Chapter 7.5
Perceptual Semantics

Andrea Cavallaro
Queen Mary University of London, UK

Stefan Winkler
National University of Singapore and Genista Corporation, Singapore

ABSTRACT

The design of image and video compression or transmission systems is driven by the need for reducing the bandwidth and storage requirements of the content while maintaining its visual quality. Therefore, the objective is to define codecs that maximize perceived quality as well as automated metrics that reliably measure perceived quality. One of the common shortcomings of traditional video coders and quality metrics is the fact that they treat the entire scene uniformly, assuming that people look at every pixel of the image or video. In reality, we focus only on particular areas of the scene. In this chapter, we prioritize the visual data accordingly in order to improve the compression performance of video coders and the prediction performance of perceptual quality metrics. The proposed encoder and quality metric incorporate visual attention and use a semantic segmentation stage, which takes into account certain aspects of the cognitive behavior of people when watching a video. This semantic model corresponds to a specific human abstraction, which need not necessarily be characterized by perceptual uniformity. In particular, we concentrate on segmenting moving objects and faces, and we evaluate the perceptual impact on video coding and on quality evaluation.

INTRODUCTION

The development of new compression or transmission systems is driven by the need of reducing the bandwidth and storage requirements of images and video while increasing their perceived visual quality. Traditional compression schemes aim at minimizing the coding residual in terms of mean squared error (MSE) or peak signal-to-noise ratio (PSNR). This is optimal from a purely mathematical but not a perceptual point of view. Ultimately, perception is the more appropriate and more relevant benchmark. Therefore, the objective must be to define a codec that maximizes perceived visual quality such that it produces better quality at the same bit rate as a traditional encoder or the same visual quality at a lower bit rate (Cavallaro, 2005b).

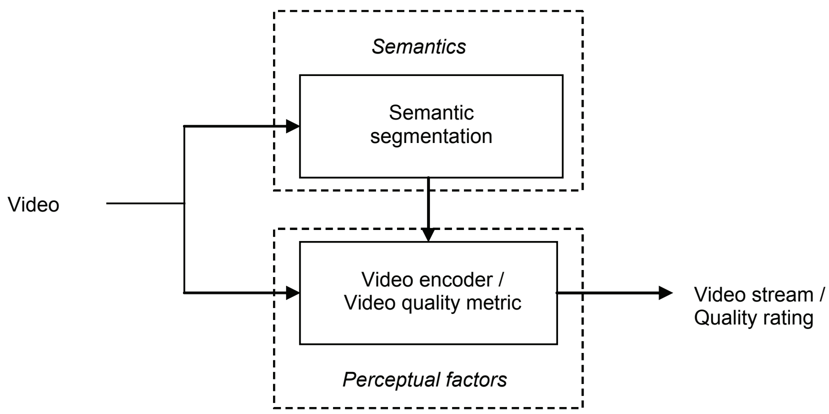
In addition to achieving maximum perceived quality in the encoding process, an important concern for content providers is to guarantee a certain level of quality of service during content distribution and transmission. This requires reliable methods of quality assessment. Although subjective viewing experiments are a widely accepted method for obtaining meaningful quality ratings for a given set of test material, they are necessarily limited in scope and do not lend themselves to monitoring and control applications, where a large amount of content has to be evaluated in real-time or at least very quickly. Automatic quality metrics are desirable tools to facilitate this task. The objective here is to design metrics that predict perceived quality better than PSNR (Winkler, 2005a).

One of the common shortcomings of traditional video coders and quality metrics is the fact that they treat the entire scene uniformly, assuming that people look at every pixel of the image or video. In reality, we focus only on particular areas of the scene, which has important implications on the way the video should be analyzed and processed.

In this chapter, we take the above observations into account and attempt to emulate the human visual system. The idea is to prioritize the visual data in order to improve the compression performance of video coders and the prediction performance of perceptual quality metrics. The proposed encoder and quality metric incorporate visual attention and use a semantic segmentation stage (Figure 1). The semantic segmentation stage takes into account some aspects of the cognitive behavior of people when watching a video. To represent the semantic model of a specific cognitive task, we decompose each frame of the reference sequence into sets of mutually- exclusive and jointly-exhaustive segments. This semantic model corresponds to a specific human abstraction, which need not necessarily be characterized by perceptual uniformity. Since the semantics (i.e., the meaning) are defined through human abstraction, the definition of the semantic partition depends on the task to be performed. In particular, we will concentrate on segmenting moving objects and faces, and we will evaluate the perceptual impact on video coding and on quality evaluation.

The chapter is organized as follows: The section "Cognitive Behavior" discusses the factors influencing the cognitive behavior of people watching a video. The section "Semantic Segmentation" introduces the segmentation stage that generates a semantic partition to be used in video coding and quality evaluation. In "Perceptual Semantics for Video Coding" and "Perceptual Semantics for Video Quality Assessment", we describe how

Figure 1. Flow diagram of the encoder and the quality metric that incorporate factors influencing visual attention and use a semantic segmentation stage



Related Content