

# IMPACT OF VIDEO TRANSCODING ARTIFACTS ON THE SUBJECTIVE QUALITY

*Lutz Goldmann, Francesca De Simone,  
Frederic Dufaux, Touradj Ebrahimi*

Ecole Federale Polytechnique de Lausanne  
EPFL/STI/IEL/GR-EB, Station 11  
1015 Lausanne, Switzerland

*Rudolf Tanner,  
Mauro Lattuada*

Abilis Systems Sarl  
Chemin des Aulx 18  
1228 Plan-Les-Ouates, Switzerland

## ABSTRACT

Video transcoding is an important step to enable interoperability between different networks, terminals, applications, and services for video communication. This paper studies the influence of typical video transcoding artifacts due to frame rate reduction and drift error on the subjective quality. Given a realistic dataset for a DVB-T to DVB-H transcoding scenario, the subjective quality before and after the transcoding is compared against each other. In order to quantify the influence of both artifacts, a pixel domain and an open loop transcoding solution have been considered. Since the strength of both artifacts depends largely on the initial encoding parameters, additional experiments have been conducted to quantify the influence of the distance between I frames and the number of consecutive B frames on the subjective quality.

**Index Terms**— video transcoding, quality of experience, subjective quality, drift error, jerkiness

## 1. INTRODUCTION

With the increasing diversity in communication technologies (networks, terminals, applications, and services), video transcoding has become an important step to enable adaptability and interoperability [1]. In general, video transcoding refers to the conversion of a compressed video stream into another compressed video stream with different characteristics.

In most applications, video transcoding can be seen as an additional step within the video processing chain which includes acquisition, coding, transmission, and restitution. Therefore, its impact on the quality of experience should be as small as possible. However, depending on the architecture, video transcoding introduces very specific artifacts which may degrade the quality considerably.

In the past most of the developed video transcoding solutions [2] have been evaluated based on objective metrics,

such as PSNR, that may not correlate very well with the quality perceived by human observers. Furthermore, most of these metrics do not consider the specific nature and interaction of transcoding artifacts.

The goal of this paper is to quantify the impact of typical transcoding artifacts due to drift error and frame rate reduction on the subjective quality. This knowledge can be exploited to optimize transcoding algorithms or develop objective quality metrics to predict the influence of video transcoding on the perceived quality. Given a DVB-T to DVB-H transcoding scenario, a cascaded pixel-domain and an open-loop architecture for MPEG-2 [3] to MPEG-4 [4] transcoding are considered.

## 2. VIDEO TRANSCODING

Video transcoding refers to the conversion of a compressed video stream into another compressed video stream and may provide the following functionalities [5]:

- Format conversion
- Bit rate reduction
- Spatial resolution reduction
- Temporal resolution reduction
- Error resilience

While a homogeneous transcoder performs a conversion between video streams of the same format, a heterogeneous transcoder is used to convert between different video formats.

### 2.1. Application scenario

Video transcoding is applied in a digital video broadcast (DVB) scenario [6], where a source video stream from a DVB-T signal is transcoded to a target video stream suitable for DVB-H enabled devices, such as mobile phones. Since the transcoding has to be performed in real-time on a mobile device with limited memory and computational power, an appropriate trade-off between computational complexity and visual quality has to be achieved.

Table 1 provides an overview of the video coding related aspects in DVB-T and DVB-H including used video codecs,

---

The research leading to these results has been performed within the framework of CTI project 10571.1 PFES-ES (Advanced Transcoding Technology).

Criteria	DVB-T	DVB-H
Resolution	720x576, 1920x1080	352x288, 704x576
Frame rate	25 fps, 50 fps	25 fps
Scanning	Interl., progr.	Progressive
Codec	MPEG-2, H.264	MPEG-4, H.264
Profile/level	MP@ML	SP, ASP
Bit rate	3 – 5 Mbps	300 – 500 kbps
GOP structure	$N = 3 - 30 / M = 2 - 3$	-

**Table 1:** Overview of the video coding aspects in DVB-T and DVB-H.

supported resolutions, frame rates and bit rates.

Given these characteristics, a heterogeneous transcoder is required which converts the MPEG-2 encoded video stream from DVB-T to an MPEG-4 encoded video stream compatible to DVB-H. Since the transcoding will be performed directly on the mobile device, bit rate reduction is not required. Spatial resolution reduction is not considered, due to increased display resolutions (up to 800x480 pixels) offered by recent mobile phones. However, since transcoding of B frames requires a larger buffer, they will be dropped by the transcoder leading to a temporal resolution reduction.

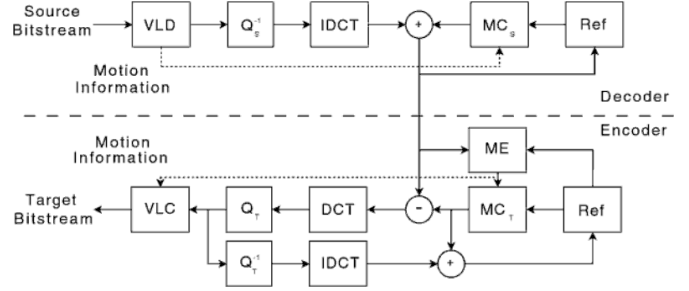
## 2.2. Transcoding architectures

Existing transcoding methods can be roughly grouped into 3 different categories [5]:

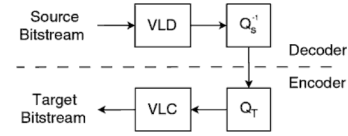
**Cascaded pixel domain transcoder:** This most straightforward approach is to cascade the decoder and encoder directly, as shown in figure 1. The source video stream is fully decoded and then encoded to the target video stream considering the desired format, bit rate, frame rate and spatial resolution. Due to the encoding of the video stream, pixel domain transcoding is the most complex approach but also leads to the highest possible quality.

**Open loop transcoder:** This is the least complex transcoding architecture, as shown in figure 2. It translates the source bitstream directly into the target bitstream by applying variable length decoding to obtain the macroblock information, requantizing the DCT coefficients to meet the target bitrate, remapping the motion vectors and macroblock types, and applying variable length encoding. Since this may introduce a mismatch between the prediction before and after the transcoding, open loop transcoding approaches are subject to error accumulation which is commonly referred to as drift error.

**Closed loop transcoder:** This transcoding architecture aims at eliminating the mismatch between the predictions by approximating the cascaded pixel domain transcoder. This is achieved through a single feedback loop which compensates the error introduced by the initial predic-



**Fig. 1:** Architecture of pixel domain transcoding [2].



**Fig. 2:** Architecture of open loop transcoding [2].

tion. Closed loop architectures provide a good trade-off between computational complexity and visual quality.

In order to study the range of possible quality degradations caused by the video transcoding, only the cascaded pixel domain and the open loop transcoding architectures are considered here. In both cases only I and P frames are transcoded while B frames are skipped leading to a possible frame rate reduction.

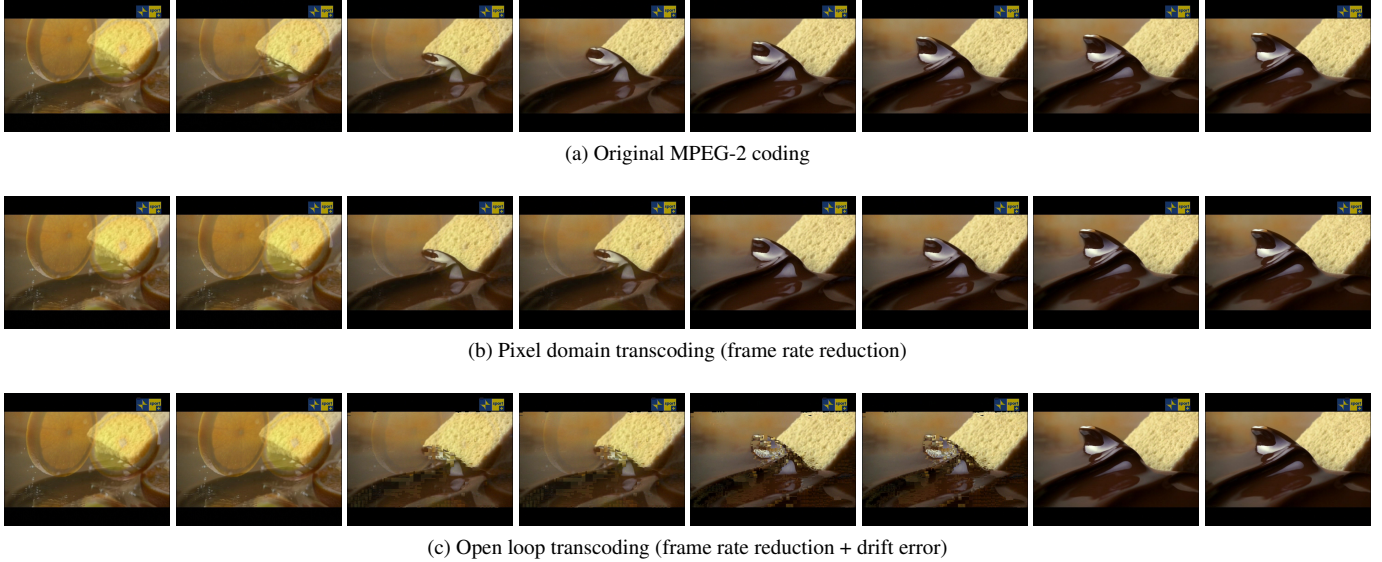
## 2.3. Transcoding artifacts

For any multimedia service, the quality of experience (QoE) for the user is directly influenced by the different stages within the multimedia processing chain. Considering a video stream within DVB-T or DVB-H, artifacts may be introduced by the acquisition, coding, transmission, and restitution.

While the initial encoding may lead to typical video coding artifacts, such as blocking and blurring, the transcoding may introduce very specific artifacts which are caused by the drift error and frame rate reduction. Since the goal of this work is to assess their influence on the overall quality, they will be discussed in more detail below.

### 2.3.1. Frame rate reduction

Depending on the parameters of the MPEG-2 encoder, the source video stream may contain a different amount of consecutive B-frames. Dropping them in the transcoding leads to a decreased frame rate in comparison to the original as it is shown in figure 3b. If the frame rate becomes too low, motion is not perceived as continuous anymore, which is referred to as jerkiness.



**Fig. 3:** Visual sample illustrating the typical transcoding artifacts of pixel domain and open loop transcoding in comparison to the original MPEG-2 coding. The frame rate reduction for both transcoding approaches leads to less fluid motion (jerkiness). The drift error in the open loop approach causes local blocking artifacts with increasing strength over time.

### 2.3.2. Drift error

The drift error is due to a mismatch between the prediction and the residual in the motion compensated prediction which may be caused by a combination of requantization, DCT coefficient elimination, arithmetic errors and the non-commutative property of the motion compensation. Regardless of the magnitude of this error for a single frame, it will accumulate over consecutive P frames. As it can be seen from figure 3c, it causes local blocking artifacts. The strength of the artifact increases gradually for consecutive P frames until the corresponding macro blocks are intra coded.

## 3. SUBJECTIVE QUALITY EVALUATION

### 3.1. Datasets and experiments

Two different datasets were used to perform the subjective quality evaluation. The DVB-T dataset contained captured DVB-T streams in order to assess the overall transcoding quality for realistic content and coding parameters. The MPEG dataset contained raw video sequences that were encoded with different MPEG-2 coding parameters in order to assess their influence on the transcoding quality.

#### 3.1.1. DVB-T dataset

The DVB-T dataset consisted of ten MPEG-2 encoded video sequences that were extracted from a large variety of captured DVB-T broadcasts. The original transport streams (TS) were cut into representative sequences with a resolution between 480x576 – 720x576 pixels, a frame rate of 25 fps and a length



**Fig. 4:** Representative frames of the test sequences within the DVB-T dataset. From top left to bottom right: cartoon (CA), commercial (CO), fashion (FA), nature (NA), programme (PR), quiz (QU), series (SE), soccer (SO).

of 200 frames. The parameters of the video coding were the following: main profile at main level (MP@ML), a bit rate in the range 3 – 15 Mbps, distance between I frames (N) 3 – 33, number of consecutive B frames (M) 2 – 3. The dataset was split into a training and a test set with 2 and 8 sequences, respectively. Representative frames of the test sequences are shown in figure 4.

Based on this dataset the first experiment aimed at assessing the subjective quality of the video transcoding on a realistic dataset. The 8 test sequences were transcoded with both the pixel domain and open loop transcoding approach, leading to a set of  $8 \times (1 + 2) = 24$  test stimuli, including the MPEG-2 encoded source sequences.



**Fig. 5:** Representative frames of the test sequences within the MPEG dataset. From left to right: city, harbour, ice, soccer.

### 3.1.2. MPEG dataset

The MPEG dataset contained five raw video sequences with a resolution of 704x576 pixels, a frame rate of 30 fps and a length of 300 frames. Each sequence contained only a single scene without any camera movement. The dataset was split into a training and a test set with 1 and 4 sequences, respectively. Representative frames of the test sequences are shown in figure 5. The sequences were all MPEG-2 encoded at MP@ML and a bit rate of 5 Mbps.

Based on this dataset, the goal of the second experiment was to assess the influence of I frame distance (N) on the subjective quality. Since the drift error grows incrementally until the next I frame, an increasing N will lead to more spatial artifacts and subsequently to a decrease of the video quality. Therefore, the distance between I frames (N) was varied in the set {1,2,5,10,15} and the number of consecutive B frames (M) was fixed to 1. Thus, considering the MPEG-2 coded sequences and both the pixel domain and the open loop transcoded sequences, a set of  $4 \times 5 \times 3 = 60$  test sequences was obtained.

Additionally, a third experiment was performed to assess the influence of the number of consecutive B frames (M) on the subjective quality. Since both transcoding approaches do not consider the B-frames within a MPEG-2 stream, M directly influences the temporal resolution (framerate) of the MPEG-4 transcoded video. Therefore, N was fixed to 12 and M varied in the set {1,2,3}, obtaining a set of  $4 \times 3 \times 3 = 36$  test sequences.

## 3.2. Environment and subjects

The subjective quality tests were conducted at the MMSPG Quality Test Laboratory which is compliant to the ITU-R recommendation [7]. The walls of the room are painted in gray 128 and the controlled lightning system consists of neon lamps with 6500 K color temperature. The tests were carried out on a 30" LCD screen (Eizo CG301W) with a native resolution of 2560x1600 pixels. The screen was calibrated according to the following profile: sRGB gamut, D65 white point,  $120 \text{ cd/m}^2$  brightness and minimum black level. A picture of the test room is shown in figure 6.

The subject was seated directly in line with the center of the video display at a viewing distance equal to 3-4 times the height of the video window. Fifteen non-expert subjects (5 female and 10 male) took part in the experiments. Their age



**Fig. 6:** Standard compliant test laboratory with controlled lighting system and calibrated screen.

ranged from 24 to 31 years.

Each of the subjects has participated in all 3 experiments. The lengths of the individual experiments were 10, 18, and 12 minutes, respectively. In order to avoid fatigue of the subjects, the experiments were conducted in individual sessions with approximately an hour break in between. For each of the datasets a separate training set was used for instructing the subjects and explaining the range of qualities.

## 3.3. Methodology

Several subjective quality evaluation methodologies have been standardized by ITU for the quality assessment of audio, video and audiovisual data within different application scenarios (television, multimedia) [7].

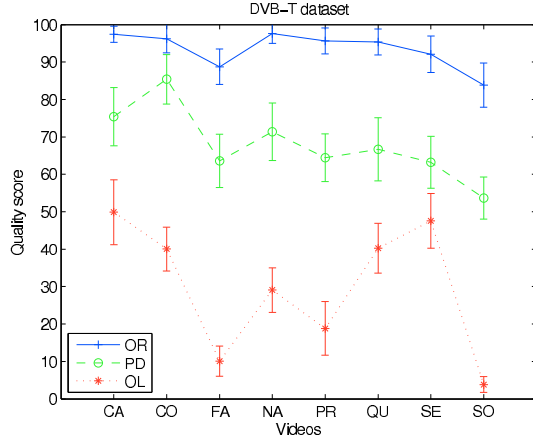
Existing methods for video quality evaluation can be compared based on the following characteristics. The rating criteria may be either quality (excellent to bad), impairment (imperceptible to very annoying) or a comparison of two video sequences (much better to much worse). Independent from the rating criteria, the rating scale may be categorical (typically 3,5,7,9,11 levels) or continuous (1-100). If a reference is available the subject may be informed about its position within the session (explicit reference) or not (implicit reference). Depending on the method, each video can be rated individually (single stimulus) or two sequences are rated together (double stimulus).

In the experiments described in this paper, a single stimulus methodology with continuous quality scale was used [7].

## 3.4. Analysis of subjective data

### 3.4.1. Outlier detection

Outlier detection or screening is performed to detect and discard the scores of unreliable subjects and improve the accuracy of the statistics. The screening of subjects was performed according to the guidelines described in [7]. No outliers were detected in any of the experiments.



**Fig. 7:** Subjective quality scores across the testing videos of the DVB-T dataset for the original MPEG-2 coded video (OR) and the MPEG-4 transcoded video using the pixel domain (PD) or open loop (OL) approach.

### 3.4.2. Mean opinion score and confidence interval

After the outlier detection, the retained scores are processed to obtain statistics which show the overall judgment by the sample of subjects representative of the entire population. Particularly, the scores obtained for each test condition are averaged in order to obtain the mean opinion score (MOS):

$$MOS_j = \sum_{i=1}^N m_{ij} / N$$

where  $N$  is the number of subjects and  $m_{ij}$  is the score by subject  $i$  for the test condition  $j$ . The confidence interval (CI) of the estimated mean for each test sequence is also computed, to quantify the relationship between the estimated mean values based on a sample of the population and the true mean values of the entire population. The 95% CI was computed using the Student's t-distribution, as follows:

$$CI_j = t(1 - \alpha/2, N) \cdot \sigma_j / \sqrt{N}$$

where  $t(1 - \alpha/2, N)$  is the t-value corresponding to a two-tailed t-Student distribution with  $N - 1$  degrees of freedom and a desired significance level  $\alpha = 0.05$ .  $N$  corresponds to the number of subjects, and  $\sigma_j$  is the standard deviation of a single test condition across subjects.

## 3.5. Results

### 3.5.1. Transcoding quality

Figure 7 provides the MOSs and CIs resulting from the first experiment, which compares the performance of the two transcoding approaches with respect to the MPEG-2 coding. The overall quality between the original and the two transcoding approaches differs considerably. While all the original

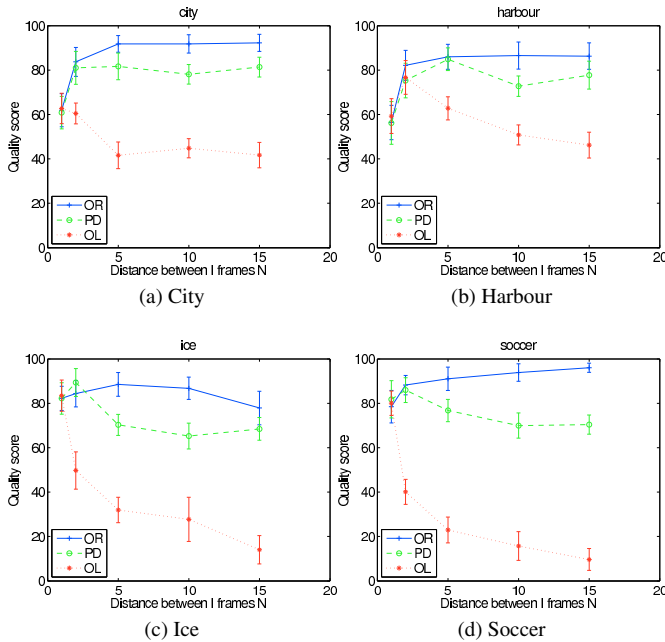
video sequences are judged as “excellent” (4.3 – 4.9), the pixel domain transcoding achieves mostly “good” quality (2.7 – 4.2) followed by the open loop transcoding between “bad” and “fair” (0.2 – 2.5). Although the subjective quality varies across the sequences for all the approaches, the variation is larger for the open loop approach than for the pixel domain approach which shows that the former is more influenced by the visual characteristics of the video. Since the pixel domain approach does not lead to noticeable spatial distortions the decreasing quality in comparison with the original video sequences can only be explained by the temporal resolution reduction due to the skipping of the B-frames. The “poor” quality of the open loop approach is caused by the combination of this temporal artifacts with the spatial artifacts due to the drift error which is inherent to this transcoding approach. While for most of the sequences the spatial artifacts seem to have the largest influence on the subjective quality (e.g. commercial, fashion, soccer), for some video sequences the temporal artifacts are dominant (e.g. quiz, series).

### 3.5.2. Distance between I frames

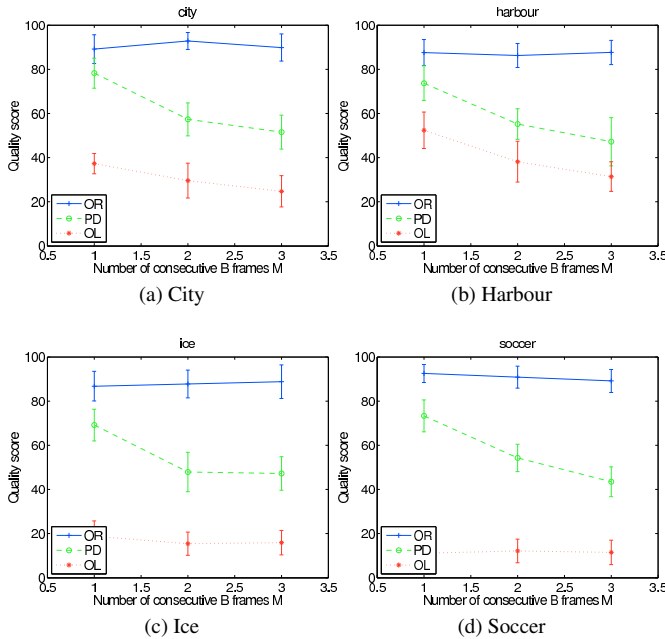
Figure 8 shows the results of the second experiment for the original MPEG-2 coding, the pixel domain and the open loop transcoding, across all the video sequences from the MPEG dataset. Several interesting observations can be made from this figure. For  $N = 1$  the subjective quality of the original video and both transcoding approaches is essentially the same due to the absence of the drift error for the open loop transcoding. However, for larger  $N$  the drift error increases and leads to a drop of the perceived quality. The loss of subjective quality depends largely on the visual characteristics of the sequences. The quality of video sequences with high spatial details and small amounts of global motion (e.g. city, harbour) suffers less than the quality of sequences with lower spatial details and high amounts of individual object motion (e.g. ice, soccer). While the quality of the pixel domain transcoding stays between “good” and “excellent” across the whole range of  $N$ , the quality of the open loop transcoding drops below “fair” for  $N > 5$ .

### 3.5.3. Number of consecutive B frames

Finally, figure 9 shows the results of the third experiment for the original MPEG-2 coding, the pixel domain and the open loop transcoding, across all the video sequences from the MPEG dataset. Several interesting observations can be made from this figure. From the figure, it can be noticed that there is virtually no influence of  $M$  on the quality of the original MPEG-2 encoded video. The influence of  $M$  on the quality of the pixel domain transcoding is much larger than for the open loop transcoding. The reason for that is that the latter suffers already from the spatial artifacts caused by the drift error and therefore the temporal resolution reduction does not have a large influence on the already bad quality. However



**Fig. 8:** Subjective quality scores vs. distance between two consecutive I frames  $N$  across the test video sequences of the MPEG dataset for the original MPEG-2 video (OR) and the MPEG-4 video transcoded using the pixel domain (PD) or open loop (OL) approach.



**Fig. 9:** Subjective quality scores vs. number of consecutive B frames  $M$  across the test video sequences of the MPEG dataset for the original MPEG-2 video (OR) and the MPEG-4 video transcoded using the pixel domain (PD) or open loop (OL) approach.

this experiment shows that both spatial and temporal artifacts must be reduced to achieve a “good” subjective quality of the transcoded video.

#### 4. CONCLUSION

Video transcoding is a very important step for universal multimedia access. In order to provide the best quality of experience for the user it is crucial to understand the influence of the transcoding on the perceived quality. Beside typical video coding artifacts such as blocking and blurring, transcoding introduces some very specific artifacts such as drift error and jerkiness. This paper presents the results of a comprehensive study on how these artifacts influence the subjective quality. Furthermore, it analyzes how specific encoding parameters (GOP size, number of B frames) influence the strength of these artifacts and thus subsequently the quality.

This work may be extended into several directions. Since most of the current objective quality metrics have been developed for general video quality assessment, it is interesting to evaluate their performance for video transcoding by measuring the correlation with the subjective quality scores. Considering the application scenario, it is very interesting to explore different transcoding approaches to find the best tradeoff between visual quality and computational complexity.

#### 5. REFERENCES

- [1] A. Vetro, C. Christopoulos, and T. Ebrahimi, “Universal multimedia access,” *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 16, 2003.
- [2] S. Moiron, M. Ghanbari, P. Assunção, and S. Faria, “Video Transcoding Techniques,” in *Recent Advances in Multimedia Signal Processing and Communications*. Springer, 2009.
- [3] ISO/IEC, “Information technology – generic coding of moving pictures and associated audio information: Video,” ISO/IEC 13818-2:2000, 2000.
- [4] ISO/IEC, “Information technology – coding of audiovisual objects – part 2: Visual,” ISO/IEC 14496-2:2004, 2004.
- [5] I. Ahmad, X. Wei, Y. Sun, and Y.Q. Zhang, “Video transcoding: an overview of various techniques and research issues,” *IEEE transactions on multimedia*, vol. 7, no. 5, pp. 793–804, 2005.
- [6] U. Reimers, *DVB: the family of international standards for digital video broadcasting*, Springer Verlag, 2005.
- [7] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R BT.500-11, 2002.