

Document perceptual quality ground truth creation

Vincent Rabeux¹, Nicholas Journet¹, Anne Vialard¹, Jean Philippe Domenger¹

¹University of Bordeaux, LaBRI, Bordeaux, France.



Context

Quality ground truthing of documents :

- quality evaluation of digitized documents.
- used for learning and training steps of algorithms.
- performance evaluation of algorithms.

General way to create a human quality perception ground truth [3] :

- each page is rated by judges.
- each judge needs a global knowledge of the data set to be ground truthed.
- the over all data set needs to be ground truthed in a unique time frame.
- the ground truth is freezed in time.

Objectives

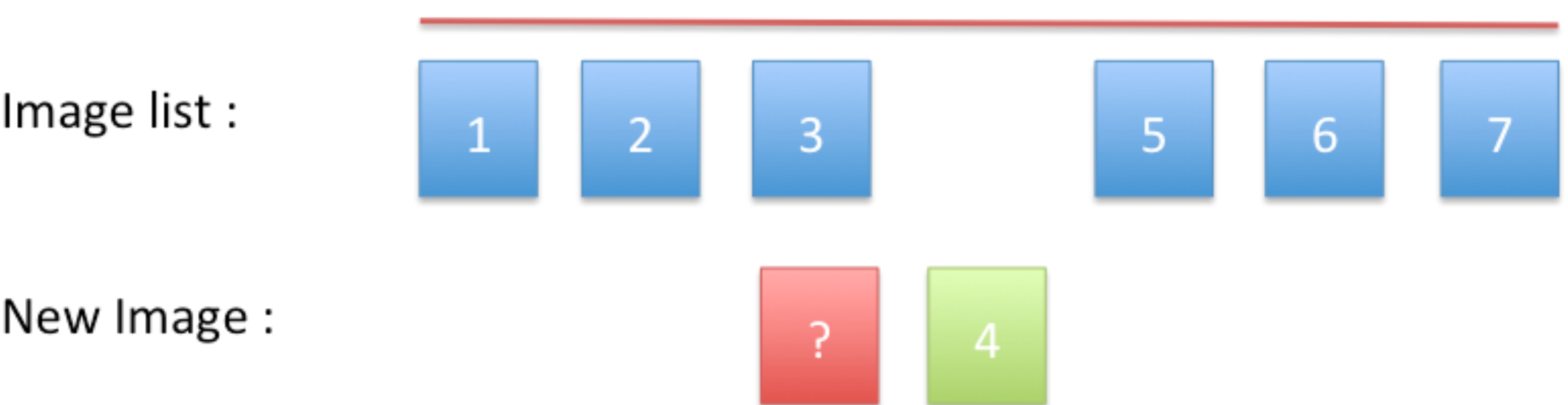
- No reference images.
- Fast, yet accurate method.
- Maintainable ground truth.
- Collaborative.

Over-all algorithm

- Relative ground truth.
- Each image is sorted in a list with a quality criteria.
- Ground truthing a new page consists in inserting it in the sorted list.



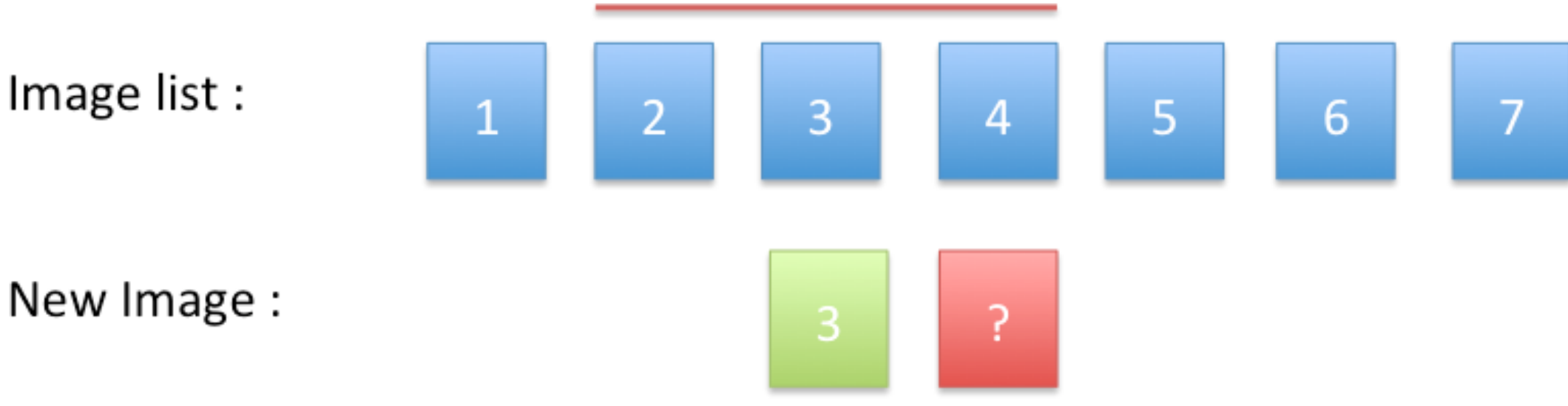
- Binary search algorithm.
- The user is asked which image has the lowest quality.



Which Image has the lowest quality ?



Which Image has the lowest quality ?



Which Image has the lowest quality ?

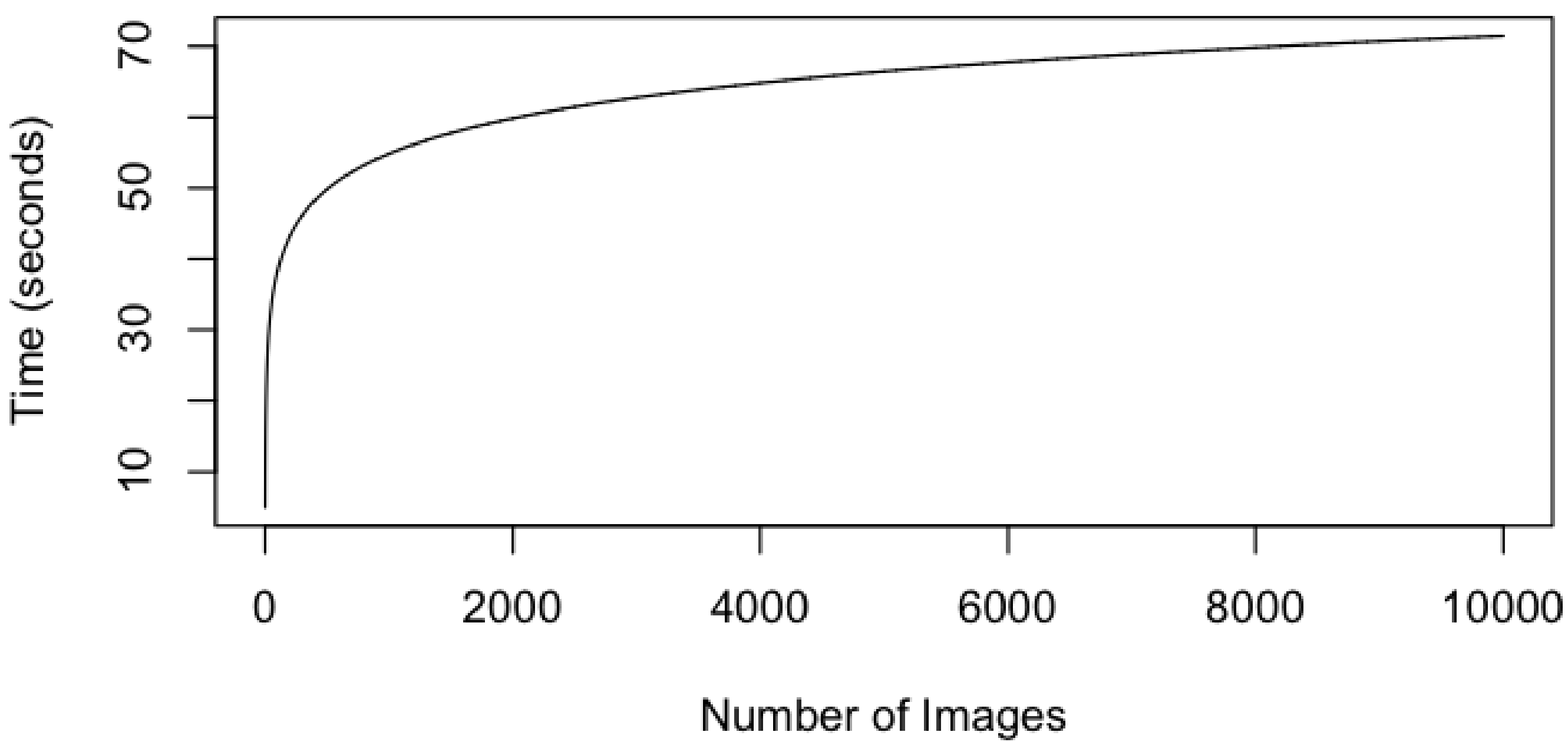
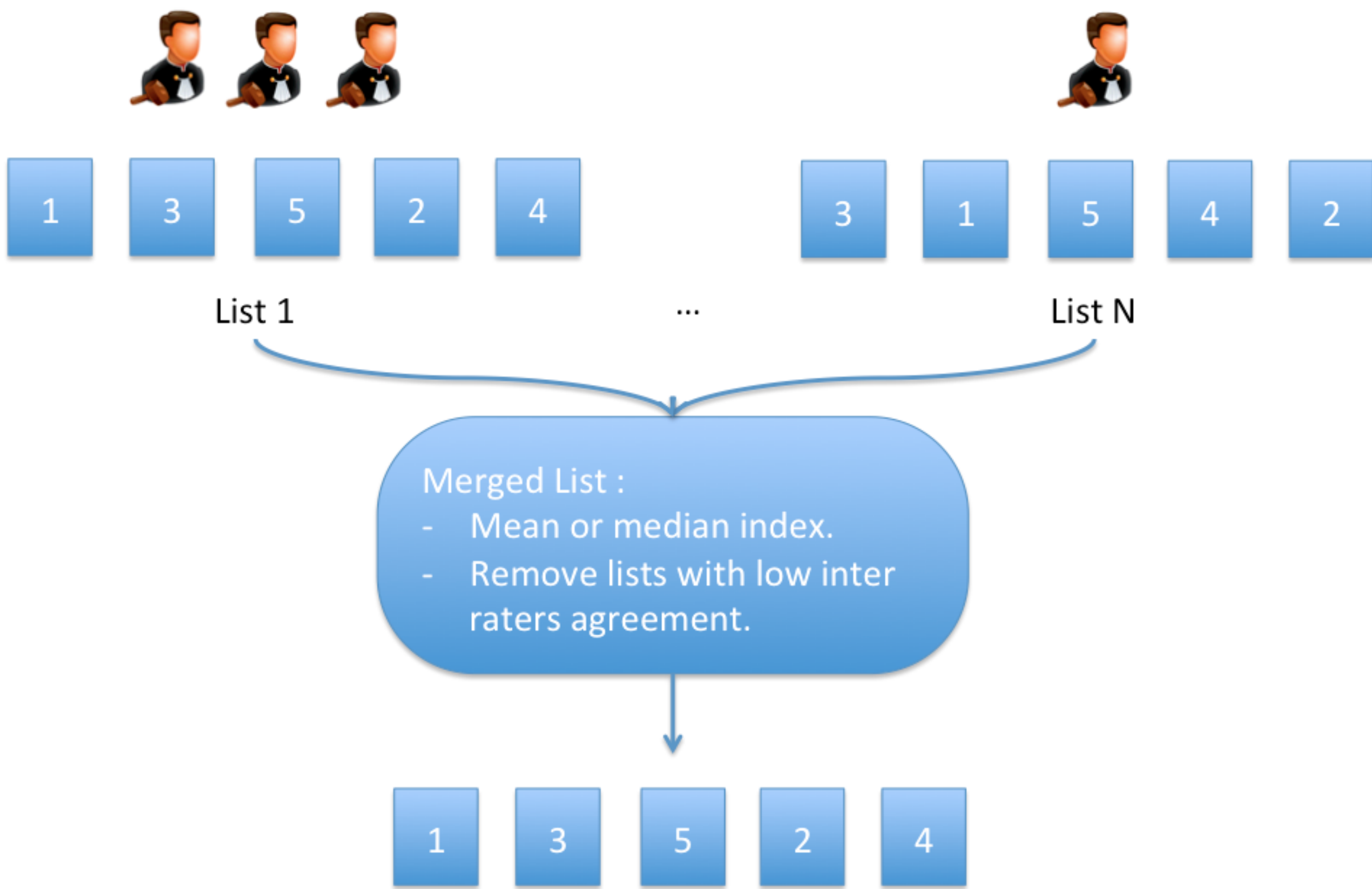


Figure: Time for a user to insert a new image in a list. Each comparison is estimated to take 5 seconds. Even with a list containing 10000 images, the page is ground truthed in about 60 seconds.

Collaborative environnement

- Allows to speed up the all ground truth creation.
- Minimizes the list subjectivity : several users can contribute to the creation of a list.
- Minimizes a page index subjectivity : several list are created.



Experiments and Results

- Implemented with web services (merged list can be used with tools such as Taverna [2]).
- Unified raters environment (iPad).
- How to know if the created ground truth is good enough ?
- Data set 1 : 24 images compressed with JPEG2000 (8 quality levels from 0 to 100).
- Data set 2 : 100 images with 4 different bleed through levels.



Measure the agreement between the *real* ground truth and the one created by 4 different users (Kappa test [1]).

List	1	2	3	4	Merged
Kappa (JPEG2000 dataset)	0.78	0.82	0.86	0.68	0.88
Kappa (bleed-through dataset)	0.88	0.93	0.85	0.75	0.93

Table: Every user agrees with the real ground truth (> 0.60). Kappa values that are higher that 0.80 can be considered as excellent agreements [1].

Conclusion

- Fast and easy ground truthing of document images.
- Maintainable ground truth.
- Collaborative.
- Really represents the human quality perception even if it is subjective.

Perspectives

- Use algorithms to minimize the number of comparisons.
- Detect and minimize the outliers within the merge step.

References

- J. Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.
- D. Hull, K. Wolstencroft, R. Stevens, C. Goble, M.R. Pocock, P. Li, and T. Oinn. Taverna: a tool for building and running workflows of services. *Nucleic acids research*, 34(suppl 2):W729–W732, 2006.
- Anush Moorthy and Alan Bovik. Visual quality assessment algorithms: what does the future hold? *Multimedia Tools and Applications*, 51:675–696, 2011.