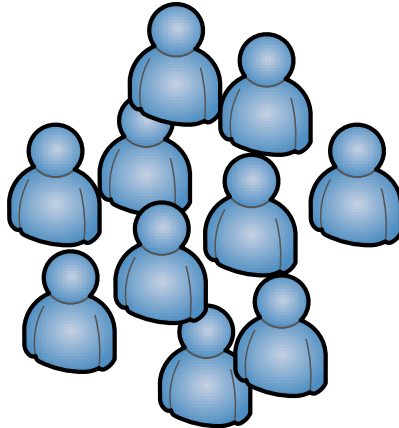# Irrelevant Testers Removal for Recognition Task
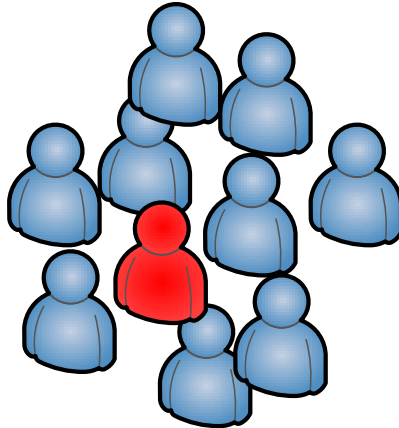
Lucjan Janowski, VQEG, Atlanta 2010

November 16, 2010
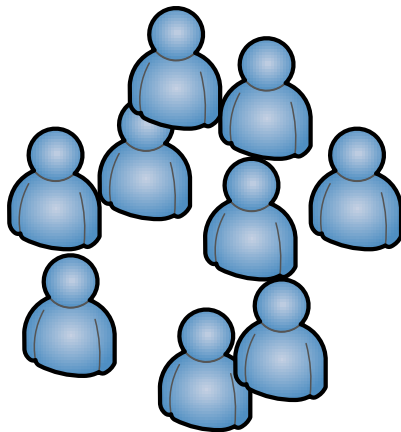
Testers

Testers

Red Paint

Standard Solution

- Each subject judges the same movies

Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range

Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable

Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
- If a subject does not correlate with other subjects he/she is removed

Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
- If a subject does not correlate with other subjects he/she is removed
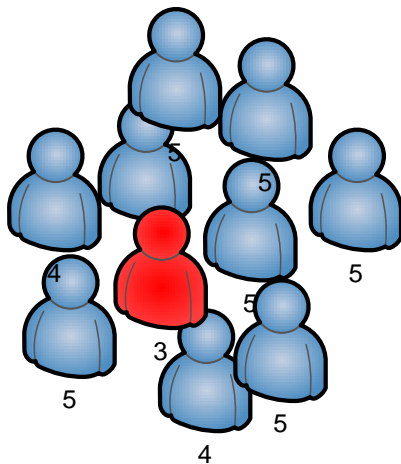- What does it mean "a subject does not correlate?"
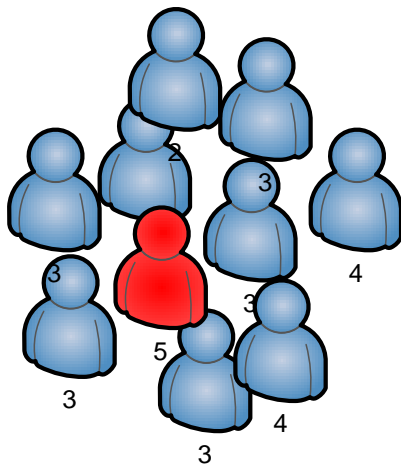
Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
- If a subject does not correlate with other subjects he/she is removed
- What does it mean "a subject does not correlate?"
- VQEG assumes 0.85 - it is a very weak assumption

Standard Solution

Standard Solution

Recognition Task

- Each subject judges the same movies - *very difficult to obtain*

Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
- The score is a value from at least 1-5 range - *more probable is 0-1 answer*

Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
- The score is a value from at least 1-5 range - *more probable is 0-1 answer*
- Pearson correlation starts to be reasonable - *not true*

Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
- The score is a value from at least 1-5 range - *more probable is 0-1 answer*
- Pearson correlation starts to be reasonable - *not true*
- If a subject does not correlate with other subjects he/she is removed - *yes, but correlation has to be defined*

Correlation

We cannot use correlation, why?

Correlation

We cannot use correlation, why?
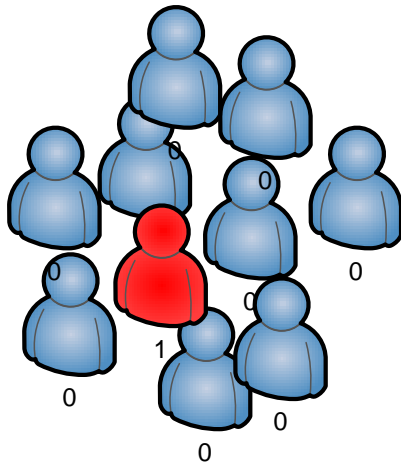
Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
- The score is a value from at least 1-5 range - *more probable is 0-1 answer*
- Pearson correlation starts to be reasonable - *not true*
- If a subject does not correlate with other subjects he/she is removed - *yes, but correlation has to be defined*

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
- The original sequence (not compressed) is not used since the plates reading is too easy in this case

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
- The original sequence (not compressed) is not used since the plates reading is too easy in this case
- The consequence of the above conditions is that the total number of HRC is 30

Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
- The original sequence (not compressed) is not used since the plates reading is too easy in this case
- The consequence of the above conditions is that the total number of HRC is 30
- We have 30 SRCs i.e. each subject sees each HRC and SRC only once!
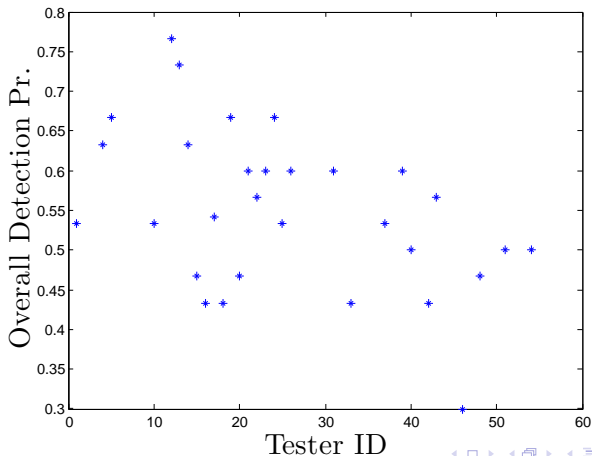
Plate Recognition Interface
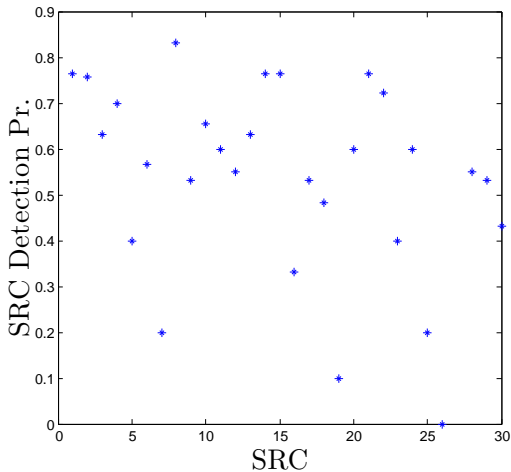
The Simplest Solutions

The simplest subject quality metric is over all detection probability which is

SRC Detection

SRC strongly influences the overall detection probability

Assumption

If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.

Assumption

If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.

The assumption means that a partial order can be defined for the PVSes set.

Assumption

If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.
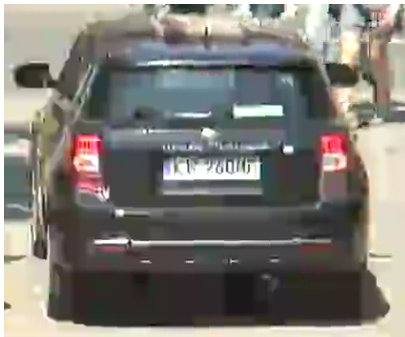
The assumption means that a partial order can be defined for the PVSes set.

It is not obvious so I have investigated this assumption manually.

Assumption Investigation

Assumption Investigation

Subject Quality Metric

$$Sq_i = \sum_{j=1}^{30} ssq_{i,j} \qquad (1)$$

where $ssq_{i,j}$ is subject sequence quality

Subject Quality Metric

$$Sq_i = \sum_{j=1}^{30} ssq_{i,j} \qquad (1)$$

where $ssq_{i,j}$ is subject sequence quality and is given by

$$ssq_{i,j} = \begin{cases} 0 & \text{if } rec_{i,j} = 1 \\ n & \text{if } rec_{i,j} = 0 \end{cases} \qquad (2)$$

Subject Quality Metric

$$Sq_i = \sum_{j=1}^{30} ssq_{i,j} \tag{1}$$

where $ssq_{i,j}$ is subject sequence quality and is given by

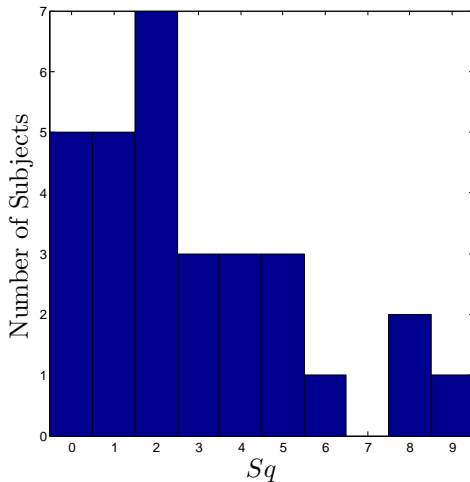$$ssq_{i,j} = \begin{cases} 0 & \text{if } rec_{i,j} = 1 \\ n & \text{if } rec_{i,j} = 0 \end{cases} \tag{2}$$

where

$$n = \sum_{k \in S, l \in A_j} rec_{k,l} \tag{3}$$

$rec_{k,l}$ is 1 if kth subject recognized jth sequence and 0 otherwise, S is all subjects set, A_j is a set of all sequences with the same resolution and view but higher or equal QP than jth sequence.

Sq Metric's Results

The Out layers Errors

ID	Entered number	Original number	Possible error explanation
18	KR102L	KR1002L	typo error
18		KR650LR	unjustified error
18	KR99ES	KR992ES	typo error
18		KR9764S	unjustified error
40	KR97645	KR9764S	similar character
40	KR308	KR3084M	probably typo error
40	KR439HS	KR439HA	typo error
40	RR2492K	KR2492K	typo error
40	KR3527	KR3527L	probably typo error
48	KR97645	KR9764S	similar character
48	KR6966N	KR6986N	probably typo error
48	KR450GF	KR150GF	probably typo error
48	KR249ZK	KR2492K	similar character
48	KR925JG	KR9253G	similar character
48	W67045W(albo	W67045W	additional information

Generalization

A single character error can be justified. The solution is Levenshtein distance.

$$Sql_i = \sum_{j \in A_j} ssql_j \tag{4}$$

where $ssql_{i,j}$ is ith subject quality according to jth sequence and is given by

$$ssql_{i,j} = \begin{cases} 0 & \text{if } leb(i,j) \leq leb(j) \\ leb(i,j) - leb(j) & \text{if } leb(i,j) > leb(j) \end{cases} \tag{5}$$

where $leb(i,j)$ is Levenshtein distance of sequence scored by subject i and having lower or equal QP than sequence j and $leb(j)$ is Levenshtein distance of jth sequence.

Sql Metric's Results
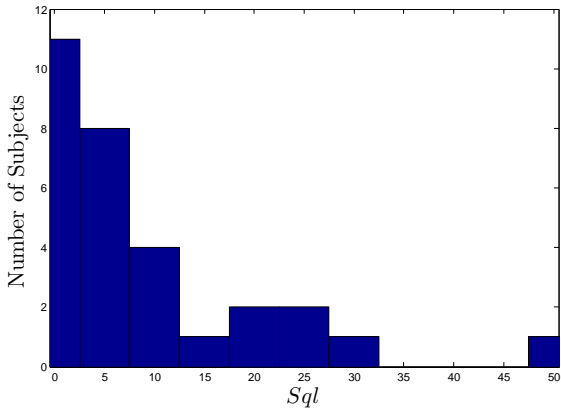
Sq and Sql results

- Sq helps to reveal subjects making typo errors

Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score "not recognized" very easily

Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score "not recognized" very easily
- Subjects make some strange errors probably by almost not seeing the sequence

Sq and *Sql* results

- *Sq* helps to reveal subjects making typo errors
- *Sql* shows that some subjects are willing to score "not recognized" very easily
- Subjects make some strange errors probably by almost not seeing the sequence
- It is difficult to say that a subject should be removed, more likely some of his/her answers should be

Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score "not recognized" very easily
- Subjects make some strange errors probably by almost not seeing the sequence
- It is difficult to say that a subject should be removed, more likely some of his/her answers should be
- It is our goal to build a better interface
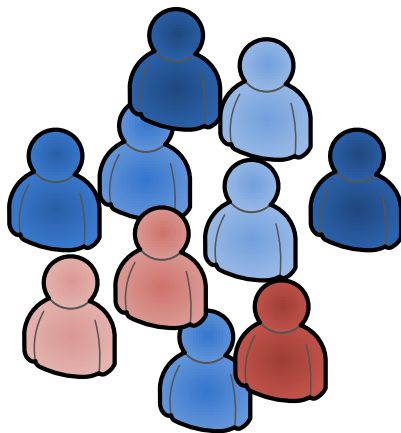
True Subjects Set Quality

Plate Recognition Interface

General Metrics

- Both Sq and Sql metrics can be generalized

General Metrics

- Both Sq and Sql metrics can be generalized
- In Sq case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!

General Metrics

- Both Sq and Sql metrics can be generalized
- In Sq case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!
- In Sql case we need a quality metric also, i.e. we have to be able to measure the error strength

General Metrics

- Both Sq and Sql metrics can be generalized
- In Sq case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!
- In Sql case we need a quality metric also, i.e. we have to be able to measure the error strength
- Manual investigation is still needed but it is limited to the marked sequences and subjects

General Metrics

- Both Sq and Sql metrics can be generalized
- In Sq case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!
- In Sql case we need a quality metric also, i.e. we have to be able to measure the error strength
- Manual investigation is still needed but it is limited to the marked sequences and subjects
- More answers to a single PVS make the metrics more precise. Moreover, we can remove assumption and use only the answers for the same sequence

Any questions/suggestions ?
Lucjan Janowski
janowski@kt.agh.edu.pl