Exploiting Temporal Redundancy of Visual Structures for Video Compression

Georgios Georgiadis Stefano Soatto

Department of Computer Science,
University of California, Los Angeles,
Los Angeles, CA 90095, USA
{giorgos,soatto}@cs.ucla.edu

We present a video coding system that partitions the scene into "visual structures" and a residual "background" layer. The system exploits the temporal redundancy of visual structures to compress video sequences. We construct a dictionary of track-templates, which correspond to a representation of visual structures. We subsequently choose a subset of the dictionary's elements to encode video frames using a Markov Random Field (MRF) formulation that places the track-templates in "depth" layers (Fig. 1). Our video coding system offers an improvement over H.265/H.264 and other methods in a rate-distortion comparison (Fig. 2) in four sequences from MOSEG[1].

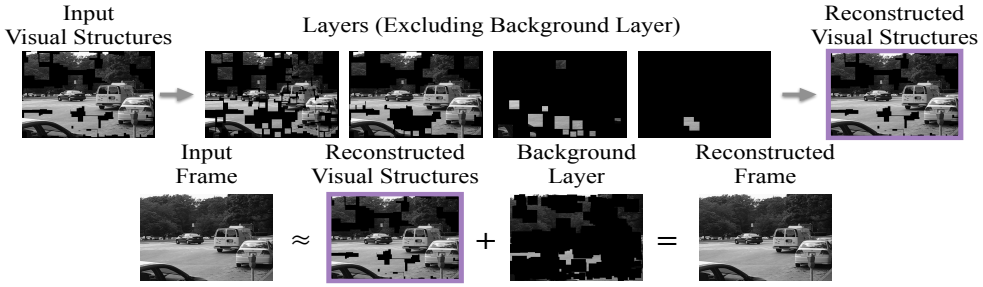

Figure 1: *Reconstructing a frame. Visual structures are decomposed into depth layers and reconciled by overlaying them. The input frame is reconstructed by adding back to the visual structures the background layer.*

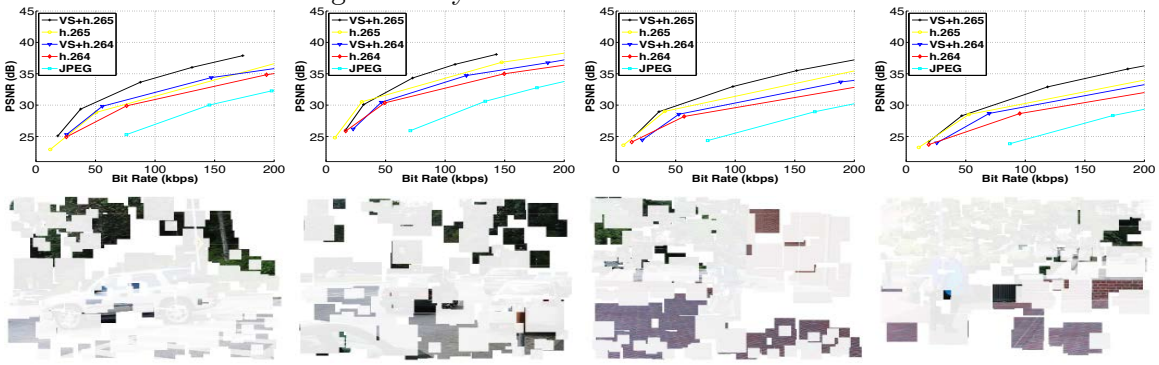

Figure 2: *Typical results on four sequences (Cars-1,2, People-1,2). Top: PSNR against bit rate. "VS+H.265" (black) and "VS+H.264" (blue) outperform H.265 (yellow) and H.264 (red) respectively. Bottom: Propagated and newly-created tracks. Non-transparent tracks correspond to tracks that are motion-predicted from previous frames. Semi-transparent tracks are tracks that start in this frame.*

[1] http://lmb.informatik.uni-freiburg.de/resources/datasets/