On the unsolved problem of
Shot Boundary Detection for Music Videos

Alexander Schindler
Information Management
Center for Digital Safety & Security
AIT Austrian Institute of Technology GmbH
alexander.schindler@ait.ac.at

Andreas Rauber
Information and Software Engineering Group
Institute of Information Systems Engineering
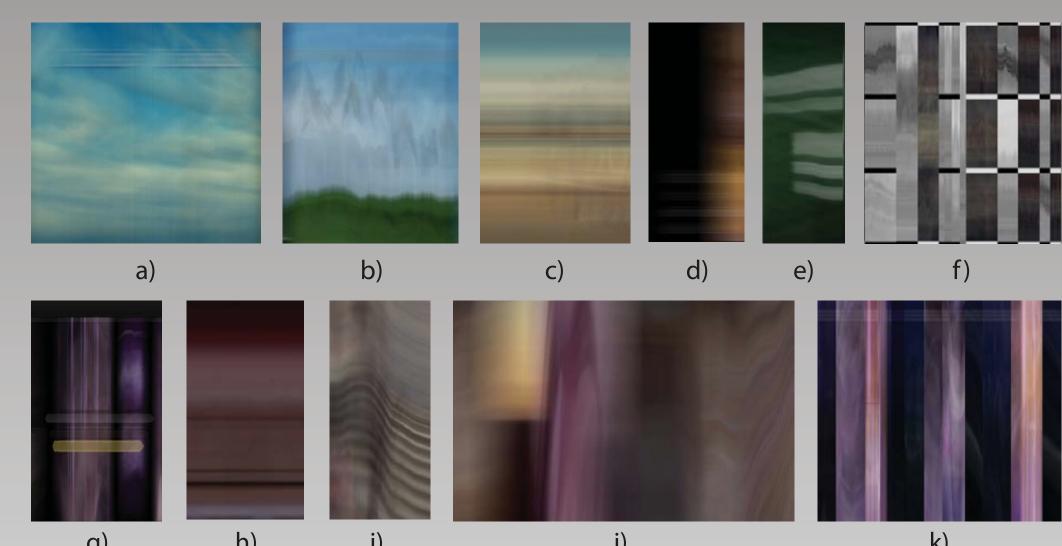Vienna University of Technology
rauber@ifs.tuwien.ac.at

Motivation

Audio-Visual approach to Music Information Retrieval

♦ Music Genre Classification, Emotion Recognition, Artist Identification
♦ Number and type of shot transition as visual feature
♦ Shots per minute (analogous to Beats per minute), Average Shot Length

Open Issues with Shot Detection in Music Videos

♦ characteristic styles in editing music videos for certain music genres
♦ style and complexity of shot transition changed over time
♦ state-of-the-art approaches only apply to the limited set of shot transition
♦ music videos utilize shot transitions in a far more artistic way
  • to create tension
  • express emotions (e.g distress, horror, melancholy)
♦ biggest challenge in shot-detection for music videos
  • vast number of transition styles
  • unconventional camera work
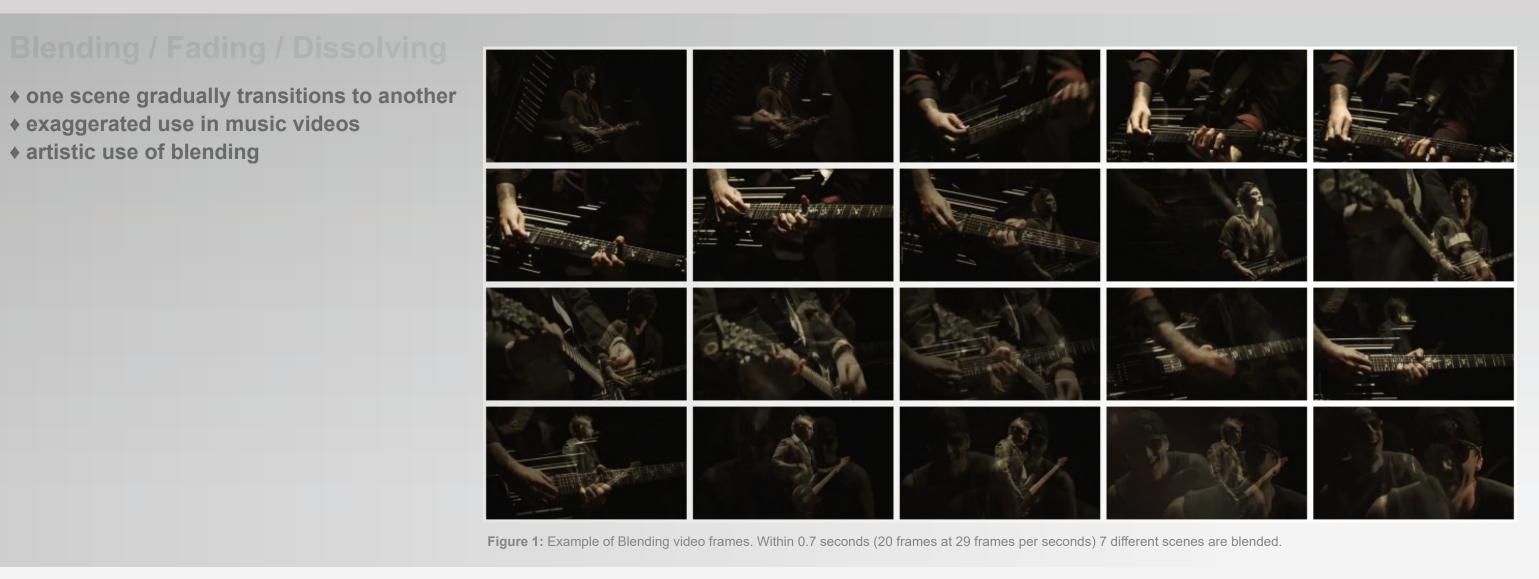 • definition of a shot-boundary
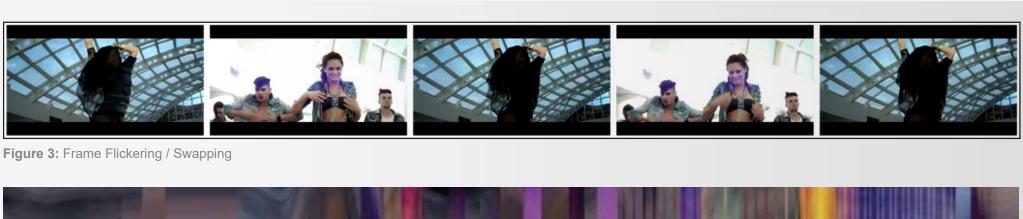
Transition Types in Music Videos

♦ Sharp Cuts
♦ Gradual Transitions
♦ Fade-In / Fade-Out
♦ Skip-Frames
♦ Jumping back and forth
♦ Abrupt tempo changes
♦ Frame-Swapping / Flickering
♦ Fast zoom in/out
♦ Abrupt focal changes

♦ Split-Screen
♦ Camera Tilts
♦ Freezing on a frame
♦ Spotlights
♦ Blending / Fading / Dissolving
♦ Overlays
♦ Distortion Effects
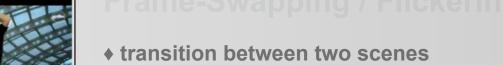♦ Dropping to black
♦ Dancing in front of Green-Box

a)   b)   c)   d)   e)   f)

g)   h)   i)   j)   k)

Skip-Frames

♦ a few video frames are skipped
♦ create rhythmic effect
♦ increase pace of visual flow

Figure 2: Example of Skip Frames. This scene was recorded in a single camera movement, starting from the left corner of the bar and panning to the right until the focus is on two women. The depicted frames show that a large segment was skipped after frame number 3 and a small segment after frame number 11. This is perceived as "jump" replaying this with 25 frames per seconds.

Frame-Swapping / Flickering

♦ transition between two scenes
♦ transition pattern:
ca few frames (milliseconds) of scene A
  • a few frames of scene B
♦ often used with build-ups in EDM
  • rescending part to create tension
  • dissolves in main theme or chorus
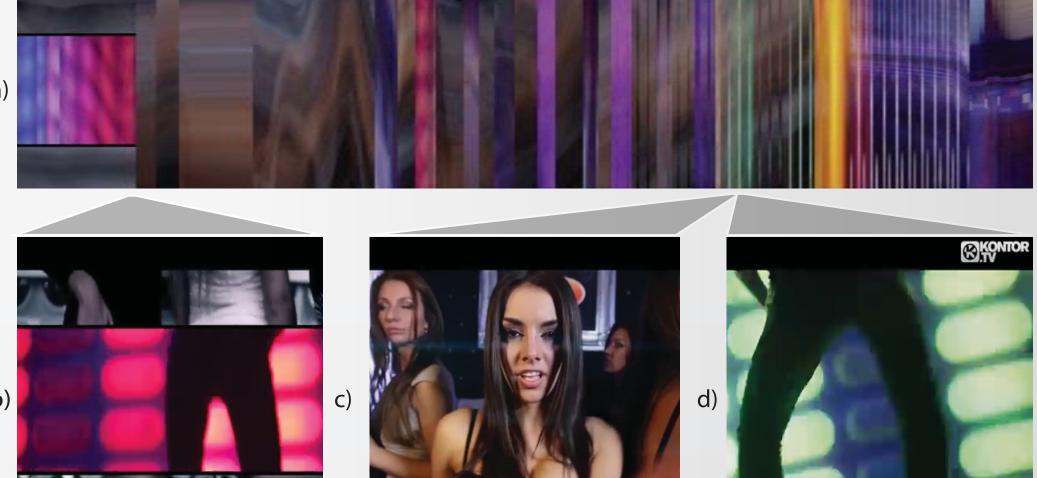

Figure 3: Frame Flickering / Swapping

Figure 4: a) Mean-color-bar to visualize music video activity over time. b) vertical split-screen section (first segment in a). b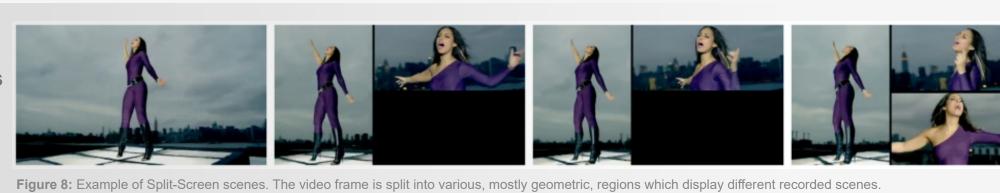) and c) in the greenish segment of a) the video swaps quickly between scene b) (darker columns) and scene c) (brighter columns).

Distortion Effects

♦ visual effects applied to video frames
♦ various types
  • heavy blurring
  • distortion
  • diffusion
  • rippling
  • white noise
♦ simulation of analog TV screen errors
  • vertical roll
  • horizontal/vertical synchronization failures
  • vertical deflection problems

Figure 5: Example of distortion effects. Blurring and rippling is applied to the video.

Figure 6: Example of distortion effects. Blurring and rippling is applied to the video.

Multiple mini-cut scenes

♦ Multiple mini-cut scenes
♦ static camera view
♦ multiple independently shot takes
♦ to express distress
♦ controversy towards the current definition of shot boundaries
  • large positional changes of the singer are clearly recognizable as sharp cuts
  • static camera position creates the impression of a coherent scene

Figure 7: Example 2: Cut scene of multiple independent takes with static camera. a) Mean-color bar to illustrate the different shots. Positional changes of the singer through the per-frames of the consecutive shots which are not longer than a few video frames. Singer abruptly changes position and orientation with every single cut.

Blending / Fading / Dissolving

♦ one scene gradually transitions to another
♦ exaggerated use in music videos
♦ artistic use of blending

Figure 1: Example of Blending video frames. Within 0.7 seconds (20 frames at 29 frames per seconds) 7 different scenes are blended.
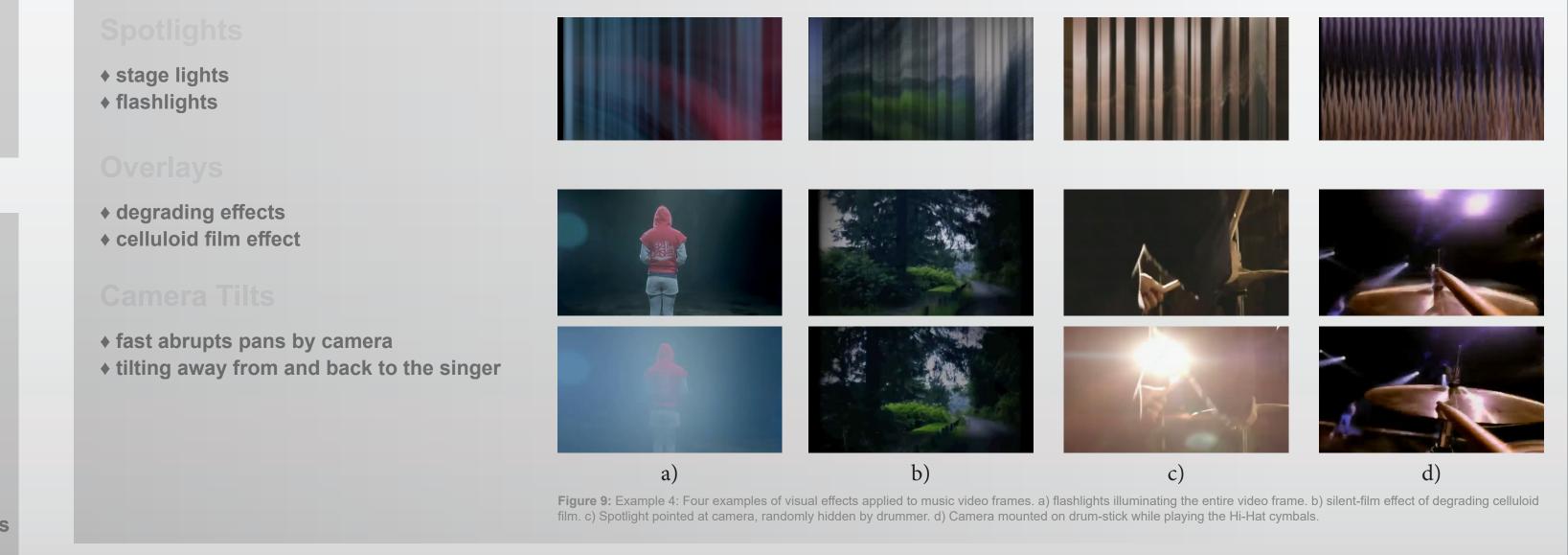
Split-Screen

♦ video frame subdivided into multiple regions
♦ usually rectangular
♦ display different scenes
♦ number of split-segments can change too

Figure 8: Example of Split-Screen scenes. The video frame is split into various, mostly geometric, regions which display different recorded scenes.

Spotlights

♦ stage lights
♦ flashlights

Overlays

♦ degrading effects
♦ celluloid film effect

Camera Tilts

♦ fast abrupts pans by camera
♦ tilting away from and back to the singer

a)   b)   c)   d)

Figure 9: Example 4: Four examples of visual effects applied to music video frames. a) flashlights illuminating the entire video frame. b) silent-film effect of degrading celluloid film. c) Spotlight pointed at camera, randomly hidden by drummer. d) Camera mounted on drum-stick while playing the Hi-Hat cymbals.

One-Shot Music Videos

♦ consist literally of one long take
♦ differnt scenes prepared along a trail
♦ use of shadows and zoom

a)

b)

Figure 10: Example 3: One-shot music videos. a) Different scenes along the route that the camera is following. b) Multiple shots to illustrate the increasing activity.