# Perceptual Semantics for Video in Situation Awareness

María Alejandra Pimentel-Niño, María Ángeles Vázquez-Castro and Iñigo Hernáez-Corres

Dept. of Telecommunications and Systems Engineering
Universitat Autònoma de Barcelona
Bellaterra, Spain 08193
Email: {mariaalejandra.pimentel, angeles.vazquez}@uab.es, inigo.hernaez@e-campus.uab.cat

***Abstract*—We introduce novel perceptual semantics for video adaptation in multimedia communications. The target is to enhance situation awareness in non-computer aided processes as in emergency operations. Our proposed perceptual semantics relate to end user requested resolution in the temporal domain for a better assessment of event's evolutions seen from streaming video. Adaptation is enabled at transmission via a perceptual semantics feedback loop to adapt source coding on-the fly in terms of frame rate. The overall framework contemplates the use of an underlying cross-layer optimization that copes with network congestion and erasures in best effort scenarios. We show through simulations that within the proposed framework, the perceptual semantics are preserved. Moreover, we show it complies with information-centric-networking philosophy and architecture, such that it is in line with content-aware trends in networking.**

***Keywords–Perceptual semantics; adaptive video; situation awareness; QoE.***

## I. Introduction

The motivation of the work presented here is the possibility of enhancing situational awareness using video streaming through constraint networks. We target non-computer-aided scenarios, where there is no artificial intelligence behind interpreting sensory information from the received video.

We consider band-limited, wireless best effort networks, as possible means of communications for point-to-point live video streaming during scenarios such as in emergency operations. Such networks pose a number of constraints affecting Quality of Service (QoS), namely, congestion and erasures. The topology envisioned is that of live user-generated content being upstreamed to proper decision-makers.

Transmission alternatives that cope with the aforementioned network constraints include UDP- based frameworks for live or real-time applications, with quality based [1], Quality of Experience (QoE) driven cross-layer optimization [2], or TCP-friendly [3] adaptive algorithms. Dynamic Adaptive Streaming for HTTP over TCP streaming is suitable for server-client architectures and video on demand applications, but with degraded performance in lossy networks with large propagation delays [4]. With regards to the lossy nature of the network, forward erasure correction methods are able to provide enough protection and maintain QoE [5][3][6].

While these transmission frameworks can be satisfactory to guarantee QoS/QoE in video-for-entertainmet scenarios, we propose an additional dimension to target specific user demands in scenarios using video for other purposes, such as in emergency operations. We propose to improve non-computer-aided situation awareness beyond standard improvements by means of perceptual semantics.

In multimedia, "classic" semantics deals with heterogeneous metadata that sensors observe and/or tag when capturing video. As such, it has applications in information retrieval, integration and aggregation of varied data types as in semantic-aware delivery of multimedia [7]. Further, semantic tagging describing pure observations is used in computer-based systems with artificial intelligence to perceive and abstract situations [8]. Rather than doing perception through classic semantics, we propose a novel human-analysis-driven perceptual semantics to tag the videos, based on the temporal/spatial characteristics a user is perceiving as means to improve situation awareness.

The term perceptual semantics has been used by Cavallaro and Winkler [9] for automatic feature extraction of video based on image segmentation and target internal changes within the video source coding mechanisms. Our approach differs, as it does not limit to particular feature extraction and it focuses on offering a solution for video communications in a non-intrusive manner towards video codecs. Further, we involve the user in tagging first-level perceptual features, for perceptual-based networking, rather than use only observations as in [10]. To the best of our knowledge, such diversion from classic semantics has not been explored before.

Finally, we frame the novelty of perceptual semantics such that it complements cross-layer optimization schemes that help cope with network constraints. Within this framework we enable a perceptual semantic adaptation loop, which will target the specific user's demand for improved perception, comprehension and further projection in situation awareness processes. Additionally, this framework can be mapped to current content-centric approaches of information-centric-networking.

The structure of the paper is as follows. We present the perceptual semantics model in Section II, followed by the integration of perceptual semantics within a video adaptation framework in Section III. We show simulation results in Section IV, and draw final conclusions in Section V.

## II. Perceptual Semantics Model

In this section, we derive the model for perceptual semantics in the context of situation awareness and how we propose to perform semantic tagging.

### A. *Spatial/temporal decoupling for situation awareness*

Situation awareness enables good decision-making [11] and hence, it is a major asset in, e.g., emergency operations. A

 **ISBN: 978-1-61208-368-1** 

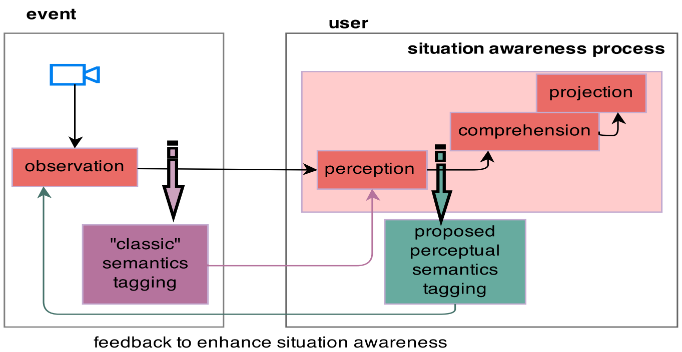



Figure 1. Perceptual semantics vs "classic semantics"

broadly accepted definition is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." [12]. The three-level model is thus inferred, namely: perception, comprehension, and projection.

We focus on the spatial and temporal advantages of video as a source of information for situation awareness. First, the possibility of capturing dynamic scenes improves the assessment of temporary evolving events. Second, video can provide visual spatial accurate accounts of an ongoing situation [13].

If the temporal and spatial perceptual characteristics of the video satisfy the situation-dependent specific user resolution, then the user satisfaction will be fulfilled, and further higher level cognitive processes will be benefited.

### B. Semantic tagging

Based on the spatio/temporal identification of perceptual features, our proposal is to utilize the end-user's (analyst) perception, to do semantic tagging that enables an enhancement of the received video stream signal tailored to the user's demand.

Semantic tagging is hence performed to describe perceptual features in the video and as such represents more complex abstractions of a viewed scene. In comparison, classic semantics tagging would focus on unprocessed sensorial observations [8]. The difference between both approaches in semantics is shown in Figure 1.

In scenarios where perception is not achieved by artificial intelligence, it is the human analysis that will interpret the sensory information and follow the three steps in the situation awareness model. Hence, the semantic tagging is performed by the user, as he is ultimately the one perceiving and foreseeing what might be of interest in the video.

We propose a tagging that would indicate the temporal/spatial predominance according to the level of perception of the user. A tag indicating predominance of temporal features, means the user is perceiving a situation that demands more attention to the dynamics of the scene (e.g., rapid movements, evolution of an environmental hazard). On the other hand, a predominance of spatial features indicate moments of less movement but densely overloaded frames that requires more detail to identify features (e.g., identifying persons or details in a emergency scene).

## III. Integration of Perceptual Semantics to Video Adaptation

We propose to integrate the perceptual semantics with video adaptation in order to provide to the user the required perceptual level for situation awareness. Further, we propose to map the tags to actions, such that the specific perceptual features are enhanced.

Following, we describe how our perceptual semantics model can be mapped to video coding characteristics and propose an algorithm to meet the end-user's demands. We comment on protocol aspects in the implementation and propose an integrated framework with an adaptive video solution.

### A. Mapping

We focus on using our proposed perceptual semantics for enhancement at source coding level. In single layer or scalable layer video encoding of state-of-the-art codecs, three types of resolution are defined, namely temporal (frame rate), amplitude (quantization step), and spatial (frame size).

We map enhancement of temporal features to higher frame rates, and predominance of spatial features to higher spatial and amplitude frame resolution. In this way, dynamics of the scene can be more closely followed (temporal preference) and details of a scene can be better identified (spatial preference). The mapping is intuitive and relies on the intrinsic architecture of video codecs currently in use in a non-intrusive manner, to facilitate the video communications. Finally, we show an example architecture within Information-Centric Networking (ICN) networking of a typical emergency scenario.

### B. Algorithm

We propose to map the perceptual semantics to a system quantified with the variable $\alpha \in [0, 1]$. $\alpha = 0$ and $\alpha = 1$ express full preference of the spatial and temporal perceptual features, respectively. Intermediate values of $\alpha$ represent weighed combinations of spatial and temporal preferences.

We denote the feasible set of finite values of frame rate, as $F_T(r_{APP})$, while $F_S(r_{APP})$ is the feasible set for the spatial factors, both a function of video coding rate $r_{APP}$. Note that higher frame rates and frame sizes are possible to attain with higher $r_{APP}$ [14], hence the feasible sets $F_S(r_{APP})$ and $F_T(r_{APP})$ corresponding to higher values of $r_{APP}$ will contain more number of possible values that can be chosen from. For example, in the case of scalable video coding, if temporal dyadic scalability is performed, the available values of frame rate contained in $F_T(r_{APP})$ would be the base layer frame rate and the frame rates from enhancement layers that would add up to i.e. a full 30Hz frame rate if $r_{APP}$ is sufficient: $F_T(r_{APP}) = \{$3.75Hz, 7.5Hz,15Hz,30Hz$\}$.

In order to choose the appropriate value of frame rate and resolution according to our mapping of perceptual semantics, we formulate the following optimization function:

$$\begin{aligned}(r^*_{fr}, s^*_{fr}) &= \max\ (\alpha \bar{r}_{fr} + (1-\alpha)\bar{s}_{fr}) \\ s.t.\ & \ r_{fr} \in F_T(r_{APP}) \text{ and } s_{fr} \in F_S(r_{APP})\end{aligned} \qquad (1)$$

where $\bar{r}_{fr} = r_{fr}/r^{max}_{fr}$, and $\bar{s}_{fr} = s_{fr}/s^{max}_{fr}$ are the normalized values of frame rate $r_{fr}$ and spatial/amplitude resolution

 ISBN: 978-1-61208-368-1 

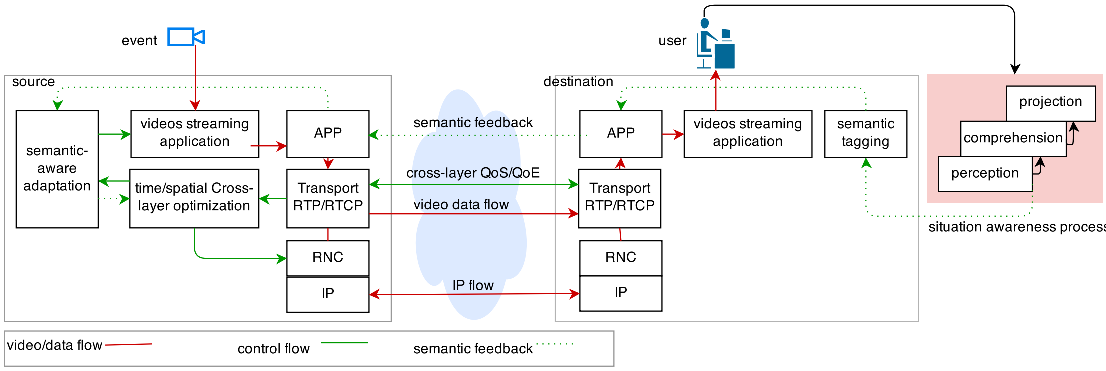



Figure 2. Block diagram proposed cross-layer framework and APP-to-APP perceptual semantics loop

$s_{fr}$ with respect to maximum available values set for the application. Note that the optimization in (1) can be applied to single layer video coding or scalable video coding.

### C. Implementation and compliance with standards

Figure 2 shows the diagram of the proposed framework.

*1) Cross-layer framework:* We assume an underlying standard cross-layer framework, that uses transport layer feedback and interacts at transport-APP layers and transport-network layers of the IP protocol stack.This framework provides the application layer rate $r_{APP}$ that can be used by the codec (such that the video coding rate equals the application layer rate), for an on-the-fly adaptive video subject to network constraints. Moreover, it provides to the network layer the necessary parameters to perform forward erasure protection. The cross-layer optimization is handling feedback with the standard RTP/RTCP protocol [15] (Real Time Protocol/Real Time Control Protocol). Note that we have assumed Forward erasure protection being performed at network layer, in particular using Random linear Network Coding (RNC).

The cross-layer optimization has been designed such that it copes with the network impairments that directly affect negatively the spatial/temporal aspects of video and is therefore QoE-driven. This time/space cross-layer optimization associates congestion with temporal impairments in video playback such as freezes. In addition, it associates erasures with artifacts degrading video quality. Further, we keep in mind that higher video quality is achieved with higher video codec rate, $r_{APP}$.

The above assumption is relevant in the design, given that the cross-layer optimization is able to mitigate the negative effects of network degradations due to congestion and erasures. Therefore, network degradations will minimally affect the job of the perceptual semantics when enhancing the temporal and spatial features necessary for situation awareness. We will further discuss this with the numerical results in Section IV.

*2) Perceptual semantics:* Figure 2 further shows how the cross-layer optimization is integrated with the perceptual semantics loop. The video streaming application uses a state-of-the-art codec such that the frame rate, frame size and codec rate can be configured on-the-fly, either as a single layer or a scalable layer coding.

In order to facilitate the perceptual semantics role, we use a return path to send the tags chosen by the user according to the perceptual semantics. The semantics-aware adaptation block in Figure 2 interprets the semantic tags coming from the end-user by mapping it to the proper decisions and forwarding to the video codec, as explained in Section III.

Following the trends in current network architectures, we propose to use semantic web protocols to enable the APP-to-APP cross talk of the semantic tagging [10]. At the transport layer, the application-specific information can be encapsulated into RTCP feedback packets compliant with the extended reports defined in RFC4585. This way, the perceptual semantics feedback loop is coherent with the cross-layer optimization.

*3) Coherence with ICN networking:* Considering future architectures, our framework complies with a semantic information-based network [16]. The aim of the Information-Centric Networking (ICN) approach is to integrate content delivery as a native network feature, where focus is not on the network as an enabler of communication links but as a platform for information dissemination. ICN could allow for future enhancements to the perceptual semantics as proposed in this paper. In particular, our approach is coherent to the receiver-driven nature of ICN. Further, caching, one of the appealing attributes of ICN in data delivery, could enable more actions at intermediate nodes concerning the incoming video stream. The philosophy of ICN by which content information is available to network/forwarding layers will allow the semantic loop we have created to trigger further actions at these intermediate nodes, such as adaptive network coding to enhance last-mile network reliability.

Figure 3 shows the topology of our framework mapped to the publish/subscribe ICN architecture for live streaming by Tsilopoulos et al. [17]. It is a typical example of an emergency application over a satellite access network, whose gateway can be mapped to the *rendevouz* (RN) and *topology manager* (TM) nodes. The publisher (operator in the ground during an emergency) announces that it has a publication available to the RN node. The subscriber (end-user/decision maker) issues a subscription, as he is interested in obtaining live feed of the current on-going events of the emergency. The RN and TM nodes find the publisher and resolve the publisher/subscriber path. The subscriber can issue petitions or unsubscribe, and in our framework, issue perceptual semantics tagging, which the publisher will receive through the RN nodes.

 ISBN: 978-1-61208-368-1 

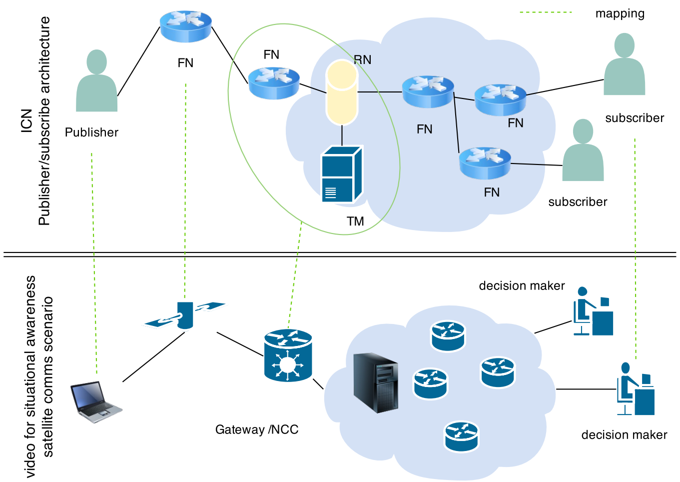



Figure 3. Publish/Subscribe architecture suitable for our proposed perceptual semantics framework

TABLE I. FEASIBLE SETS CONSIDERED FOR SIMULATION

| $r_{APP}$ (in kbps) | Feasible Set $F_T$ | Feasible set $F_S$ |
|---|---|---|
| $r_{APP} \leq 64$ | {3.75, 7.5,10,15} | {QCIF} |
| $64 < r_{APP} \leq 192$ | {3.75, 7.5,10,15} | {QCIF,CIF} |
| $192 < r_{APP} \leq 384$ | {3.75, 7.5,10,15} | {CIF,QCIF} |
| $384 < r_{APP} \leq 500$ | {3.75, 7.5,10,15,30} | {QCIF,CIF,640x360} |

## IV. SIMULATION RESULTS

We simulate a realistic scenario typical of emergency operations, where mobile satellite services are used to upstream live video from field, to proper decision makers remotely located. To our knowledge, there is no similar framework in the literature to match our proposed perceptual semantics framework and hence comparison to solutions that do not have such aim would be unfair. Therefore, our results are compared to not having such kind of framework.

### A. Setup

We use a simulation system that allows to test the proposed framework shown in Figure 2. The video streaming application is simulated by generating packets of size $l$ encoded at a rate $r_{APP}$ and frame rate $r_{frame}$.

*1) Cross-layer optimization setup for congestion and erasures:* This block receives as inputs the feedback from the receiver on current network conditions, and outputs the rate $r_{APP}$ that the video streaming application is allowed to use, and the code rate $\rho$ to be used for erasure correction, such that the transmission rate is $R = {}^{r_{APP}}/_{\rho}$.

The transmission rate is online optimized through a QoE delay-driven optimization at the sender side that uses receiver feedback, as the one proposed in [2]. The resulting discrete rate control update is given by (2)

$$R(t_{k+1}) = R(t_k) + f(\tau(t_k - \tau_D)) \tag{2}$$

where $f(\cdot)$ is a function of the delay $\tau$ measured at time $t_k - \tau_D$, a delayed value due to the propagation delay $\tau_D$ in the feedback loop. Updates on network measurements are received every $T_{samp} = t_{k+1} - t_k$ seconds.

Further, our additional novelty to cope with erasures, is the use of adaptive network coding with Systematic Random linear Network Coding, (SRNC). We use SRNC due to similar performance to optimal forward erasure correction codes like Reed-Solomon [5], but higher flexibility and compliance with future network-coded networks. For a rate budget given by $R$ in (2), the code rate $\rho = r_{APP}/R$, chosen for SRNC is maximized such that the performance meets a target residual erasure rate given the current erasure rate $\epsilon$ of the network. Hence the application layer rate $r_{APP}$ is maximized.

*2) Network simulation:* We simulate a network as a FIFO finite queue of available rate $r_{av}$ with erasure rate $\epsilon$. Simulated packets are transmitted at the obtained rate $R$.

SRNC uses the allocated code rate $\rho$ to meet the complete budget rate $R$, such that its performance meets the residual erasure rate $\epsilon^{res}$.

Congestion events are simulated as a drop (step-like) in maximum available rate $r_{av}^{max}$ to $r_{av}^{min}$ that occurs halfway through one streaming session, at $T/2$ such that $\eta = \frac{r_{av}^{max} - r_{av}^{min}}{r_{av}^{max}}$, with $\eta \in (0, 1]$. (Higher $\eta$ means higher congestion). Each simulation, corresponding to one streaming session, lasts 300s, one corresponding value of $\eta$ and $\epsilon$.

The values used correspond to a realistic satellite network commonly used in emergency operation, operating in the L-band offering up to 500kbps uplink in best effort mode.

*3) Perceptual semantics:* We model the user's semantic tagging from temporal/spatial features with the parameter $\alpha$. $\alpha$ may vary over time throughout one single streaming session, such that the sender is receiving feedback of this changes and will adapt to them using, e.g., (1). We assume these tags are changed by the user with a period of at least 10s. Three cases are considered for variation of semantic tagging, namely, $TAG_T$: only temporal tagging for the entire session, $TAG_S$: only spatial tagging, $TAG_{TS}$: alternating tags, each of 10s.

Table I summarizes the feasible sets for values of frame rate dependent on $r_{APP}$, in order to solve the algorithm in (1). The values chosen correspond to typical feasible combinations in current state-of-the art codecs.

### B. Metrics

The following metrics relate to the effects of the network constraints in terms of Quality of Experience.

*1) $QoE_A$. :* This metric is related to degradations due to erasures in the network, that cause artifacts in the image: $QoE_A = 1 - \bar{p}$, where $\bar{p}$ is the average packet loss rate at the receiver. $QoE_A \in [0, 1]$.

*2) $QoE_F$. :* This metric is related to degradation due to congestion, that cause freezes in video playback. $QoE_F = 1 - \bar{f}$ where $\bar{f}$ is the probability of freezes occurring in the playback. A freeze is the event where the time elapsing between two consecutive frames displayed exceeds a tolerated threshold. $QoE_F \in [0, 1]$.

*3) $\hat{\alpha}$ and $\Delta_\alpha$.:* These metrics are related to the performance of the adaptation through perceptual semantics. We measure the value achieved by the algorithm as $\hat{\alpha}$, and the mean absolute error with respect to the user's demanded $\alpha$, as $\Delta_\alpha = |\hat{\alpha} - \alpha|$.

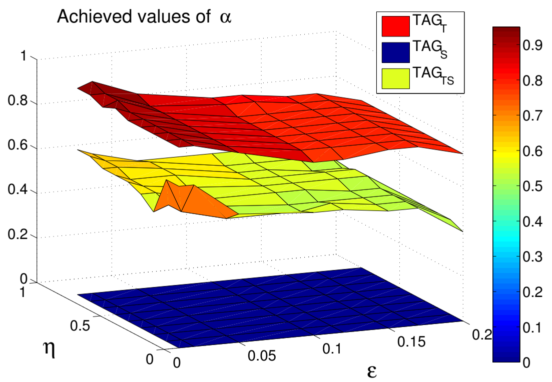



Figure 4. Achieved values of $\alpha$ vs. $\eta$ vs. $\epsilon$, using cross-layer optimization

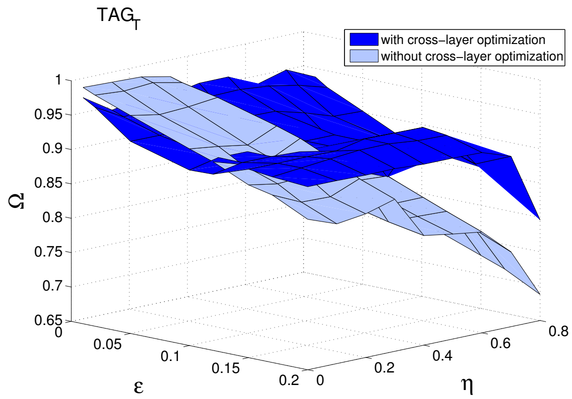



Figure 6. $\Omega$ metric for $TAG_T$

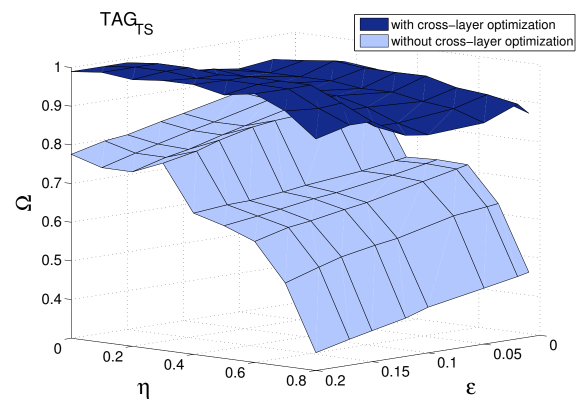



Figure 7. $\Omega$ for time-varying semantic tagging $TAG_{TS}$

*4) $\Omega$.:* Combined metric to measure tradeoffs of using perceptual semantics with and without cross-layer optimization. It is defined as:

$$\Omega = w_1 \cdot QoE_A + w_2 \cdot QoE_F + w_3 \cdot (1 - \Delta_\alpha)$$

with $w_1 + w_2 + w_3 = 1$. $\Omega \in [0, 1]$. The best performance, i.e., $\Omega = 1$, occurs when no losses degrade the video ($QoE_A \rightarrow 1$), freezes in playback are minimal ($QoE_F \rightarrow 1$) and the perceptual semantic adaptation matches the one requested by the user ($\Delta_\alpha$).

### C. Results

*1) Perceptual semantics with and without cross-layer optimization:* Figure 4 shows the performance with respect to metric $\alpha$ of the perceptual semantics together with the cross-layer optimization, as a function of congestion drops, $\eta$, and erasures $\epsilon$. Each surface corresponds to one of the three cases of time varying semantic tagging. $TAG_T$ achieves high values of $\alpha$ close to the tagged from the user, representative of preference on temporal features, while $TAG_{TS}$ offers an intermediate values, corresponding to the alternating tags. $\alpha$ only reflects on the performance of the perceptual semantics algorithm, and whether it is capable to achieve the expected user demand. However, it does not reflect the effects on $QoE_F$ and $QoE_A$, directly affected by network degradations. $\Omega$ will express the full performance as a whole.

In order to observe the combined effects of the adaptation through perceptual semantics with an underlying cross-layer optimization, we observe the individual metrics. The comparison is made between using cross-layer optimization to cope with the network constraints, or no use of it.

Figure 5a shows the value of $\Delta_\alpha$ as a function of $\eta$ and $\epsilon$. In order to achieve high QoS and QoE with the cross-layer optimization, the application layer rate $r_{APP}$ is sacrificed, as more rate is needed to protect from network erasures. Hence, the feasible set of frame rates is reduced, and the obtained $\alpha$ can not achieve the highest expected value. This can be observed with higher values of $\Delta_\alpha$ as $\epsilon$ increases.

Nevertheless, the cross-layer optimization is guaranteeing very low packet losses, as Figure 5b shows, which translates into minimal artifacts in the video. Hence, while seemingly $\Delta_\alpha$ is not as low as expected, the user is guaranteed a seamless video playback.

Figure 6 shows the combined metric $\Omega$, where the above trade-off result into higher performance when using cross-layer optimization in combination with the perceptual semantics loop, especially for highly degraded networks.

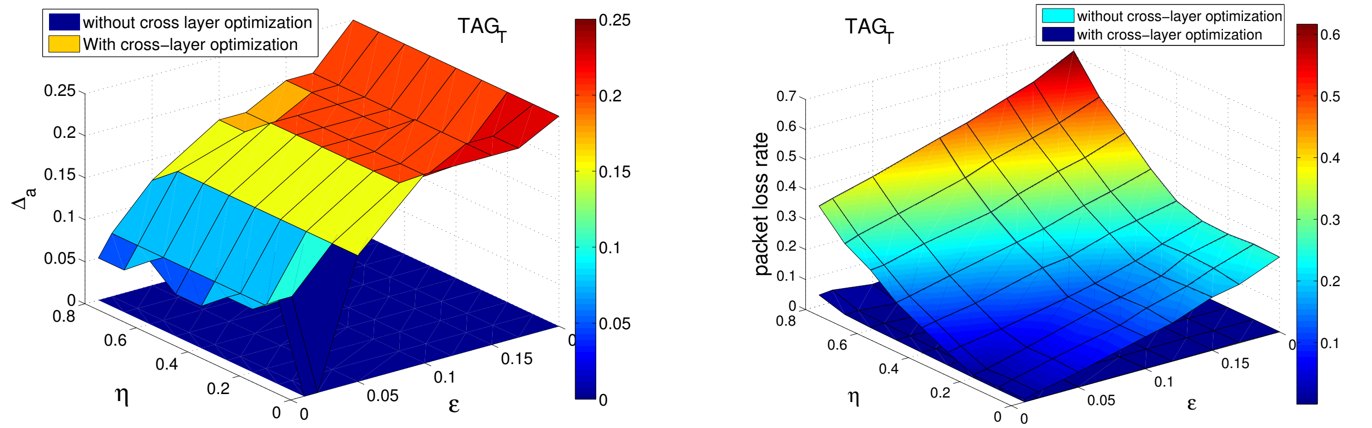
*2) Time varying perceptual semantics tagging:* We analyze the effects of time-varying perceptual tagging, representing a realistic case where the user identifies different situations that demand attention towards temporal or spatial features. These variations are represented as alternations of temporal and spatial tagging. Figure 7 shows the performance in terms of the combined metric $\Omega$.

In addition to achieving the expected $\alpha$ demanded through the semantic tagging, the performance is above 80% regardless of the degradations of the network, thanks to the cross-layer optimization. The performance is highly degraded due to congestion, as well as erasures when no cross-layer optimization is used, with performance dropping to 40%. In conclusion, Figure 7 shows that the cross-layer optimization preserves the perceptual semantics.

## V. Conclusion and Future Work

We have presented in this paper a framework where we introduced perceptual semantics for video adaptation. Percep-

 ISBN: 978-1-61208-368-1 




(a) $\Delta_\alpha$ for case 1

(b) $\bar{p}$

Figure 5. $\Delta_\alpha$ and $\bar{p}$ for $TAG_T$

tual semantics are used to acknowledge the user's demand in the context of situation awareness, where special attention is required when using video as means to perceive, comprehend and project ongoing situations, in particular for emergency scenarios. We have presented a novel model for perceptual semantics, based upon these demands, and propose a framework to be integrated into a video adaptive solution, for non-computer aided situation awareness. We discussed how to practically implement perceptual semantics into an adaptive loop that works with an underlying cross-layer optimization in charge of coping with network constraints typical of best effort wireless scenarios. Further, we have shown an adaptive algorithm that translates the perceptual semantics into temporal and spatial resolutions at codec level. Finally, our framework is contextualized for information-centric-networking. Our simulation results show how the perceptual semantic tagging achieves the expected user demands while the underlying cross-layer optimization preserves such performance. Future work includes extensions of perceptual semantics in the ICN context. Moreover, we will study more pertinent QoE metrics to match user's satisfaction when using perceptual semantics. Finally, the presented framework will be further developed for practical usage to implement a potential prototype.

## References


[1] O. Habachi, Y. Hu, M. van der Schaar, Y. Hayel, and F. Wu, "MOS-based congestion control for conversational services in wireless environments," Selected Areas in Communications, IEEE Journal on, vol. 30, no. 7, 2012, pp. 1225–1236.

[2] M. A. Pimentel-Niño, P. Saxena, and M. A. Vázquez-Castro, "QoE Driven Adaptive Video with Overlapping Network Coding for Best Effort Erasure Satellite Links," in 31st AIAA International Communications Satellite Systems Conference, Florence,Italy, 2013.

[3] H. Seferoglu, A. Markopoulou, U. C. Kozat, M. R. Civanlar, and J. Kempf, "Dynamic FEC algorithms for TFRC flows," Trans. Multi., vol. 12, no. 8, Dec. 2010, pp. 869–885.

[4] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, "Multimedia Streaming via TCP: An Analytic Performance Study," ACM Trans. Multimedia Comput. Commun. Appl., vol. 4, no. 2, May 2008, pp. 1–22.

[5] P. Saxena and M. A. Vázquez-Castro, "Network Coding Advantage over MDS Codes for Multimedia Transmission via Erasure Satellite Channels," in Personal Satellite Services (PSATS), 2013 International Conference on, October 2013, pp. 199–210.

[6] R. Immich, E. Cerqueira, and M. Curado, "Adaptive video-aware fec-based mechanism with unequal error protection scheme," in Proceedings of the 28th Annual ACM Symposium on Applied Computing, ser. SAC '13. New York, NY, USA: ACM, 2013, pp. 981–988.

[7] J. Thomas-Kerr, C. Ritz, and I. Burnett, "Semantic-aware delivery of multimedia," in Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on, Sept 2009, pp. 1498–1503.

[8] C. Henson, A. Sheth, and K. Thirunarayan, "Semantic perception: Converting sensory observations to abstractions," Internet Computing, IEEE, vol. 16, no. 2, 2012, pp. 26–34.

[9] A. Cavallaro and S. Winkler, "Perceptual semantics," in Multimedia Technologies: Concepts, Methodologies, Tools, and Applications, S. M. Rahman, Ed. Information Science Reference, Jun. 2008, ch. 7.5, pp. 1441–1455.

[10] S. B. Kodeswaran and A. Joshi, "Content and context aware networking using semantic tagging," in Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on. IEEE, 2006, pp. 77–77.

[11] J. Strassner, J. Betser, R. Ewart, and F. Belz, "A Semantic Architecture for Enhanced Cyber Situational Awareness," in Secure and Resilient Cyber Architectures Conference, 2010.

[12] M. R. Endsley, "Theoretical underpinnings of situation awareness: A critical review," Situation awareness analysis and measurement, 2000, pp. 3–32.

[13] S. S. Krupenia, C. Aguero, and K. C. Nieuwenhuis, "The value of different media types to support command and control situation awareness," in Proceedings of 9th International ISCRAM Conference, 2012, pp. 1–5.

[14] Z. Ma, F. Fernandes, and Y. Wang, "Analytical rate model for compressed video considering impacts of spatial, temporal and amplitude resolutions," in Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on, July 2013, pp. 1–6.

[15] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications, IETF RFC 3550," United States, 2003.

[16] H. Wirtz and K. Wehrle, "Opening the loops - towards semantic, information-centric networking in the internet of things," in Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Conference on, June 2013, pp. 18–24.

[17] C. Tsilopoulos, G. Xylomenos, and G. Polyzos, "Are Information-Centric Networks Video-Ready?" in Packet Video Workshop (PV), 2013 20th International, Dec 2013, pp. 1–8.