Region of Interest-based
Adaptive Multimedia Streaming Scheme

Gabriel-Miro Muntean, *Member, IEEE,* Gheorghita Ghinea, *Member, IEEE,* Timothy Noel Sheehan

Abstract—**Adaptive multimedia streaming aims at adjusting the transmitted content based on the available bandwidth such as losses that often severely affect the end-user perceived quality are minimised and consequently the transmission quality increases. Current solutions affect equally the whole viewing area of the multimedia frames, despite research showing that there are regions on which the viewers are more interested in than on others. This paper presents a novel region of interest-based adaptive scheme (ROIAS) for multimedia streaming that when performing transmission-related quality adjustments, selectively affects the quality of those regions of the image the viewers are the least interested in. As the quality of the regions the viewers are the most interested in will not change (or will involve little change), the proposed scheme provides higher overall end-user perceived quality than any of the existing adaptive solutions.**

Index Terms—**multimedia streaming, content adaptation, user perceived quality, region of interest.**

I. INTRODUCTION

A sustained growth in the number of broadband connections to residential users and business premises is expected, as part of a trend towards all-IP networks [1]. They provide a low cost, high bandwidth infrastructure that enables the distribution of rich content services, many of them multimedia-based such as digital and interactive video and audio – including video on demand and voice over IP. The success of this trend relies heavily on the users' quality of experience and on the price they have to pay. In order to reduce this price, the network operators strive to ensure a high utilisation of their infrastructure and simultaneously serve many customers. However, this reduces the quality of their services and is at odds with the operators' wish to deliver high quality streamed multimedia to their customers. Therefore there is a need for a solution to stream multimedia data – with its high bandwidth and stringent timing requirements - such that it accommodates these constraints. A typical architecture for multimedia streaming over IP networks is presented in Figure 1.

Adaptive solutions [2] were proposed for streaming

multimedia in the quest to maintain high end-user perceived quality even when loaded delivery conditions cause data loss that severely affects the quality of the transmission. In these conditions the adaptive schemes reduce in a controlled manner the amount of multimedia data to be streamed. By sending less data, the pressure on the delivery network is released and eventually the loss rate decreases, increasing the end-user perceived quality.

Existing solutions, though, affect equally the whole viewing area of multimedia frames when adjusting the multimedia stream. However research has shown [3] that there are regions of the multimedia display image on which the viewers are more interested in than on others. Consequently these multimedia display areas are named regions of interest (ROI). Figure 2 illustrates possible ROIs with highest user interest in a multimedia frame.

This paper proposes a novel region of interest-based adaptive scheme (ROIAS) for multimedia streaming that when performing adaptive transmission-related quantity and consequently quality adjustments, selectively affects the quality of those regions of the image the viewers are the least interested in. As the quality of the regions the viewers are the most interested in will not change (or will involve little change), the proposed scheme provides higher overall end-user perceived quality than any of the existing adaptive solutions that adjust equally the whole image area.

The structure of the paper is as follows: some existing adaptive multimedia streaming solutions are described next,
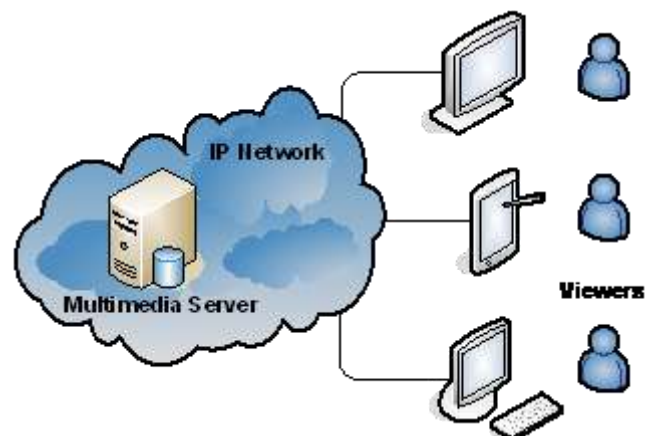
G.-M. Muntean is with the School of Electronic Engineering, Dublin City University, Ireland (phone: +353-1-700-7648; e-mail: munteang@eeng.dcu.ie).

G. Ghinea is with the School of Information Systems, Computing and Mathematics, Brunel University, United Kingdom (e-mail: george.ghinea@brunel.ac.uk).

T. N. Sheehan is with Eircom, Ireland (e-mail: sheehant@eircom.ie).

Fig. 1 Architecture for IP multimedia streaming

along with some ROI-based research. The presentation of the

Fig. 2. a) Original multimedia frame
b) Eye-tracking based regions of interest

state of the art in this area is followed by the description of the proposed region of interest-based adaptive multimedia streaming scheme. The model that deploys it and the simulation testing setup are then presented along with some simulation results. Subjective testing setup, method and scenario are presented next along with the results and a brief discussion of the results. At the end, conclusions are drawn and possibilities for future work are indicated.

II. RELATED WORK

A. Adaptive Multimedia Streaming Solutions

The large majority of the proposed adaptive schemes for multimedia streaming are **sender-based**, giving a significant role to the server in taking adjustment decisions, such as the *Loss-Delay based Adjustment algorithm (LDA)* proposed in [4]. It relies on RTCP reports and on a packet-pair technique to estimate round trip delay, loss rate and the bottleneck link bandwidth. The scheme controls the transmission rate using these estimates and some user parameters. The *enhanced Loss-Delay Adaptation algorithm (LDA+)* [5] also makes use of RTCP reports to collect loss and delay statistics and to adjust the transmission rate like TCP connections subject to equal losses and delays. The *Rate Adaptation Protocol (RAP)* proposed in [6] uses TCP-like acknowledgement of the packets to estimate loss rates and delays. In case of no loss, the rate is additively increased function of round trip delay, whereas in case zero loss, the rate is halved as TCP does. *Layered Quality Adaptation (LQA)* [7] is one of the most significant schemes that make use of the properties of layered-encoding in supporting rate-controlled adaptations. It modifies the bitrate and consequently the quality of the transmitted multimedia by adding and removing a layer respectively. In [8] a *TCP-Friendly Rate Control Protocol (TFRCP)* is presented, based on a TCP model. In case of losses, the rate is limited to the equivalent TCP rate computed according to the TCP model otherwise the rate is doubled. One of the most significant

sender-based adaptive schemes as it introduces an estimation of the end-user perceived quality in the adaptation loop is the *Quality Oriented Adaptation Scheme (QOAS)* [9].

The **receiver-based schemes** provide mechanisms that allow for the receivers to select the service quality and/or rate such as *Receiver-driven Layered Multicast (RLM)* [10] and *Receiver-driven Layered Congestion Control (RLC)* [11].

The *TCP Emulation At Receivers (TEAR)* scheme, described in details in [12] is a significant **hybrid adaptive mechanism** that involves both the sender and the receiver in the adaptation process. **Transcoder-based solutions** constitute another category. They focus on matching the available bandwidth of heterogeneous receivers through transcoding or filtering [13, 14].

All these schemes perform adjustments of the streamed multimedia data such as it affects equally all the frames' regions, regarding of viewer's interest. Unlike them the proposed region of interest-based adaptive scheme takes into account the differentiated importance these regions have for the end-user.

B. Region of Interest-based Research

The essential idea behind ROI video encoding is to display regions of the screen where the user is more likely to focus on with higher quality parameters than its surrounding areas, and considerable research has been done in the field. To this end, as well as providing a comprehensive introduction to the area of ROI video encoding (covering topics including medical applications of and perceptually lossless RoI video coders), [15] presents an elegant pixel shader algorithm showing that real-time ROI video processing can be used on both still image and video content. The implication is the proposed solution can be implemented for real-time video feeds, which would support the use of video in bandwidth constrained environments.

One of the key questions in ROI video encoding research is the issue of perceptual disruptions, and [16] considered the issue of continued perceptual disruptions in ROI video encoding – specifically examining perceptually acceptable update delays in multi-resolution displays. This research suggests that encoding video using ROIs, although useable, could introduce considerable perceptual distraction that can interrupt normal attentive processes. Moreover, in [17] it was found that when adapting a high-resolution window at the point-of-gaze and degraded resolution in peripheral areas, participants took longer to identify a visual target, than when a low resolution was uniformly displayed across the whole display window. This implies that use of degraded peripheral resolution can lead to longer search times and therefore impact user pre-attentive processes. In support of earlier studies [18], it was also found that, if degradation is increased in peripheral areas, the size of the adapted high-resolution window at the point of gaze also needs to be increased, if the users level of performance is to be maintained [19]. Sadly, this increase in the high-resolution window cancels out any bonus of peripheral degradation and limits any gain of using a RoI video coder.
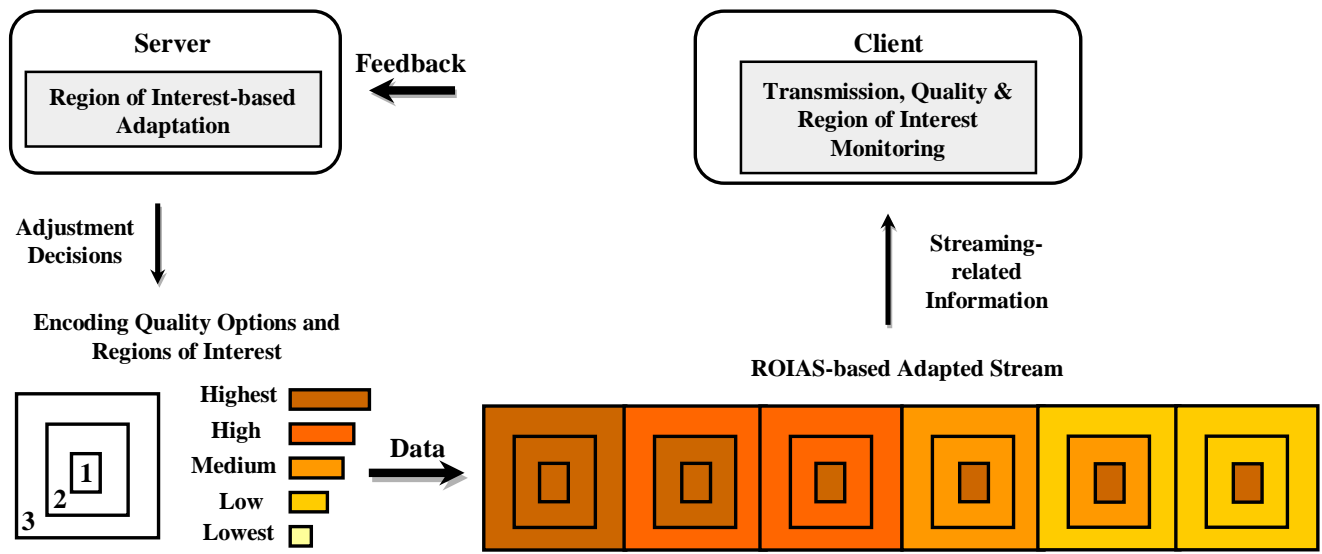
Fig 3. Illustration of ROIAS principle

Lastly, two other key issues in ROI video encoding research must also be mentioned: how to determine the regions of interest themselves, and what constitutes an ideal (*high quality, low quality*) combination in such coders. As regards the first question, determination of ROIs is usually done either by applying models of human vision [20], [21] to establish perceptually relevant areas of display, or by empirically determining these through eye tracking experiments [3]. In respect of the latter question, although there is no clear consensus in respect of what constitutes an ideal display combination, there is agreement on what parameters are to be used, and examples put forward mainly centre on varying either bit rates or frame rates [3], [20], [22], [23].

III. REGION OF INTEREST-BASED ADAPTATION

A. ROIAS Overview

The Region of Interest-based Adaptive Scheme (ROIAS) for multimedia streaming is an unicast rate-based adaptive solution for delivering high quality multimedia. Its goal is to increase the end-user perceived quality when viewing remotely streamed multimedia sequences in highly loaded delivery conditions by taking into consideration viewer's interest in certain multimedia frame regions and consequently differentiating their tratment during the adaptation process.

B. ROIAS Principle

ROIAS involves a client-server architecture and includes client and server-located components that are involved in the bi-directional exchange of video data and control packets through the delivery network. The client monitors some transmission-related parameters and regularly computes the quality of delivery scores, which are sent as feedback to the server. The server analyses these scores and proposes content-related adjustment decisions in order to increase user Quality of Experience in existing delivery conditions. Content adaptation is performed as research has shown that viewers of streamed multimedia content prefer controlled reduction in quality to the effect of random losses [24]. Therefore the transmitted quantity of multimedia data can vary during the streaming process.

Existing adaptive multimedia streaming schemes involve content modifications that affect equally the whole viewing area of the multimedia frames being transmitted. However as eye-tracking research has shown [3], there are some regions within multimedia streams' frames the viewers are more interested in than in others. Consequently ROIAS enhances the classic network condition-based adaptive solution for streaming multimedia with a novel approach.

When required to reduce the quantity and consequently the quality of transmitted multimedia-related information in order to meet the available bandwidth constraints, ROIAS affects the streamed data in terms of some compression-related parameters such as resolution and frame-rate differently based on the ROIs and on the user interest level on them. As result those ROIs the user is highly interested in are transmitted at high quality, whereas those on which the user interest is lower are streamed at lower quality, saving bandwidth.
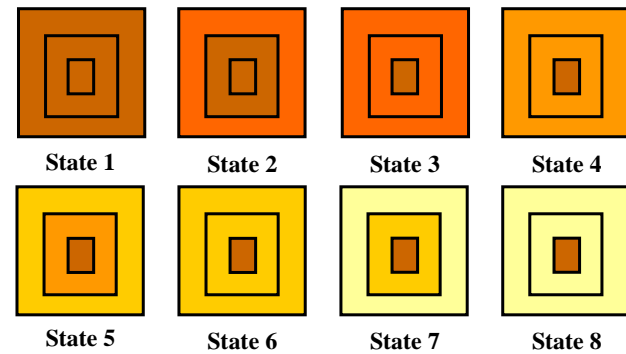

Fig. 4 – ROIAS server eight-state model

ROIAS's server side component maintains a viewer region of interest model that is updated regularly by this feedback. Based on the information from this model, ROIAS selectively adjusts the quality of those regions the viewer is the least interested in when transmission-related quality adaptations are required to be performed. As the quality of the regions the viewers are the most interested in will not change (or will involve little change), the proposed scheme will provide much higher overall end-user perceived quality than any of the

existing adaptive solutions, with significantly reduced bandwidth requirements. This is illustrated in Figure 2, which shows that the coverage of eye-tracking based regions of interest to be far less than that of the original frame, and highlights the opportunity to selectively stream only those perceptually relevant regions at the highest quality. Surrounding areas can be transmitted at lower quality in a step-wise process as the distance to the regions of highest interest to the user increases.

C. ROIAS Architectural Components

Figure 3 presents schematically the ROIAS architecture. ROIAS involves server-client communication via a bi-directional channel in order to exchange both multimedia data and feedback information.

Based on the multimedia stream resolution and on eye-tracking research results, the ROIAS server defines the overall multimedia viewing area and, within this area, a number of different regions of interests (ROIs). The placement of these ROIs is highly dependent on the multimedia sequence content and may also vary within the same multimedia stream from one scene to another. Following the observation that most professionally captured multimedia content includes the areas of highest user interest approximately in the middle of the viewing rectangle, in order for the proposed solution to be independent from the delivered content, ROIAS considers only concentric ROIs and associates the highest user interest to the ROI closest to the centre of the image. ROIAS orders these ROIs based on the decreasing user interest on them. For example Figure 3 illustrates three such ROIs.

The server also introduces a number of different potential multimedia quality levels which could be applied on the variously defined ROIs. Figure 3 exemplifies five possible quality levels: highest, high, average, low and lowest. The different quality versions of the same content are to be obtained in real time by adjusting some compression-related parameters such as resolution and frame rate.

The ROIAS server has an associated finite state model. Each server state indicates what multimedia quality level is to be associated with each of the ROIs in terms of the pre-defined quality levels. Figure 4 presents an eight-state ROIAS server model which always maintains the highest quality for the ROI on which the users have the highest interest. This model gracefully degrades the quality of neighboring ROIs in order to offer a smooth transition from higher to lower quality regions and consequently to maintain a high end-user perceived quality. Other state models can also be defined which may adjust the quality of all the ROIs, including that with highest user interest.

The ROIAS client monitors multimedia delivery in terms of loss, delay, jitter and estimated end-user perceived quality. It grades short-term and long-term Quality of Delivery grades (QoD$_{grades}$) which are sent to the server via feedback. The same mechanism was employed by QOAS and was described in detail in [9].

D. ROIAS Feedback-based Adaptation

During transmission, according to delivery con user quality quality state data sent. proposed RO interest. As adaptation, s that signific stream's end transmitting less information. In improved conditions, the server gradually increases the quality of the delivered stream and if the loss rate is kept low, this determines an increase in the end-user perceived quality. As a direct consequence of the ROIAS-based multimedia stream adaptation the viewers will receive an "adapted" stream as illustrated in Figure 3.

An example of ROIAS-based dynamic content adaptation involves a situation when, at the beginning, the server is in state 1 and the highest quality stream is transmitted. The stream's frames have all their ROIs, including those of little user interest, played at 25 fps and at maximum resolution. Due to the fact that ROIAS client feedback reports network congestion via low QoD$_{grades}$, ROIAS server dynamically enters in state 2. The resulted stream will have the region on which the user has least interest displayed at 20 fps and at a resolution degraded with 20%. If feedback continues to report non-optimal multimedia stream delivery, the ROIAS server
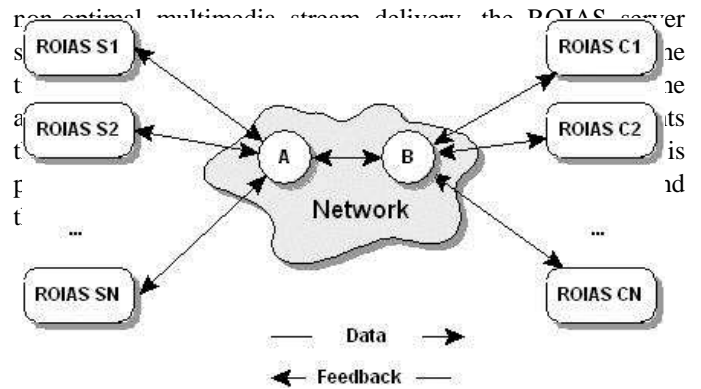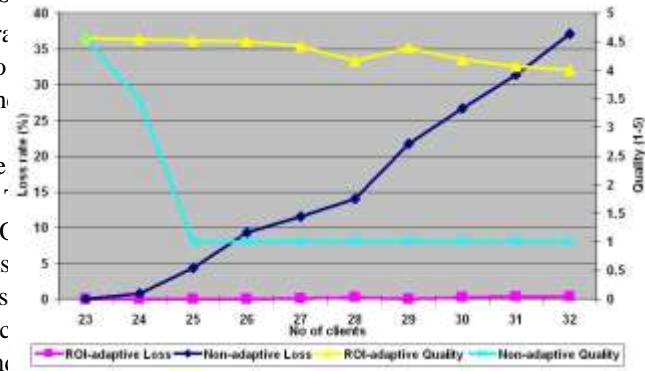

Fig 6. Comparative loss rate and estimated end-user quality when streaming multimedia using ROIAS and a Non-adaptive solution

Fig. 5 – Simulation testing setup

When compared with existing research in the area, ROIAS-based adaptation combines several attractive features. Firstly, instead of adjusting multimedia content across the whole viewing area as performed by the solutions put forward in the literature, ROIAS employs a novel adaptation approach. It performs quality-based adjustment selectively on different areas of the image depending on user interest on them. Secondly ROIAS allows high flexibility in its adaptation mechanism. The server state definition, the number of regions of user interest employed and their location and the number of different multimedia quality levels used can be modified. Last but not least, ROIAS incorporates an innovative client-based monitoring of multimedia delivery quality which includes an estimation of the end-user perceived quality which was already used successfully by QOAS [9]. The information collected is actively used by ROIAS in its feedback based adaptation loop.

IV. SIMULATION-BASED OBJECTIVE TESTING

Preliminary tests involved simulations using Network Simulator version 2 (NS-2) [25]. The "Dumbbell" topology that assumes a single shared bottleneck link (A-B) presented in Figure 5 was used. The sources of traffic are located on one side of the bottleneck link, whereas the receivers are on the other side. ROIAS's components were deployed at both the server and client (ROIAS S_i, and ROIAS C_i, respectively, where $1<=i<=N$). No other traffic was involved. Buffering at the bottleneck link uses a drop-tail queue of length proportional with the product between the round trip time and the bottleneck link's bandwidth. During simulations this bandwidth was set to 100 Mbps and the bottleneck link's delay was set to 0.1 s. Apart from the bottleneck link, the other links are over-provisioned such as they will not influence the simulation results.

Simulations involved ROIAS-based adaptive and non-adaptive multimedia streaming. *diehard1* – a multimedia sequence with very high motion content – was MPEG2 encoded at five different rates between 2 Mbps and 4 Mbps using the same frame rate (25 frames/s) and the same IBBP frame pattern (9 frames/GOP). Traces were collected and used during NS-2 simulations. ROIAS-based adaptation involved five server quality states, each having associated one of the encoded multimedia stream quality versions. The non-adaptive streaming used the maximum quality sequence with average bitrate 4 Mbps. Table I presents the properties of all multimedia versions used during simulations.

The simulations started with a number of clients randomly selecting the starting point from within the multimedia clip in order to allow for independence from the natural multimedia bitrate variations. The resulting streaming sessions lasted 500 sec from which 50 s transitory periods at the beginning and at the end were not considered. Simulation results were assessed in terms of average throughput, average loss rate and estimated end-user perceived quality. The quality is assessed using the non-reference moving picture quality metric proposed in [26] and expressed on the 1-5 ITU-T R. P.910 subjective quality

scale [27], where 1 represents "Bad" quality and 5 - "Excellent". Table II presents comparative simulation results when ROIAS and the non-adaptive scheme were used in turn with an increasing number of clients. At the same time Figure 6 graphically compares ROIAS and the non-adaptive solution in terms of loss rate (expressed as percentage and measured on the left hand vertical scale) and estimated end-user perceived quality (expressed as a value on the ITU-T P.910 five point scale and measured on the left hand vertical scale).

It could be clearly seen how when using ROIAS the number of clients streaming multimedia at least at "Good" quality (level 4 on the ITU-T 1-5 scale) can be increased with 40 % in comparison with the non-adaptive case, reaching 32. By using adaptation in cases when streaming rate rapidly number of end-user pe
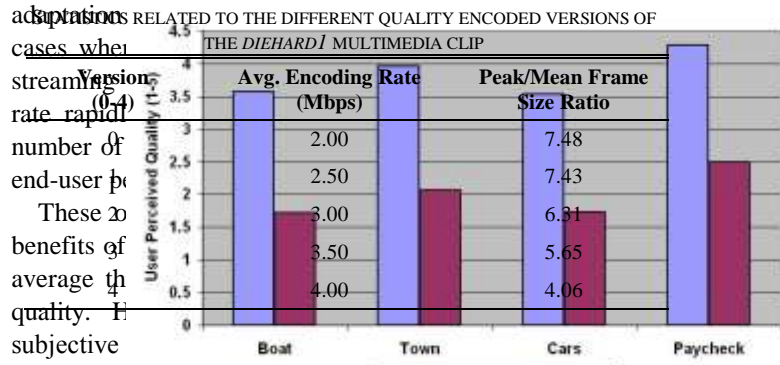
These re benefits of average th quality. H subjective when strea

The goal of the subjective tests performed was to confirm the good simulation testing results in terms of user perceived quality when streaming multimedia. Three different region of interest-based levels of adaptation and three different full-frame quality degradations that emulate the random effect of loss are considered and consequently the effect of the proposed ROIAS and of random noise on image quality was comparatively studied.

A. Testing Setup and Multimedia Sequences

Four MPEG-4 encoded clips with 640x480 pixels resolution representative for a class of video sequences in terms of spatial and temporal motion content (measured in terms of Spatial Index and Temporal Index respectively [28]) were selected as follows:

- Low spatial and low temporal motion content: *Boat* - features a boat moving slowly into a harbor.

TABLE I
PROPERTIES RELATED TO THE DIFFERENT QUALITY ENCODED VERSIONS OF THE *DIEHARD1* MULTIMEDIA CLIP

Version	Avg. Encoding Rate (Mbps)	Peak/Mean Frame Size Ratio
	2.00	7.48
	2.50	7.43
	3.00	6.31
	3.50	5.65
	4.00	4.06

Fig. 7. Comparative effect of ROIAS and noise on average subjective user perceived quality involving different types of clips

TABLE II
AVERAGE THROUGHPUT, LOSS AND QUALITY WHEN INCREASING THE NUMBER OF ROIAS-ENABLED STREAMING CLIENTS

Client No.	ROIAS			Non-adaptive		
	Through. (Mbps)	Loss (%)	Qual.	Through. (Mbps)	Loss (%)	Qual.
23	4.00	0	4.56	4.00	0	4.56
24	4.00	0	4.54	3.89	0.81	3.48
25	4.00	0.02	4.52	3.52	4.36	1
26	4.00	0.03	4.51	3.48	9.34	1
27	3.62	0.05	4.42	3.35	11.57	1
28	3.33	0.25	4.17	3.18	14.03	1
29	3.26	0.04	4.39	3.12	21.73	1
30	3.48	0.22	4.19	3.06	26.64	1
31	3.12	0.32	4.07	3.00	31.34	1
32	3.09	0.37	4.01	2.84	37.12	1

TABLE III
SUBJECTIVE TESTING RESULTS - MEAN AND STANDARD DEVIATION VALUES
OF USER PERCEIVED QUALITY SCORES

Clips		Orig.	ROI1	ROI2	ROI3	Noise1	Noise2	Noise3
Boat	Mean	4.40	3.87	3.63	3.20	2.13	1.87	1.17
	StDev	0.74	0.90	0.77	0.65	0.61	0.67	0.65
Town	Mean	4.40	4.20	3.90	3.77	2.50	2.27	1.43
	StDev	0.57	0.60	0.66	0.84	0.63	0.80	0.62
Cars	Mean	4.17	4.00	3.53	3.07	2.17	1.77	1.27
	StDev	0.56	0.68	0.86	1.03	0.77	0.88	0.59
Paycheck	Mean	4.87	4.43	4.10	4.30	2.53	2.70	2.27
	StDev	0.52	0.62	0.78	0.84	0.95	0.75	1.05

TABLE IV
T-TEST-BASED COMPARISON BETWEEN ROIAS AND NOISE-AFFECTED CLIPS

Clip	ROI Mean	Noise Mean	ROI Var.	Noise Var.	α	P two-tail	Significance
Boat	3.57	1.72	0.66	0.56	0.001	1.241×10^{-18}	Very
Town	3.96	2.07	0.51	0.67	0.001	1.884×10^{-18}	Very
Cars	3.58	1.73	0.87	0.69	0.001	1.843×10^{-15}	Very
Paycheck	4.28	2.50	0.56	0.85	0.001	5.200×10^{-15}	Very

- High spatial and low temporal motion content: *Town* - features a town scene with cars moving slowly down the street and pedestrians walking on footpaths.
- Low spatial and high temporal motion content: *Cars* - features cars moving quickly on a road outside a town.
- High spatial and high temporal motion content: *Paycheck* - features a high-speed car chase from a movie.

Out of the seven versions of each video clip used which included sound, the first was the original (marked *Orig* - e.g. *Boat_Orig*), the next three versions (denoted *ROI1*, *ROI2*, and *ROI3*) included different ROI-based levels of adaptation (decreasing image resolutions were used for the regions with less user interest), while the last three versions involved simulated quality degradations applied equally on the whole image. These three versions (named *Noise1*, *Noise2*, and *Noise3*) were so generated in order to be affected by increasing levels of salt-and-pepper noise. In order to produce the clips with different ROI quality levels, sixty degradations were defined between the image center and the frame edge. For the clips with the least quality modification (*ROI1*), quality degradations in terms of resolution were step-wise and were performed with a step of 1, whereas *ROI2* and *ROI3* involved steps of 2 and 3 respectively. The clips the least affected by noise (*Noise1*) were produced by generating noise with a probability of 0.008, whereas the *Noise2* and *Noise3* clips were produced by increasing the noise probability to 0.01 and 0.04 respectively.

The tests involved 15 subjects mostly in the 30 to 50 year age bracket, three over 50 and two less than 30, with approximately half of the participants being familiar with multimedia streaming. None of the participants had any visual impairments that could affect their perception of video quality, but nine were wearing glasses.

For each of the seven versions of the four video clips the participants were asked to grade the clips on a modified version of the ITU-T R. P.910 scale with half point granularity. The chosen testing method is a combination between the Absolute Category Rating (ACR) and the Degradation Category Rating (DCR) presented in [27]. ACR involves the subjects grading separately each clip, but an implicit reference must be well known by all assessors, which is not expected in this case. DCR involves showing the reference clip before each test clip, which in this case would make the overall test time very long. Here, by combining ACR and DCR, the reference clip is shown first and then the multimedia sequences that have to be assessed.

Before starting the actual test, the fifteen participants where shown a training clip which contained a ten second example of each modification type, which helped to familiarize the participants with both the test procedure and the quality degradations involved. The overall goal of the experiment was outlined in addition to what was required from the participants.

During the experiment the participants were first shown the unaltered version of each video clip. Then the remaining clips with varying levels of ROI adaptation and noise probability were shown in a random order. In order to enable the participants to accurately appreciate and evaluate the quality of each clip, all clips had a duration of approximately one minute.

To ensure the experimental conditions remained consistent the same test equipment was used for all experiments. A Compaq Laptop with an inbuilt 15-inch LCD display and Windows Media Player 10 application were used to display all the video clips using the original resolution. To ensure a consistent audio level a constant audio level (approximately 70 dB) was used for all participants. Once the laptop was appropriately setup and the user felt comfortable the experimental process was started.

B. Testing Results

The subjective testing results presented in Table III indicate that ROIAS-based multimedia streaming was appreciated by

the human subjects and scored around 4, the ITU-T R. P.910 "good" quality level, across all movie clip types. It is significant to note that for the *Paycheck* clip, in spite of the fact that it has the highest temporal and spatial motion content among all the clips used, scored the best, achieving on average 4.28. The explanation may be that as this clip is a sequence from a commercial movie that used professional film shooting of content in order to attract the viewer's attention to the areas where action is performed, by introducing little quality modification to those regions, the adjustments on the areas of little user interests remain unnoticed. The opposite explanation stands for the low result obtained by the *Boat* sequence with the lowest temporal and spatial motion content. Since this clip contains the view of a scenic harbor entrance with a boat moving in the middle of the picture, viewers tend to explore the entire panoramic view shown and not just the regions of activity located around the boat. This result is especially important for ROIAS assessment as most other adaptive schemes such as QOAS for example [29], would achieve the worst results in terms of end-user perceived quality for the clips with the highest action content and the best when the degree of motion in the clips' content is low.

The low standard deviation values presented for each case in Table III indicate that the results obtained are consistent, despite the coarse granularity of the grading process (i.e. lower than 0.5 point fractional grades were not accepted).

When comparing the scores obtained by the ROIAS clips with those received for the noise affected videos it can be seen that while the ROI clips' grades are on average at "good" level, the other scores are between 1.7 and 2.5 corresponding to "poor" quality level on the ITU-T 1-5 scale. The t-tests performed on the scores given by the participants in each case confirmed that there is a very significant statistical difference between the ROIAS and the noise-affected clips' quality grades, in favor of the region of interest-based scheme. This result is valid regardless of the clips motion content level. A summary of the t-test results is presented in Table IV.

By analysing these subjective testing results – graphically summarized in Figure 7 - there is no doubt about the users' preference for region of interest-based adaptivity in comparison with a full-image quality adjustments.-based solution. It is also important to note the average ROIAS end-user perceived quality score is only 13.73 % adrift than the average grade obtained by original sequences which were not affected by any modifications due to transmission related constraints.

VI. Conclusions and Future Work

This paper introduced ROIAS, a frame region of interest-based adaptive scheme for multimedia streaming, in which the adaptation is performed at the level of certain regions within clip frames based on user interest obtained as a result of eye-tracking monitoring. ROIAS selectively adjusts the quality of those regions from the multimedia frames the viewer is the least interested in when transmission-related quality adaptations to existing delivery network conditions are required to be performed. As the quality of the regions the viewers are the most interested in will either not change or will involve little adjustment, the proposed scheme achieves high overall end-user perceived quality.

Simulation results involving multiple clients streaming multimedia simultaneously show how ROIAS performed much better in terms of average client throughput, loss and estimated end-user perceived quality than when a non-adaptive streaming scheme was used.

Subjective testing results involving clips with different levels of motion content show how ROIAS-based adaptive multimedia streaming was appreciated by the tests subjects, which, on average, graded their perceived quality as "good".

Work in progress compares the performance of ROIAS-based multimedia streaming with that of other adaptive schemes such as LDA+ and TFRCP. Adapting ROIAS in order to enable multimedia streaming over wireless networks is also envisaged.

Future work aims at enhancing ROIAS with an eye-tracking-feedback loop that determines the regions of interest for the viewers in real-time. Information related to the location of the region of maximum interest will be collected by the ROIAS client and sent regularly as feedback to the server. ROIAS server will be able to perform content adaptation based on dynamic ROIs.

References

[1] S. Dravida, D. Gupta, S. Nanda, K. Rege, J. Strombosky, M. Tandon, "Broadband Access over Cable for Next-Generation Services: A Distributed Switch Architecture", IEEE Communications Magazine, Vol. 40, No. 8, August, 2002, pp. 116–124

[2] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, J. M. Peha, "Streaming Video over the Internet: Approaches and Directions", IEEE Trans. On Circ. and Sys. for Video Tech., vol. 11, no. 3, 2001, pp. 282–300

[3] S.R. Gulliver, G. Ghinea, "Stars in their eyes: what eye-tracking reveals about multimedia perceptual quality", IEEE Transactions on Systems, Man and Cybernetics, Part A, vol. 34, No. 4, July 2004, pp. 472 – 482

[4] D. Sisalem, H. Schulzrinne, "The Loss-Delay Based Adjustment Algorithm: A TCP-Friendly Adaptation Scheme", Proc. ACM NOSSDAV, 1998

[5] D. Sisalem, A. Wolisz, "LDA+ TCP-Friendly Adaptation: A Measurement and Comparison Study", ACM NOSSDAV, 2000

[6] R. Rejaie, M. Handley, D. Estrin, "RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet", Proc. IEEE INFOCOM, 1999

[7] R. Rejaie, M. Handley, D. Estrin, "Layered Quality Adaptation for Internet Video Streaming", IEEE Journal on Selected Areas of Communications, vol. 18, no. 12, Dec. 2000, pp. 2530-2543

[8] J. Padhye, J. Kurose, D. Towsley, R. Koodli, "A Model Based TCP Friendly Rate Control Protocol", ACM NOSSDAV, 1999

[9] G.-M. Muntean, P. Perry and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks", IEEE Trans. Broadcasting, vol. 50, no. 1, March 2004, pp. 1-10

[10] S. McCanne, V. Jacobson, M. Vetterli, "Receiver-Driven Layered Multicast", Proc. ACM SIGCOMM, Stanford, California, USA, August 1996, pp. 117-130

[11] L. Vicisano, J. Crowcroft, L. Rizzo, "TCP-like Congestion Control for Layered Multicast Data Transfer", Proc. IEEE INFOCOM, vol. 3, San Francisco, CA, USA, March 1998, pp. 996-1003

[12] I. Rhee, V. Ozdemir, Y. Yi, "TEAR: TCP Emulation at Receivers - Flow Control for Multimedia Streaming", Technical Report, CS Department, NCSU, April 2000

[13] L. Wang, A. Luthra, B. Eifrig, "Rate Control for MPEG Transcoders", IEEE Trans. On Circuits and Systems for Video Technology, vol. 11, no. 2, Feb. 2001 pp. 222–234

[14] N. Yeadon, F. Garcia, D. Hutchison, D. Shepherd, "Filters: QoS Support Mechanisms for Multipeer Communications", IEEE J. on Selected Areas in Comp., vol. 14, no. 7, Sept. 1996, pp 1245-1262

[15] A.T. Duchowski and A. Coltekin, "Foveated Gaze-Contingent Display for Peripheral LOD Management, 3D Visualisation, and Stereo Imaging", accepted for publication ACM Transactions on Multimedia Computing, Communications and Applications.

[16] L. C. Loschky and G.S. Wolverton, "How Late Can You Update Gaze-contingent Multi-resolutional Displays Without Detection?" accepted for publication, ACM Transactions on Multimedia Computing, Communications and Applications.

[17] E. M. Reingold and L. C. Loschky, "Reduced Saliency of Peripheral Target in Gaze-contingent Multi-resolutional Displays: Blended versus Sharp Boundary Windows", in Proc. of the Eye Tracking Research and Applications Symposium, New Orleans, Louisiana, USA, pp 89-93, 2000.

[18] S. Shioiri and M. Ikeda, "Useful resolution for picture perception as a function of eccentricity". Perception, 18, 347-361, 1989.

[19] L. C. Loschky and G. W. McConkie, "User performance with Gaze-Contingent Multiresolutional Displays", in Proc. of ACM Eye Tracking Research & Applications Symposium (ETRA) 2000, Palm Beach Gardens, FL, pp 97-103, 2000.

[20] W. Lai, X-D. Gu, R-H. Wang, W-Y Ma, and H-J Zhang, "A Content-based Bit Allocation Model for Video Streaming", in Proc. IEEE International Conference on Multimedia and Expo (ICME 2004), Taipei, Taiwan, June 2004, pp. 1315-1318.

[21] W. Osberger and A. J Maeder. "Automatic identification of perceptually important regions in an image using a model of the human visual system." In 14th International Conference on Pattern Recognition, Brisbane, Australia, Aug 1998, pp. 701-704.

[22] W. Lai, X-D. Gu, R-H. Wang, L-R. Dai, and H-J. Zhang "A region based multiple frame-rate tradeoff of video streaming", in Proc. IEEE International Conference on Image Processing (ICIP 2004), Singapore, October 2004, pp.2067-2070.

[23] A.C-W. Wong and Y-K. Kwok "On a region-of-interest based approach to robust wireless video transmission", in Proc IEEE International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN 2004), Hong Kong, China, May 2004, pp 385 - 390

[24] G. Ghinea, J. P. Thomas, "QoS Impact on User Perception and Understanding of Multimedia Video Clips", Proc. ACM Multimedia, Bristol, United Kingdom, 1998, pp. 49-54

[25] Network Simulator-2, http://www.isi.edu/nsnam/ns

[26] O. Verscheure, P. Frossard, M. Hamdi, "User-Oriented QoS Analysis in MPEG-2 Video Delivery", Journal of Real-Time Imaging, vol. 5, no. 5, Oct. 1999, pp. 305-314

[27] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications", September 1999

[28] N. Cranley, L. Murphy, P. Perry, "Content-Based Adaptation of Streamed Multimedia", in "Management of Multimedia Networks and Services", Lecture Notes in Computer Science vol. 3271, Springer Verlag, October 2004, pp.39-49

[29] G.-M. Muntean, P. Perry, L. Murphy, "Subjective Assessment of the Quality-Oriented Adaptive Scheme", IEEE Trans. on Broadcasting, vol. 51, no. 3, September 2005, pp. 276-286