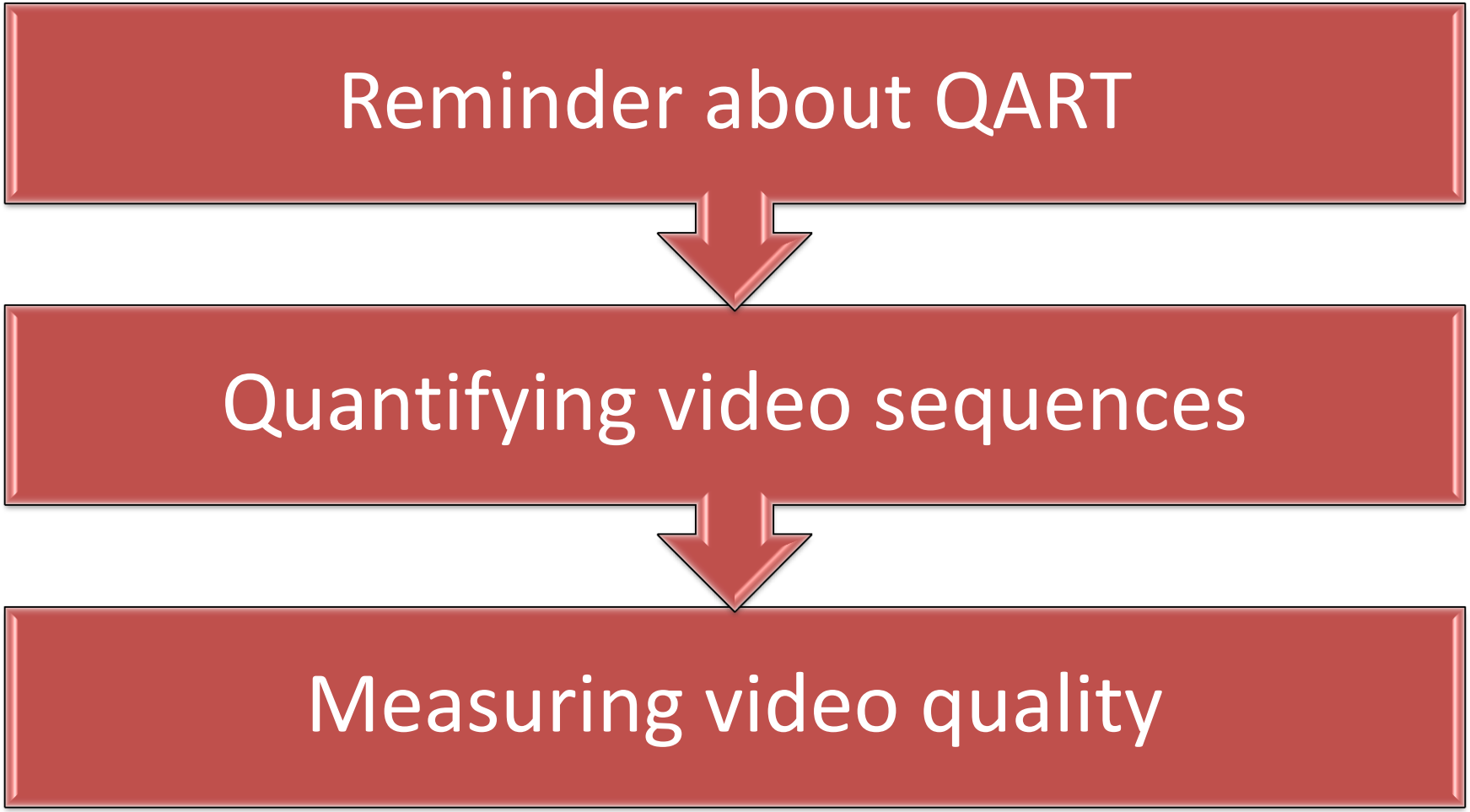
Quality Assessment Recognition Tasks (QART) – Recent Results

Mikołaj Leszczuk, Lucjan Janowski, Łukasz Dudek,
Sergio Garcia
AGH – University of Science and Technology

Joel Dumke
ITS – Institute for Telecommunication Sciences

Presentation Plan



REMINDER ABOUT QART

Quality Assessment for Recognition Tasks (QART)

- **Mission**
  - *"To study effects of resolution, compression and network effects on quality of video used for recognition tasks"*
- **Goals**
  - To perform series of tests to study effects and interactions of
    - Compression
    - Scene characteristics
  - To test existing or develop new objective measurements that will predict results of subjective tests of visual intelligibility

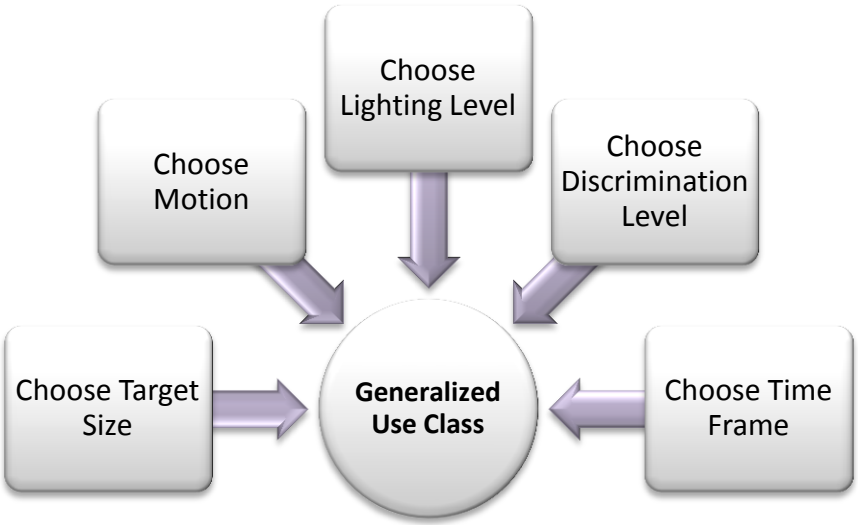
And where do we go from here?

Research topics within area of quality assessment for recognition tasks…

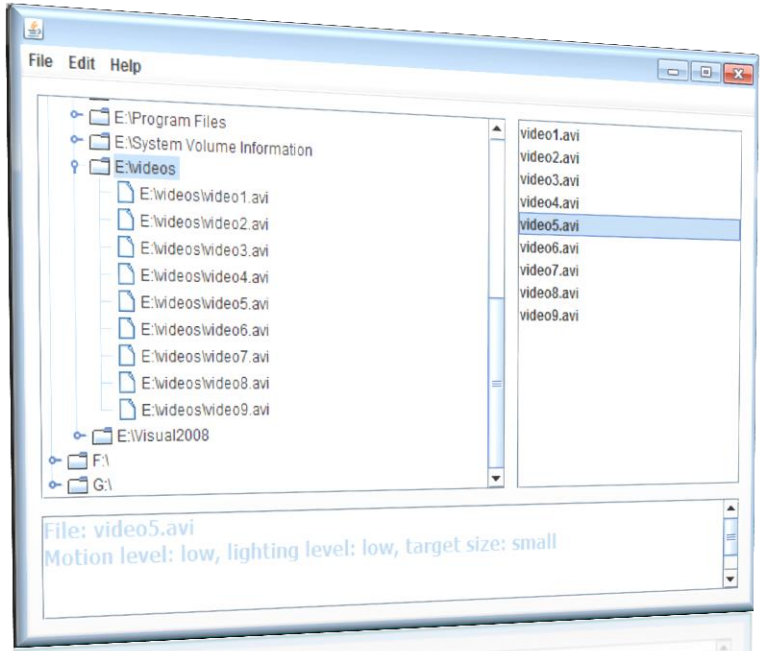
QUANTIFYING VIDEO SEQUENCES

Quantifying Video Sequences (Automatically)

Target size: 70% accuracy,
Lighting level: 93% accuracy
Motion Level: planned



Reference application for practitioners, quantifying GUC (pending) – screenshot of user interface



Target Size

- 2 classes – large/small
- Represents the sizes of appearing objects relative to frame dimensions
- The larger side of objects bounding box is compared to the respective dimension of the frame. The threshold of “large” class is 0.4
- Any object which has been large on the majority of frames it appears in, is classified as large. Otherwise it’s small.
- General results for the scene is the one that applies to more objects

Lighting Level

- Binary classification – high/low
- The threshold of gray level is 55, when range is [0, 255]
- Calculated for every objects as average luminance for every frame, as well as for whole frames
- Objects are bright if they are bright for more frames than not
- Final result – the class which more objects represent

Motion Level

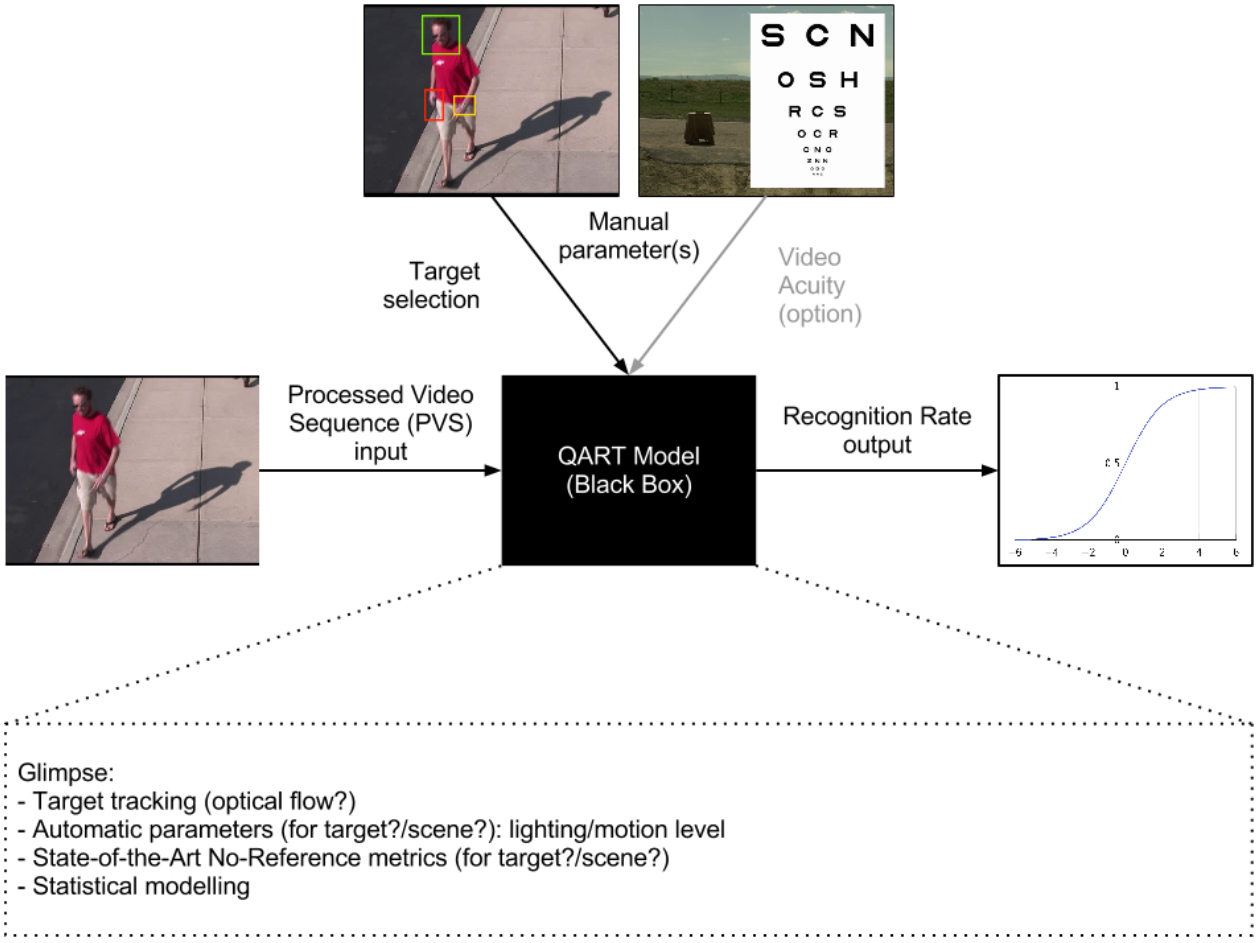
- Also 2 values – high and low
- Not defined properly, therefore not implemented – results presented as undefined
- When necessary, a dummy measure used: average magnitude of gradients in the temporal direction

Example

The only moving object in this clip (person) would be classified as **large** and **bright**. Consequently, this would be the class assigned to the scene.

MEASURING VIDEO QUALITY

QART Model


Target selection
Manual parameter(s)
Video Acuity (option)
Processed Video Sequence (PVS) input
QART Model (Black Box)
Recognition Rate output
Glimpse:
- Target tracking (optical flow?)
- Automatic parameters (for target?/scene?): lighting/motion level
- State-of-the-Art No-Reference metrics (for target?/scene?)
- Statistical modelling

Test sequence database

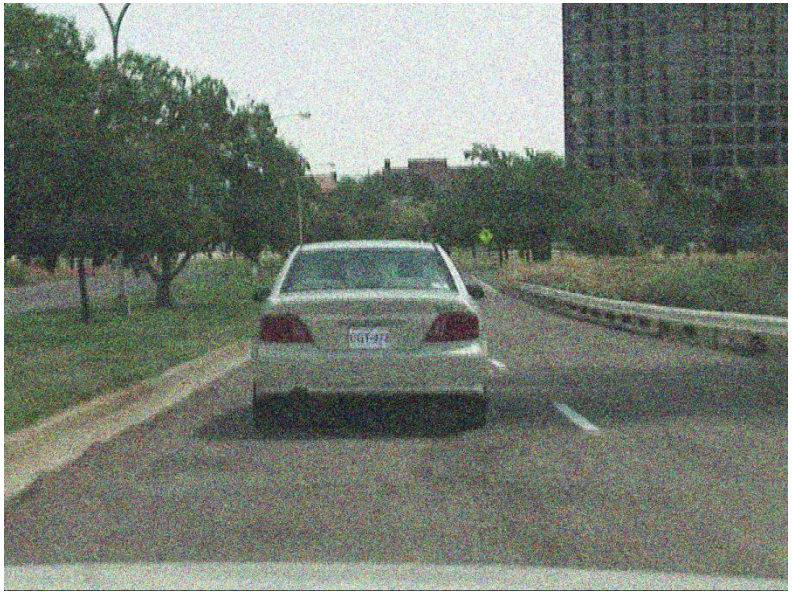
- Goal: create a database of object recognition results for a collection of videos of differing quality

- Store objective quality measures for each video
- Check how well people discern specified visual information from the movies
- **Several** original video sequences – without artifacts
- **Several** derived clips

Many clips from each source

Every original scene is downsampled, cropped or distorted in one of several ways to see how quality loss affects the accuracy of target recognition

Source



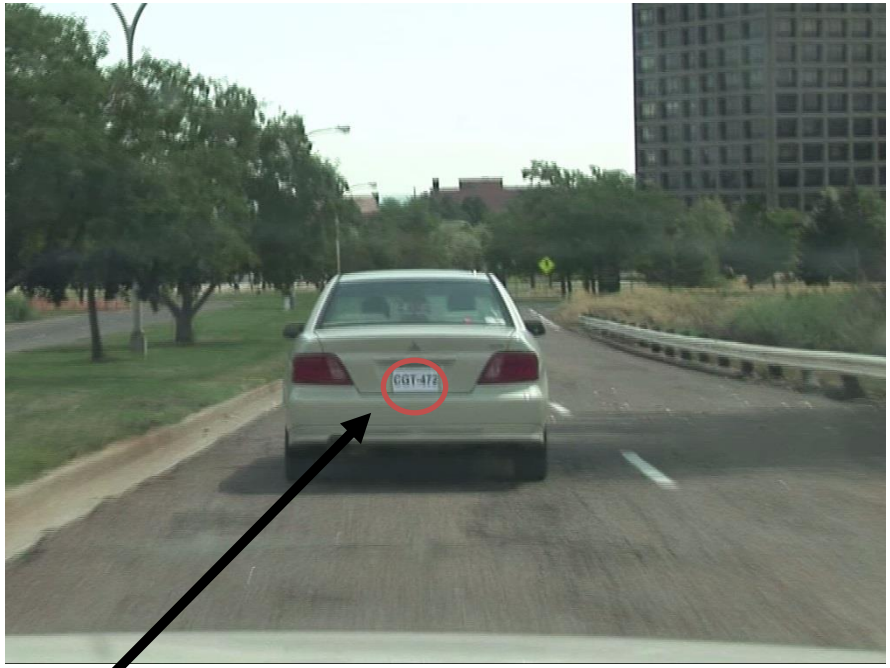
… and so on

What are we testing?

For each clip human subjects are asked about the specified information

Is it a gun? Or phone? Or a radio?

What is the license number?



Target object

How many correct answers? How many wrong?

Target Tracking

- The target object whose recognizability we are examining must be located in every moment throughout the scene
- It has to be marked by human user once in a source sequence
- A selection is stored as rectangle position and the number of frame in which it was indicated
- The bounding rectangles for other frames must then be obtained by object tracker in both time directions from that point

Parameters calculated

- Quality measures such as blur, noise, blockiness, interlacing, etc.
- The list should include enough parameters to enable finding conclusive statistical results
- Each one computed for every frame
- Stored separately for whole frames and for current target location

Object selected

Pixels located in a further frame

Quality parameters calculated for whole frame and for the object being followed

Following the object in other frames

When Work Is Done…

Measures: noise, blur, …

$m_1, m_2, m_3 \ldots m_n$

Probability of recognition:

$p(Correct) = f(m_1, m_2 \ldots m_n)$

Any comments???

THANK YOU – QUESTIONS AND DISCUSSION

LESZCZUK@AGH.EDU.PL