VQEG 3DTV
DVB 3D format comparison testplan

Marcus Barkowsky, Quan Huynh-Thu, Patrick Le Callet

POLYTECH
NANTES

Overview

- **DVB asks VQEG to evaluate three 3D frame compatible formats:**
  - Side-by-Side
  - Top-Bottom
  - Tile

- **The questions posed by DVB are:**
  1. **The best performing option from amongst the existing solutions**
  2. **Those that offer significant improvements over existing solutions already adopted by the DVB**

  3. The size of the performance advantage, if any, between (1) and (2)
  4. The performance improvement each offers

POLYTECH NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan

25/2/2015

Current test setup

- ■ 10 Video Contents selected from 21 preselected contents

- ■ 3 different bitrates manually chosen per content from the list of 1, 2, 3, 4, 6, 8, and 10 Mbit/s

- ■ Part 1 (in order to answer questions 1 and 2):
  - ❑ Direct Paired Comparison between the 3 different formats, i.e.
    - ■ Side-by-side vs. Top-Bottom
    - ■ Side-by-side vs. Tile
    - ■ Top-Bottom vs. Tile
      ➔ 3 comparisons * 3 bitrates * 10 SRC = 90 unique pairs

POLYTECH
NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan

Current test setup

- Part 2 (in order to check bitrate selection and in order to answer DVB questions 3 and 4):
  - Paired Comparison between all bitrates and all formats, that is:
    - 3 formats * (3 bitrates plus reference) = 12 combinations per content ➔ 12*11/2=66 unique pairs per content ➔ 660 unique pairs for all 10 SRC
    - Using the Optimized Square Design method, the number can be reduced to 300 pairs. The pairs of part 1 are included in this set.

Data analysis and expected results

- **Part 1:**

  - Number of votes for preference of format A as compared to format B

  - Cross-lab agreement
    Statistical analysis with Barnard test, basically:
    If x observers preferred format A to format B in lab A, the number for lab B should be contained within the interval [x-m, x+n]

  - DVB Question 1+2 « which format is preferred per content»
    Statistical analysis with Barnard test, basically:
    if at least 33 observers out of 48 voted for format A, it is significantly different from 24/24 (50%)

POLYTECH
NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan

25/2/2015

Data analysis and expected results

- **Part 2:**
    - All analysis of Part 1
    - Goodness of bitrate selection:
    Statistical significant difference between bitrates, basically:
    if at least 19 out of 24 observers voted for bitrate A, it is significantly different from 12/12 (50%)
    - Comparison of the unencoded reference in all three formats
    - Goodness of data congruence:
    Additional pairs are used to establish an uncertainty value
    (e.g. Bradley-Terry goodness of fit analysis), basically:
    If A>B and B>C, then A>C
    - DVB Question 3+4 « how large is the difference? »
    Conversion of votes to scale values (Bradley-Terry model), basically:
    For content X, the difference between format A and B was d1 while the difference between bitrate E and F was d2

POLYTECH NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan

25/2/2015

Division into subjective test sessions

- **Assuming:**
  - 10 seconds video playback « video A »
  - 0.5 seconds gray screen
  - 10 seconds video playback « video B »
  - 0.5 seconds gray screen
  - 10 seconds video playback « video A »
  - 0.5 seconds gray screen
  - 10 seconds video playback « video B »
  - 0.5 seconds gray screen
  - 4.5 seconds voting
  - 0.5 seconds gray screen
  - In total: 47 seconds per pair
  - 48 observations for part 1
  - 24 observations for part 2
  - 3 parallel observers in front of the screen

Marcus Barkowsky: DVB 3D-format comparison testplan

25/2/2015

POLYTECH
NANTES

Division into subjective test sessions

- **Part 1 requires:** 64 sessions of 18 minutes
  (3 observers every 4 sessions of 18 minutes (47 seconds * 90 pairs))

- **Part 2 requires:** 64+192 sessions of 20 minutes
  (including part 1 as common set)

- **Part 1 does not allow to:**
  - Measure the distance between the selected bitrates
  - Measure the distance of the highest bitrate to the reference
  - Provide an indication of the difference between the formats in comparison to bitrate differences (i.e. the subjective experiment may not be able to distinguish between the formats AND it may not be able to detect 50% bitrate reduction)
  - A stable indication of the reliability of the test setup using consistency checking (Bradley-Terry confidence intervals)
  - Analysis of primary questions with BTL

POLYTECH
NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan

- Part 1 example pairs:
- Bitrate 1 Format A vs. Bitrate 1 Format B
- Bitrate 1 Format A vs. Bitrate 1 Format C
- Bitrate 1 Format B vs. Bitrate 1 Format C

- Part 2 example pairs in addition to part 1:
- Bitrate 1 Format A vs. Bitrate 2 Format A
- Bitrate 1 Format A vs. Bitrate 2 Format B

POLYTECH
NANTES

Marcus Barkowsky: DVB 3D-format comparison testplan