the feedback delay value.

Perceptual coverage: Figure 12.

For the delay of 166 msec. there was almost no difference in coverage values between OPAF and USW. For the "Car" video the coverage for both methods was around 11%. For the "Shamu" video coverage was around 22%. For the "Airplanes" - 24%.

For the delay of 500 msec. OPAF had smaller coverage then for USW. For the "Car" video OPAF covered 14% and USW – 25%, that is 1.7 times decrease in coverage between these two

methods. For the "Shamu" video OPAF covered 30% and USW covered 48% - 1.6 times decrease in coverage. For the "Airplanes" video OPAF covered 23% and USW covered 36% - 1.6 times decrease in coverage.

For delay of 1 sec. for the "Car" video OPAF covered around 27% of the video frame and USW covered around 47% - 1.7 times decrease in coverage. For the "Shamu" video OPAF covered around 36% of the video frame and USW covered around 73% - 2 times decrease in coverage. For the "Airplanes" video OPAF covered around 28% of the video frame and USW covered around 46% - 1.6 times decrease in coverage.

As a conclusion we can say that the optimal PAF set selection method predicts the area of perceptual attention focus (perceptual attention window) for multiple viewers more efficiently than the method where the perceptual attention information from other participating viewers is not considered (USW method). OPAF method reduces the size of the area that requires high quality coding (perceptual attention window coverage) by 1.6-2 times, when compared to USF. At the same time the amount of perceptual attention captured by OPAF measured by gaze containment is just slightly smaller than the gaze containment for USW – less than 5%.

The actual amount of bandwidth reduction and computational burden reduction between two perceptual compression methods will depend on the two parameters: the size of the area which requires high quality coding (perceptual attention window coverage) and visual degradation of the periphery. Visual degradation of the periphery should match a human's eye acuity model to avoid visual discomfort to the viewers. We suggest that the acuity function presented in section 2 is to be used for that purpose. In this paper we concentrated on perceptual attention focus prediction and perceptual coverage reduction, which directly leads to the bandwidth and computation savings in any perceptual adaptation scheme which may involve multiple viewers and large feedback delay (transmission lag).

9 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 was 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 attention fields that predict/register attention areas created by a number of viewers. The results of our experiments show that people tend to pay attention towards same parts of the image. In our experiments we tested or scheme with five viewers, but it can be applied to any number of viewers. 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 attention 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 this paper 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 attention field set selection method proposed in this paper 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 case of high (500 msec. and higher) delays/lags in the perceptual adaptation and transmission systems.

10 Acknowledgments

References

IRWIN, D. E. 1992. Visual Memory Within and Across Fixations. In *Eye movements and Visual Cognition: Scene Preparation And Reading,* K. Rayner, Ed. Springer-Verlag, Springer Series in Neurophysiology, New-York, NY. 146-165.

KOMOGORTSEV O., KHAN J. 2004. Predictive Perceptual Compression for Real Time Video Communication. In *Proceedings of the 12th ACM International conference on Multimedia (ACM MM 04)*, 220-227.

MURPHY, H., AND DUCHOWSKI, A. T. 2001. Gaze-contingent level of detail rendering. In *EuroGraphics 2001*, EuroGraphics Association.

KORTUM, P., GEISLER, W. S. 1996. Implementation of a Foveated Image Coding System for Image Bandwidth Reduction. In *Proceedings of SPIE Vol. g657, Human Vision and Electronic Imaging,* 350-360.

LEE, S., PATTICHIS, M., BOVOK. 2001. A. Foveated Video Compression with Optimal Rate Control. In *IEEE Transaction of Image Processing, V. 10, n.7,* 977-992.

KUYEL, T., GEISLER, W. S., GHOSH, J. 1998. Retinally reconstructed images (RRIs): digital images having a resolution match with the human eye. In *Proceedings of SPIE Vol. 3299, Human Vision and Electronic Imaging,* 603-614.

LOSCHKY, L., MCCONKIE, G. 2000. User performance with gaze contingent multiresolutional displays. In *Proceedings of the symposium on Eye tracking research & applications,* 97-103.

KOMOGORTSEV, O., KHAN, J. 2006. Perceptual Attention Field Video Set. At *www.cs.kent.edu/~okomogor/ETRA06VideoSet.htm.*

DUCHOWSKI A. T. 2003. *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag, London, UK.

STELMACH, L. B., TAM, W. J., HEARTY, P. J. 1991. Static and dynamic spatial resolution in image coding: an investigation of eye movements. In *Proceedings of SPIE Vol. 1453, Human Vision, Visual Processing, and Digital Display II,* 147-152.

STELMACH, L. B., TAM, W. J. 1994. Processing image sequences based on eye movements. In *Proceedings of SPIE Vol. 2179, Human Vision, Visual Processing, and Digital Display V,* 90-98.

DALY S., MATTHEWS K., RIBAS-CORBERA J. 2001. As Plain as the Noise on Your Face: Adaptive Video Compression Using Face Detection and Visual Eccentricity Models. In *Journal of Electronic Imaging V. 10 (01),* 30-46.

KHAN J., YANG S., PATEL D., KOMOGORTSEV O., OH W., GUO Z., GU Q., MAIL P. 2002. Resource Adaptive Netcentric Systems on Active Network: a Self-Organizing Video Stream that Automorphs itself while in Transit Via a Quasi-Active Network. In *Proceedings of the Active Networks Conference and Exposition (DANCE '2002), IEEE Computer Society Press*, 409-426.

GEISLER, W. S., PERRY, J. S. 1998. Real-time Foveated Multiresolution System for Low-bandwidth Video Communication. In *Proceedings of SPIE Vol. 3299, Human Vision and Electronic Imaging III,* 294-305.

WANG, Z., LU, L., BOVIK, A. 2001. Rate scalable video coding using a foveation-based human visual system model. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP '01),* Vol. 3, 1785-1788.

WESTEN, S. J., LAGENDIJK, R., BIEMOND, J. 1997. Spatiotemporal model of human vision for digital video compression. In *Proceedings of SPIE Vol. 3016, Human Vision and Electronic Imaging II,* 260-268.

DUCHOWSKI, A. T., MCCORMICK B. H. 1998. Gaze-contingent video resolution degradation. In *Proceedings of SPIE Vol. 3299, Human Vision and Electronic Imaging III,* 318-329.
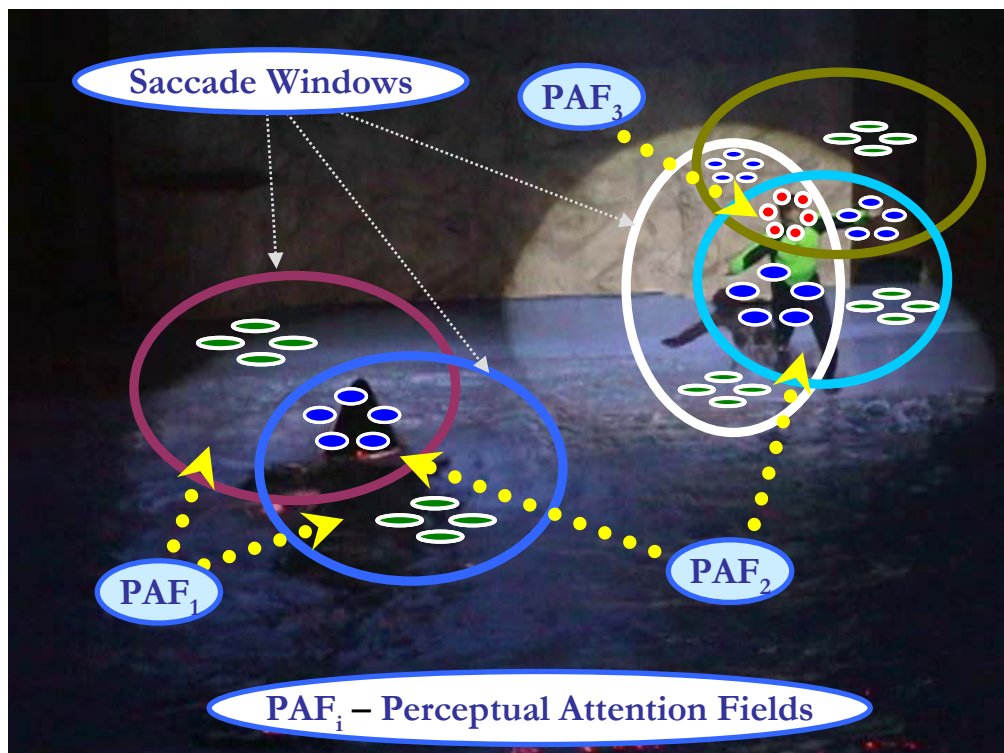
Figure 2: "Shamu" video. Diagram for perceptual attention fields construction. Five ellipses represent five saccade windows - one for each viewer. PAF_3 presented by small circles. PAF_2 by rhombi. PAF_1 by triangles. PAF_4 and PAF_5 do not exist for this frame. PAF_0 is a part of video frame not covered by any saccade window.
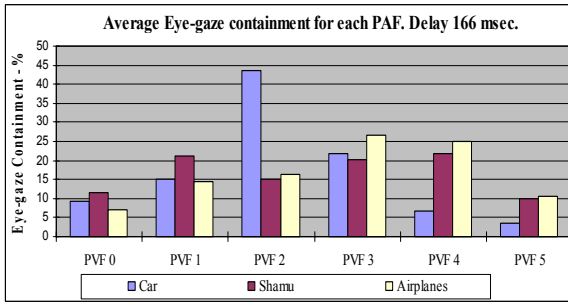

Figure 3. "Car" video.

Figure 4. "Airplanes" video.

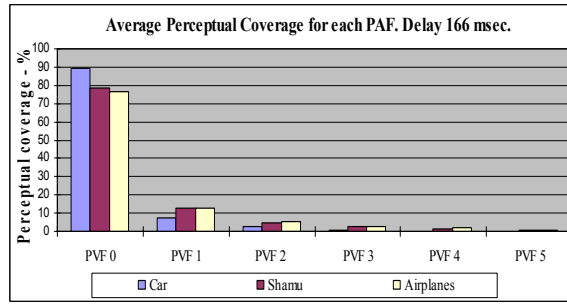Figure 5: Average eye-gaze containment for every perceptual attention field. Feedback delay is 166 msec.

Figure 8: Average perceptual coverage for every perceptual attention field. Feedback delay is 166 msec.
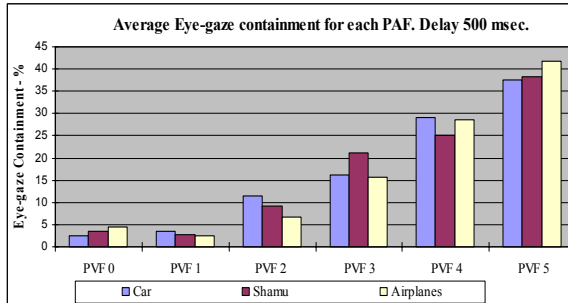
Figure 6: Average eye-gaze containment for every perceptual attention field. Feedback delay is 500 msec.
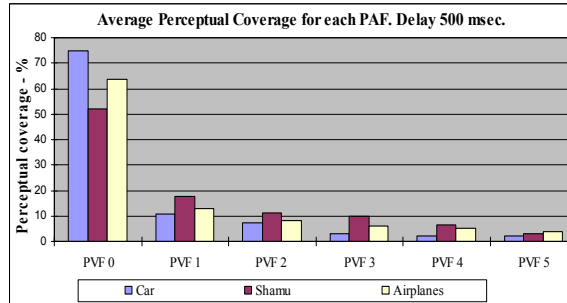
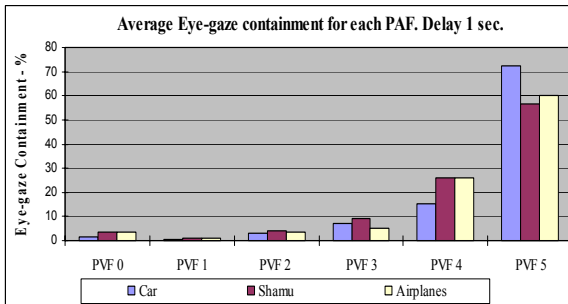Figure 9: Average perceptual coverage for every perceptual attention field. Feedback delay is 500 msec.

Figure 7: Average eye-gaze containment for every perceptual attention field. Feedback delay is 1 second.
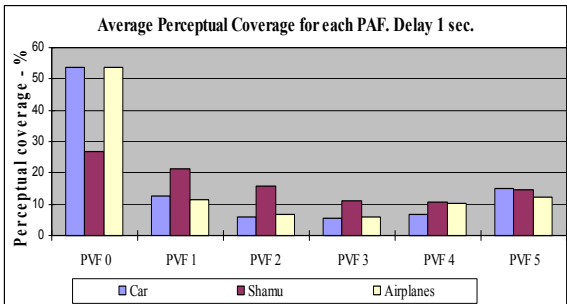

Figure 10: Average perceptual coverage for every perceptual attention field. Feedback delay is 1 second.
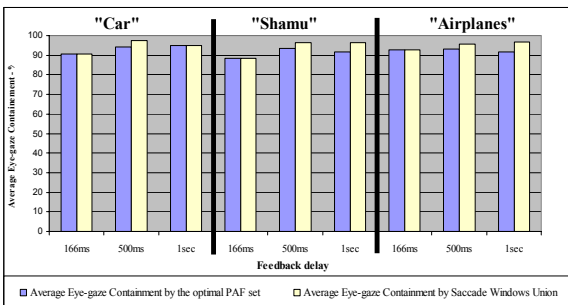

Figure 11: Average eye-gaze containments for different perceptual attention prediction methods. Delay/lag ranges from 166 msec to 1 sec.
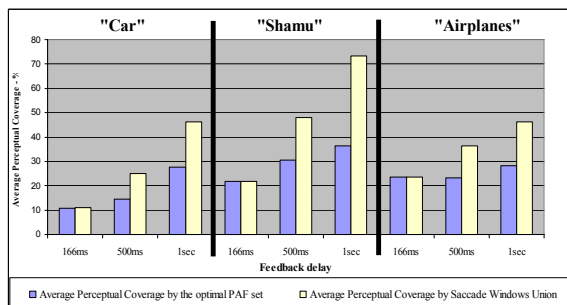

Figure 12: Average perceptual coverages for different perceptual attention prediction methods. Delay/lag ranges from 166 msec to 1 sec.