

User Perception Model for Wearable Supervision Systems

Damien Perritaz
damien.perritaz@epfl.ch

Christophe Salzmann
christophe.salzmann@epfl.ch

Denis Gillet
denis.gillet@epfl.ch

Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne
Switzerland

ABSTRACT

Wearable supervision systems ease the deployment of advanced mobile solutions for the control of industrial plants. These systems are used to provide operators with the adequate information to perform the required manual operations on industrial plants. This paper presents the adaptation strategy developed to ensure that the operator perceives accurately the plant state relayed by a distant server despite the varying network conditions. The information provided to the user is mainly in the form of Augmented Reality video rendered in a Head Mounted Display. Using experimental subjective testing, the user video Quality of Perception is modeled to determine continuously the best encoding parameters values resulting from the compromise between the fluidity and the level of detail for real-time interaction with the plant. This model is then used by an adaptation scheme to reject output bitrate disturbances due to the variations of the spatial and temporal video content. The proposed approach adapts in real-time the parameters values of the video encoder to track a given reference bitrate.

1. INTRODUCTION

Today's industrial plants are highly automated and usually managed from a centralized control room. Manipulations on the plant require the collaboration of two operators, one located near the installation to perform observations and manual operations, and one located in the control room to manage the automated operations.

In this context, the 6^{th} Sense project aims at developing a wearable supervision system for chemical plants to remove the need for an operator in the control room. The proposed wearable supervision system uses Augmented Reality (AR) video to improve the user perception of the plant by displaying computer-generated graphics in the Head Mounted Display (HMD) of the operator. A handsfree speech recognition module is implemented to remotely control the plant while performing manipulations. A wireless communication link between the mobile system (thin client) and the server

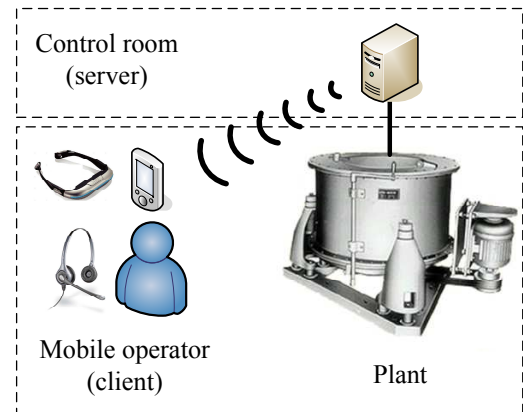


Figure 1: Typical configuration for industrial wearable supervision systems

is used to transmit information to and from the operator working on the plant (Figure 1). The real-time information streams to be provided to the operator are:

- AR video to be displayed in the HMD;
- audio speech synthesis for alarm and instruction communication;
- operating conditions of the plant relayed by the server.

This paper focuses on the video stream. By analogy the proposed solution can be extended to the other streams of information. In the context of wearable supervision system, the video stream has specific real-time constraints. The AR video should be rendered with minimal delay and at a regular pace to reproduce the dynamical behavior of the installation despite the movements of the head and to ensure safe operations.

Transmissions over a wireless link are subject to the large variations of the network characteristics. As a result, both the available bandwidth and the transmission delay are constrained. Any information transmitted that could exceed the available bandwidth would be delayed or discarded by the network. At the application level, adaptation techniques are used to ensure that the information sending rate is in accordance with the available bandwidth. The stream content is

compressed in order to adapt the amount of transmitted information to the available bandwidth. An estimation of the available bandwidth based on the network Quality of Service (QoS) is assumed to be given by a network adaptation module not presented here. Users preferring a compressed video than a corrupted video due to packet loss [1], the proposed adaptation scheme matches the output bitrate to the available bandwidth by continuously adapting the values of the parameters of the encoder.

The Quality of Perception (QoP) is defined as the quality measure for the perception experienced by users. It is somehow equivalent to a user oriented QoS. The QoP for video stream can be based on many criteria: in multimedia presentation, [2] proposed a QoP evaluation that differentiates satisfaction from understanding; [3] proposed an adaptation architecture that uses the packet loss rate as QoP measure; the subjective evaluation of video clips presented in [4] has shown that an optimal configuration of encoding parameters exists between possible configurations in a zone of equal average bitrate.

Video encoding is based on the fact that there is a strong correlation between both the successive video frames and within the single video frame elements themselves. Thus, decorrelation of these signals can lead to bandwidth compression without significantly affecting image resolution. Moreover, additional compression techniques that exploit the insensitivity of the Human Visual System (HVS) of certain spatio-temporal visual information can further reduce the amount of transmitted data. Video codecs are generally based on two fundamental redundancy reduction principles: *i*) spatial redundancy reduction compresses similar pixels within the frame with compression techniques that exploit HVS limitations; *ii*) temporal redundancy reduction removes similarities between the successive pictures. Due to the encoder characteristics, the output bitrate of an encoded video stream depends on its content. A video with poor spatial content (low spatial redundancy) has a lower output bitrate than a video with rich spatial content (image with many details). Similarly in temporal term, a static video (low temporal redundancy) has a lower output bitrate than a dynamical one (sequence with fast moving objects). This property will be exploited in the proposed QoP-based adaptation scheme.

This paper is organized as follow: Section 2 presents a model of perception to determine the optimal encoding parameters values according to a variable reference bitrate. This model of perception based on subjective evaluations is estimated in Section 3. Finally, an adaptation scheme is proposed in Section 4 to match the video output bitrate to the reference, by adapting the values of the encoding parameters: the framerate (influences the fluidity) and the compression parameter (influences the level of detail).

2. MODEL OF PERCEPTION

The QoP can be modeled as the impact of the selected encoding parameter values forming a tuple on the user perception. The video stream is encoded according to a reference bitrate. It exists many encoding tuples to achieve a given bitrate. If only one encoding parameter is used, there is a unique encoding 1-tuple matching the reference bitrate. With two encoding parameters, many 2-tuples (or pairs) are

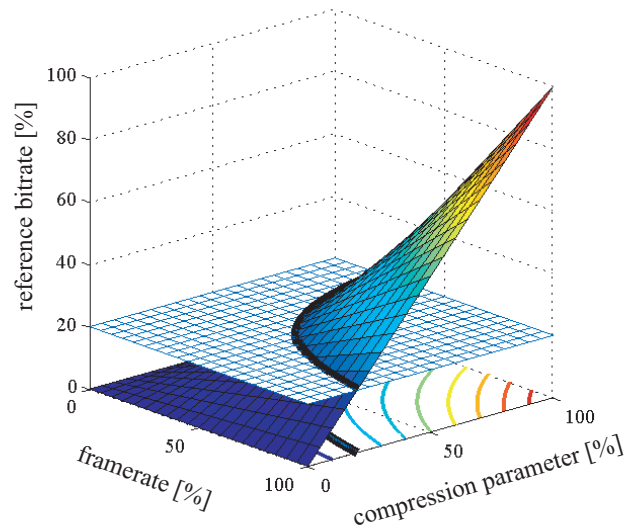


Figure 2: Isobitrate 20% is the intersection between the encoding surface and 20% bitrate plan; isobitrates are projected in the encoding space

accessible to match the reference bitrate; one degree of freedom is given for each new encoding parameter.

In Figure 2, the tuples are represented as points on a surface (encoding surface) with both encoding parameters on the base axis, and the reference bitrate for the third dimension. The encoding parameters are normalized and defined as a percentage of their corresponding maximum output bitrate for selected spatial and temporal characteristics of the stream. The reference bitrate is represented in percentage of its maximal value which corresponds to the encoder output bitrate for the tuple (100%;100%). The encoding surface represents the relation between the encoding parameters values and their corresponding reference bitrate. To find the set of tuples for a given bitrate, the intersection line between the encoding surface and the horizontal plan corresponding to the bitrate is used; this line is called an isobitrate. Thereafter, the isobitrates are projected on the parameters plan called the encoding space.

The model of perception aims at determining the optimal encoding tuple for a desired reference bitrate. The Local Adaptation Path (LAP) is the model of perception, defined as the path in the encoding space passing through the optimal encoding tuples of all the possible reference bitrates. The Local Adaptation Scheme (LAS) presented in Section 4, uses this model to select the optimal tuple for the reference bitrate by moving along the LAP.

3. MODEL ESTIMATION

In this section, the perception model determining the optimal encoding tuple is constructed. To determine the LAP, two perception evaluation methods exist: *i*) the subjective evaluation, based on user rating through experiments; *ii*) the objective evaluation, based on mathematical comparison between the original and the encoded stream. The subjective evaluation has been chosen as the objective evaluation does not provide results sufficiently correlated with human

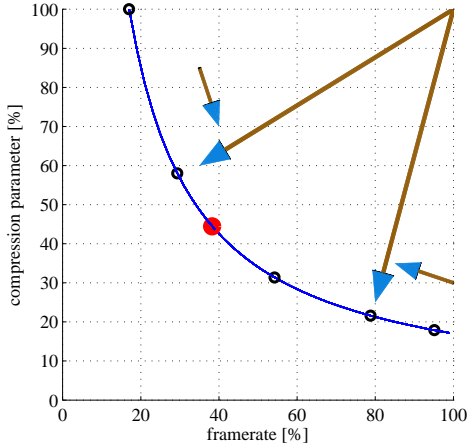


Figure 3: Forced Choice Methodology applied to an isobitrate: a pair of tuples is presented to converge on the best one (disc)

vision [5]. The user QoP varies according to the considered subject of observation, thus experiments should be performed according to specific context. The selected subjective methodology employed is the Forced Choice Methodology (FCM). This methodology has been selected for the simplicity of the statistical analysis and the absence of contextual effect influencing the rating [6].

For the sake of repeatability, the industrial plant is emulated in laboratory. It consists in a one meter high boiler with three valves and a manometer. Typical actions performed on the industrial plant are reproduced in the laboratory. The AR system consists in an opaque HMD and a camera fixed on its top. The acquired images are encoded and then displayed in the HMD to provide video see-through. The user wears the HMD and interacts with the plant by performing various operations such as reading values and manipulating valves. The experiments outcomes determine the best compromise between the fluidity (framerate) and the level of detail (compression parameter) for a given bitrate. The representation of this model is represented in the encoding space by the LAP. The selected encoder (MPEG4-10) is implemented with the help of the FFmpeg library. Its gop-size is set to 10 between key frames. The selected camera delivers 640x480 image at 30 frames per second. The resolution of the 3DVisor HMD display is 800x600 pixels.

The QoP model has been defined following the ITU recommendations [6] by proposing a set of experiments to a group of 13 users. In the FCM, the user is presented with video sequences encoded at a specific bitrate with varying values for the framerate and the compression parameter. The user has to select the best tuple between a pair of alternatives. The proposed video sequences start with tuples located at both end of the isobitrate. The selected sequence is then proposed with the sequence next to the canceled one; the procedure is repeated until a preferred QoP is determined for the given isobitrate (Figure 3). This QoP evaluation is repeated for the next isobitrates.

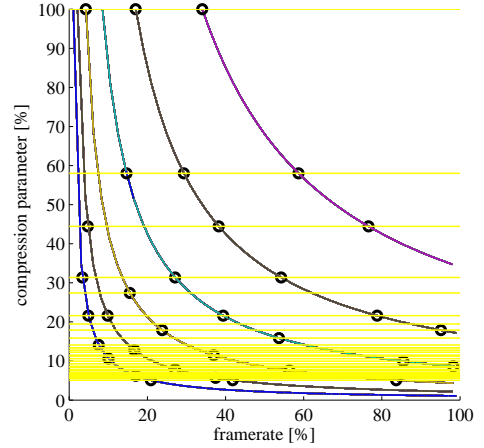


Figure 4: Tested tuples (circles) located on feasible compression parameter thresholds (horizontal lines)

Finding the LAP consists in testing all possible encoding tuples for a given bitrate in order to find (subjective quality evaluation) the optimal encoding tuple. This procedure should be repeated for all bitrates. Testing all possibilities being not feasible in a reasonable amount of time, the number of tests is reduced by sampling the encoding space. Six representative (in term of HVS sensitivity) isobitrates have been chosen, principally in the low bitrate area that is of interest for wireless transmission (1%, 2%, 4%, 8%, 17% and 34%). For these isobitrates, specific tuples have been chosen according to the spatio-temporal contrast sensitivity of the human [5]. Due to the encoder implementation limitations, the range of the compression parameter is not continuous and only 30 thresholds exist (Figure 4). As a consequence, the tuples are constrained to a finite number of feasible lines.

For each tested isobitrate, the optimal encoding tuple is computed as the average of the preferred encoding tuples given by each user, adapted on the nearest feasible compression parameter. To provide an optimal tuple for every reference bitrate, the LAP is extrapolated between the optimal tuples given by the experiments (Figure 5). The LAP is a piecewise definite path, constrained in the feasible compression parameter values, covering the whole range of bitrate. The best QoP is obtained near the bisector between the two encodings and not along a unique encoding axis; it confirms the idea of seeking the optimal tuple as a compromise between the encoding parameters. For low bitrates, results show that the QoP is maximized when the framerate is increased rapidly to around 40% (around 13 frames per second): in the proposed context, users prefer the fluidity than high level of detail. For higher bitrate, the tests show that the level of detail becomes more important, the framerate being even decreased. As a matter of fact, users are more sensitive to an increase in level of detail (giving a better perception) than an increase in the framerate that already permits a good perception of the dynamic. For higher bitrates, level of detail is preferred; this can be explained by the fact that the HVS is less sensitive to framerate if it is already higher than 20 frames per second (60%) [5].

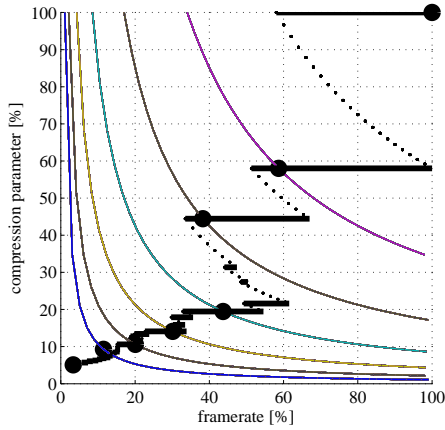


Figure 5: LAP resulting from subjective evaluation projected on the encoding space with computed optimal tuples (discs) and extrapolation on feasible compression parameter (lines)

4. ADAPTATION STRATEGY

The adaptation scheme ensures that the encoder sending rate is in accordance with the available bandwidth. Figure 6 describes the proposed adaptation scheme using a system based representation. The network is represented as a system with the encoder output bitrate as input, and the measured QoS parameters as output. The global adaptation scheme consists of a cascade controller with two loops where the network adaptation loop (outside loop) provides the stream content adaptation loop (inside loop) with the reference bitrate based on the measured network QoS [7, 8]. The content adaptation scheme tracks the provided bitrate by adjusting the encoder parameters values according to the QoP model that defines the optimal parameters path.

The stream adaptation scheme tracks the provided reference bitrate to compensate the output bitrate variations due to variations of the video content. The QoP model is used to encode the stream in an optimal manner in term of perception. The reference bitrate given to the Local Adaptation Scheme (LAS) is adapted according to the bitrate measured after the encoding process. The adaptation scheme matches the measured output bitrate to the reference bitrate by rejecting disturbances due to content variation; this controller is called the Content Adaptation Scheme (CAS).

5. CONCLUSION

This paper presents the methodology to ensure the best QoP for the real-time video to be rendered on HMD used in supervision systems. The user QoP is first modeled via experimental subjective testing using the Forced Choice Methodology for a given video encoder (MPEG4-10). This model defines within the encoding space the best encoding tuple for a given reference bitrate. The encoding space is a 2-dimensional coordinate system defined by the compression parameter and framerate values. The Local Adaptation Path is a piecewise definite path, constrained within the selected range of bitrates and the feasible compression parameter values that maximizes the user QoP. This model is used by an adaptation scheme that tracks the given bitrate

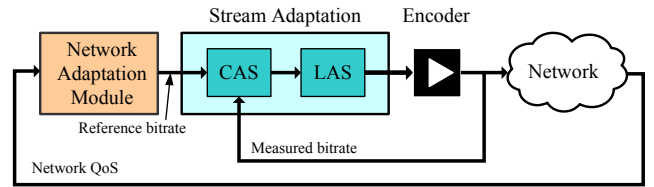


Figure 6: Cascade controller with control flows: outer loop with *Network Adaptation*, inner loop with *Stream Adaptation*

by rejecting the disturbance in the output bitrate induced by the variations of the spatial and temporal video contents. Experimental validation showed that at low bitrate the user QoP model privileges the frame rate up to about 10 frames per second, then the video quality must also be increased even at the cost of reducing the framerate.

The QoP model will be integrated within the global adaptation scheme and validated. At the same time, the proposed methodology will be used to develop models for audio and data transmission, and for a model that uses additional video encoding parameters. These models will be combined to define a multimodal perception model in order to globally maximize the user QoP.

6. ACKNOWLEDGMENTS

Hasler Stiftung: Man-Machine Interaction Program #1970

7. REFERENCES

- [1] O. Verscheure, P. Frossard, and M. Hamdi. MPEG-2 video services over packet networks: Joint effect of encoding rate and data loss on user-oriented qos. In *EEE Workshop NOSSDAV'98*, Proceedings of EUSIPCO 2002, pages 257–264. EUSIPCO, 1998.
- [2] G. Ghinea and J.P. Thomas. Quality of perception: user quality of service in multimedia presentations. *Multimedia, IEEE Transactions on*, 7(4):786–789, August 2005.
- [3] P.M. Ruiz and E. Garcia. Improving user-perceived QoS in mobile and wireless IP networks using real-time adaptive multimedia applications. In *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, volume 3, pages 1467–1471 vol.3, 2002.
- [4] N. Cranley, P. Perry, and L. Murphy. Dynamic content-based adaptation of streamed multimedia. *Journal of Network and Computer Applications*, 30(3):983–1006, August 2007.
- [5] S. Winkler. *Vision models and quality metrics for image processing applications*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 2000.
- [6] ITU-T Recommendation BT.500-11. Methodology for subjective assessment of the quality of television picture applications. January 2002.
- [7] Ch. Salzmann. *Real-time interaction over the Internet*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 2005.
- [8] Ch. Salzmann, D. Gillet, and Ph. Mullhaupt. Real-time interaction over the internet: Model for QoS adaptation. In *16th IFAC World Congress*, 2005.