Impact of Tone-mapping Algorithms on Subjective and Objective Face Recognition in HDR Images

Pavel Korshunov[*]
MMSPG, EPFL
pavel.korshunov@epfl.ch

Marco V. Bernardo
Optics Center, UBI
mbernardo@ubi.pt

António M. G. Pinheiro
Optics Center, UBI
pinheiro@ubi.pt

Touradj Ebrahimi
MMSPG, EPFL
touradj.ebrahimi@epfl.ch

ABSTRACT

Crowdsourcing is a popular tool for conducting subjective evaluations in uncontrolled environments and at low cost. In this paper, a crowdsourcing study is conducted to investigate the impact of High Dynamic Range (HDR) imaging on subjective face recognition accuracy. For that purpose, a dataset of HDR images of people depicted in high-contrast lighting conditions was created and their faces were manually cropped to construct a probe set of faces. Crowdsourcing-based face recognition was conducted for five differently tone-mapped versions of HDR faces and were compared to face recognition in a typical Low Dynamic Range alternative. A similar experiment was also conducted using three automatic face recognition algorithms. The comparative analysis results of face recognition by human subjects through crowdsourcing and machine vision face recognition show that HDR imaging affects the recognition results of human and computer vision approaches differently.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*perceptual reasoning, representations, data structures, and transforms*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*evaluation/methodology, video*

Keywords

Crowdsourcing, face recognition, HDR, database, evaluation.

1. INTRODUCTION

High Dynamic Range (HDR) imaging is able to capture a wide range of luminance values, similar to the perception of human visual system (HVS). Such ability of HDR to capture details in high-contrast environments, making both dark and bright regions clearly

[*]Currently with Idiap research institute (Martigny, Switzerland)

visible, can have a strong implication on identification and recognition tasks. Since face recognition is typically performed either by human observers or by computer vision algorithms, the effect of HDR imaging on the accuracy of recognition need to be investigated in both of these scenarios.

To be able to conduct such study, and since no publicly accessible HDR face dataset is available, we created a dataset of HDR images showing people of various gender, race, and age in indoor and outdoor environments, under highly variable lighting conditions, including deep shades, sunny outdoor, and dark indoor scenes. From these images, 149 faces of 61 different individuals were manually cropped to construct the testing or *probe* set in HDR, tone-mapped and low dynamic range (LDR) versions. Five different tone-mapping operators, including a simple Gamma-based operator *gamma*, *drago03* by Drago *et al.* [4], *reinhard02* by Reinhard *et al.* [21], *mai11* by Mai *et al.* [14], and *mantiuk06* by Mantiuk *et al.* [15], were used to adapt HDR images for typical LDR displays. A separate training set or *gallery* set is also included in the dataset. The dataset is made freely available to public for research purposes [1].

To evaluate the change in human recognition accuracy when they see a tone-mapped HDR version of an image instead of LDR, we conducted an extensive crowdsourcing study, since this approach was shown as a viable alternative to lab-based subjective assessments [10, 11]. For this purpose, an open source framework QualityCrowd2 [9][2] was adapted and crowdsourcing workers were employed from Microworkers[3] platform. The crowdsourcing workers were asked to find the best match in a known set of 9 faces for a given evaluated face. For each face in the dataset, its five tone-mapped versions and an LDR version were evaluated. In total, 860 workers took part in the crowdsourcing campaign with 42 reliable workers per each evaluated face.

To evaluate the influence of HDR imaging on face recognition algorithms, we used the following popular algorithms available in OpenCV[4] library: based on Principal Component Analysis (PCA) [23], referred to as 'Eigen', based on Linear Discriminant Analysis (LDA) [3], referred to as 'Fisher', and based on local features (LBP) [1], referred to as 'LBPH'. A recognition scenario similar to that used in the crowdsourcing evaluation was simulated for face recognition algorithms. For each tested face, algorithms were trained on 9 random images from the gallery set and the best match was used as the measure of their accuracy. The recognition was

---

[1]It can be downloaded here: http://mmspg.epfl.ch/hdr-faces
[2]https://github.com/ldvpublic/QualityCrowd2
[3]http://microworkers.com/
[4]http://opencv.org/

performed over 20 trials to insure the consistency of the results. Such close similarity between crowdsourcing and objective (i.e., using face recognition algorithms) scenarios allow us to perform a fair analysis and comparison of results.

In summary, this paper has the following main contributions:

- A comprehensive dataset of HDR and LDR images, which represent people in different scenarios and under different lighting conditions. Dataset also contains two subsets of manually cropped faces: a probe set with faces in six different versions, i.e., 5 tone-mapped versions plus an LDR, and a gallery set of typical LDR versions under good lighting condition;

- A subjective evaluation via crowdsourcing and an objective evaluation with three recognition algorithms of the impact of HDR imaging on the recognition accuracy.

2. RELATED WORK

2.1 HDR imaging

Most of the recent studies related to High Dynamic Range (HDR) focus on making the practical transition from the legacy (8 bits-based) systems to HDR-based systems easier. To understand the best way to render HDR content on legacy displays, many different subjective evaluations have been performed that compare different tone-mapping operators for HDR images and video. Main focus of these studies are either on determining better approaches to tone-mapping or establishing an evaluation methodology for subjective evaluation of HDR content. One of the first subjective evaluations of HDR images was performed by Ledda *et al.* [13]. The authors used paired comparison to evaluate the perceptual quality of six different tone-mapping algorithms. An HDR display was used as reference display for 48 subjects. The focus of this work was on the evaluation methodology for the subjective comparison of HDR images in a controlled environment. The evaluations provided the performance ranking of different tone-mapping algorithms leading to different perceptual qualities in color and gray images. Similar studies were conducted to determine the appeal of HDRi [24], usefulness of HDR for astronomical images [18], accuracy of tone-mapping algorithms to represent reality [12], objective metrics of HDR [2], and on using HDR for 3D content [14].

Also, several studies focus on acquisition, compression, and storage aspects of HDR imaging. The book by Myszkowski *et al.* [17] discusses most of the issues related to HDR video and provides overview of the research advances in this area.

However, as opposed to the above works, the focus of our study is on the effect that HDR imaging, once adopted, could have on such cognitive tasks like recognition. The closest related studies are [11] and [19], which investigate the impact of HDR on a more general aspect of privacy intrusiveness. However, we focus specifically on the face recognition task, which is a typical example of visual cognitive tasks, aiming to investigate whether HDR imaging improves the recognition accuracy of both human and computer vision.

2.2 Crowdsourcing

Crowdsourcing is an increasingly popular approach for employing large numbers of people to perform short and simple online tasks. Several commercial crowdsourcing platforms provide online workers with varying cultural and social backgrounds from around the world. Since typical payment for a crowdsourcing job is small and, often, is less than a dollar, crowdsourcing can be a powerful and cost effective tool for performing work that can be easily divided into a set of short and simple tasks, such as surveys, image tagging, text recognition, and viral campaigns.

Subjective quality assessment or QoE assessment of multimedia is another task suitable for crowdsourcing. A typical subjective test consists of a set of repetitive tasks and, hence, can be easily implemented using the crowdsourcing principle. In particular, the cost effectiveness and access to a large pool of test subjects makes crowdsourcing an attractive alternative to lab-based evaluations. Therefore, researchers in quality assessment increasingly use crowdsourcing in various research areas, including rebuffering in streaming video [7], aesthetics of images [20], emotional reaction caused by image content [8], and privacy issues in HDR images [11].

In contrast to traditional recruiting processes, where dedicated employees are selected and assigned to tasks by an employer, in crowdsourcing, the *employer* submits the task as an open call to a large anonymous crowd of *workers*. The workers can then freely decide which available task they want to work on. Usually these tasks have a smaller granularity than traditional forms of work organization and are highly repetitive, such as labeling large number of images. The tasks are usually grouped in larger units, referred to as *campaigns*. Maintaining a dedicated worker crowd, including the required infrastructure, is usually not feasible for most employers and therefore specialized crowdsourcing platforms are used to access the crowd. These platforms abstract the crowd to a certain extent, but sometimes also provide additional services, e.g., quality control or worker selection mechanism.

Crowdsourcing offers the possibility to conduct web-based tests with participants from all over the world. Such flexibility enables a faster completion compared to traditional forms of assessment as more potential participants are available. It can help to reduce the costs of the experiments, since no dedicated test lab is required. It also helps to create a realistic test environment, as the assessment is done directly on the participants' devices. The diversity of the test participants helps to avoid possible biases caused by the limited number of participants in traditional lab tests.

3. DATASET

A dataset of HDR images was created by fusing 5 bracketed images with different exposures ($-2, -1, 0, 1, and\ 2$ exposure settings of the camera) shot with Nikon D7100 and Canon 550D Rebel T2i cameras. In total, 63 images of about 5200×3500 pixels in size were collected, which depict groups of people under highly variable lighting conditions (see Figure 1 for the sample images), including deep shades, sunny outdoor and dark indoor. Although parts of the content were taken from the available PEViD-HDR image dataset [5], the majority of images were shot at premises of UBI campus (Covilhã, Portugal).

In all images, faces of people were manually annotated and by using the annotated coordinates, 149 faces of 61 different individuals were generated to construct the testing or *probe* set. HDR images cannot be displayed on conventional monitors, hence, they cannot be used in crowdsourcing evaluations. Also, HDR images cannot be used directly by face recognition algorithms, as they are often specifically designed to work with 8-bits images. Therefore, we have to apply tone-mapping operators (TMOs), which convert an HDR image to Low Dynamic Range (LDR) 8-bits image in the best possible way. To understand if the choice of tone-mapping has an effect on recognition, we have selected the following five representative tone-mapping operators:

| (a) *LDR with no tone-mapping* | (b) *drago03 TMO* | (c) *mantiuk06 TMO* |
| (d) *LDR with no tone-mapping* | (e) *mai11 TMO* | (f) *reinhard02 TMO* |

Figure 1: Tone-mapped and original LDR images from the dataset.

(**gamma**) a gamma clipping operator scales image values so that the average luminance value is equal to 1.0, then clamps all intensities to $[0, 1]$, and finally applies a gamma correction with an exponent of 2.2. This is a very simple global tone mapping operator that implements a naïve auto-exposure strategy.

(**drago03**) a global logarithmic tone-mapping operator [4], which was found to give good compression performance [16].

(**reinhard02**) a global version of the photographic operator [21], which is a popular choice in many applications.

(**mai11**) a tone-mapping optimized for the best encoding performance in a backward-compatible scheme [14].

(**mantiuk06**) a local operator with strong contrast enhancement [15].

Using these TMOs on 149 images of faces in HDR format led to 5 sets of 149 differently tone-mapped faces (see Figure 2 for examples). Also, since the focus of this study is on understanding what effect HDR imaging has on face recognition task by computers and people, we need *normal* kind of faces for the comparison with the faces obtained from HDR images. For that purpose, from each set of bracketed images that was used to fuse an HDR image, we took one that corresponded to '0 exposure' setting of the camera, since this image would be the *intended* image of the photographer. The value of '0 exposure' was chosen by taking a single LDR image with automatic settings of the camera, so it can be assumed as a default exposure given the surrounding lighting conditions. Given the face recognition scenario, the resulted tone-mapped and LDR images of faces are assumed as belonging to *probe* set of images, which are the faces that are not known and need to be recognized. The set of known images is called *gallery*.

For the gallery set, another set of images were captured with an 'iPhone 5' camera for most participants taking place in the dataset recording. For the rest of the participants, we used their photos from social networks. Faces in these images were also manually

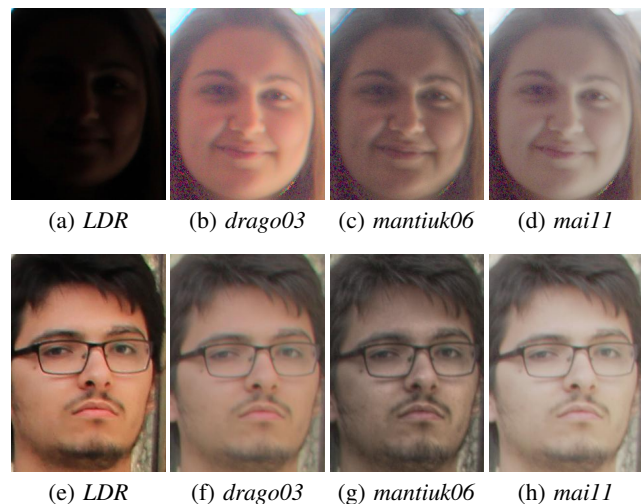| (a) *LDR* | (b) *drago03* | (c) *mantiuk06* | (d) *mai11* |
| (e) *LDR* | (f) *drago03* | (g) *mantiuk06* | (h) *mai11* |

Figure 2: Faces used in the experiments. Notice that LDR image is not always unrecognizable.

annotated and cropped to compose the training or *gallery* set of faces for our evaluations. Also, some faces from LFW dataset[6] were taken to represent the 'known' people that do not appear in the probe set (simulating a practical usage scenario), leading to a total of 105 faces in the gallery.

4. EVALUATION METHODOLOGY

In this paper, we conduct and compare results of face recognition by humans and computers. The human recognition is conducted via crowdsourcing and machine vision using three face recognition algorithms.
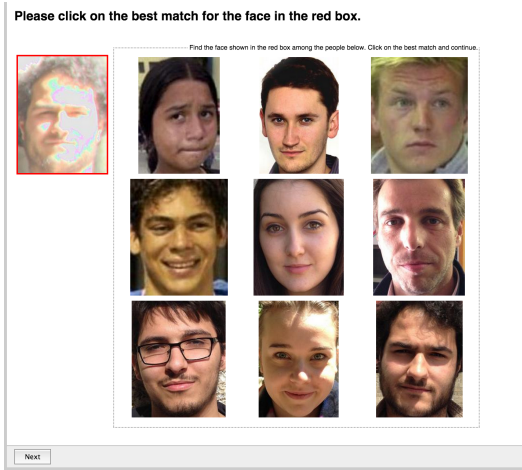
---

[6]http://vis-www.cs.umass.edu/lfw/

Figure 3: An example of the crowdsourcing interface page.

4.1 Crowdsourcing evaluation

We used crowdsourcing approach to understand how the face recognition ability of the human subjects is affected when they see faces from tone-mapped HDR images compared to the faces from typical LDR images. We chose the crowdsourcing approach instead of a more conventional subjective lab-based assessment, because we wanted to reach as wide and variable pool of subjects as possible. The lab-based evaluation is often socially and geographically restrictive to students in a university campus. Crowdsourcing allows to employ subjects with different background, race, and age from all over the world, and such variety in crowdsourcing workers is important for a more thorough and complete study of face recognition.

For the crowdsourcing study, we employed online workers from Microworkers[3], which is a commercial crowdsourcing platform. To display images to different workers provided by Microworkers and to collect evaluation results, we used a modified version of the *QualityCrowd* framework [9]. It is an open-source platform designed for QoE evaluation with crowdsourcing. We choose this framework, because it is easy to modify for our evaluation task using the provided simple scripting language for creating campaigns, training sessions, and control questions. Also, a brightness test was performed for each worker using a method similar to that described in [6].

Figure 3 shows a screenshot of the crowdsourcing interface for the face recognition experiment. Each worker is asked to find the closest match for a given face from testing (probe) set among the 9 randomly selected faces (one of which was the correct match) from gallery set. Before the actual test, short written instructions are provided to the workers to explain their tasks. Additionally, three training samples, with a different content, are displayed to familiarize workers with the assessment procedure. The training instructions and samples are presented using QualityCrowd.

It is not possible for one worker to evaluate all faces in 5 tone-mapped sets plus one set of LDR faces, since the total number amounts to 894 faces. Therefore, the evaluation was split into 18 batches with 50 faces evaluated in each batch. Each worker was allowed to take only one batch. To reduce contextual effects, the stimuli orders of display were randomized and special care was taken for the same face not to be shown consecutively. Subjects were recruited only from the countries, supported by Microworkers, where English is a dominant language, with either more than

Table 1: Human recognition accuracy from crowdsourcing results.

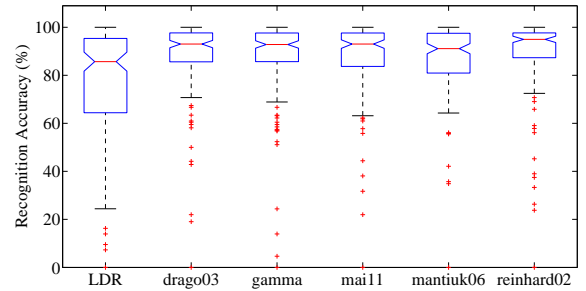| LDR	drago03	gamma	mai11	mantiuk06	reinhart02
76.6%	87.6%	87.3%	87.8%	86.9%	88.2%

Figure 4: ANOVA for the crowdsourcing results.

50% of population or more than 10 million of people speaking English, according to Wikipedia. Based on this criteria and availability of the country in Microwrokers, the workers were employed from the following 13 countries: Germany, Australia, United States, Egypt, France, United Kingdom, Malaysia, Bangladesh, Canada, Poland, Pakistan, India, and Philippines. Each worker was paid 15 US cents for the test that needed about 10 minutes to complete.

Since the major shortcoming of the crowdsourcing-based subjective evaluation is the inability to supervise participants behavior and to control their test conditions, there are several techniques to exclude unreliable workers [6]. To identify a worker as 'trustworthy', the following four approaches were used in our crowdsourcing evaluation:

- Two 'Honeypot' questions were inserted in each batch. These are the obvious easy-to-answer questions (i.e., 'What is the color of the sun?' and 'How much is $6+2$?' to detect people who do not pay attention;

- Task completion time, mean time spent on each question and deviation of the time spent on each question by a worker were measured and analyzed for anomalies.

Based on these factors, 756 out of 860 workers were found to be reliable with 42 reliable workers evaluating each stimuli (a face), which ensures the statistical significance of the evaluation results.

4.2 Face recognition evaluation

We investigated the influence of the HDR imaging on performance of three face recognition algorithms implemented in OpenCV: Principal Component Analysis (PCA)-based [23], referred to as 'Eigen', Linear Discriminant Analysis (LDA)-based [3], referred to as 'Fisher', and local features (LBP)-based [1], referred to as 'LBPH'.

The experiments with face recognition algorithms were conducted with the aim to match the crowdsourcing experiment as close as possible. For each algorithm and for each given face from the testing (probe) set, the algorithm was trained on 9 random faces (one of them was the correct match) from the training (gallery) set. The gallery faces were selected randomly for each probe face and 50 different trials were run to insure the fairness of the results. The resulted score for the algorithm was computed by averaging the recognition accuracy (the number of correctly recognized faces divided by the total number of faces) for each probe face across all trials, which is essentially a true positive measure.

Table 2: The accuracy of recognition algorithms for LDR and HDR tone-mapped images.

TMO	Eigen	Fisher	LBPH
LDR	19.66%	19.09%	21.21%
drago03	20.27%	20.67%	27.68%
reinhart02	20.97%	21.17%	28.36%
maniuk06	22.15%	20.94%	27.32%
mai11	22.81%	22.28%	29.30%
gamma	21.38%	22.65%	29.09%

5. RECOGNITION RESULTS

Table 1 shows the face recognition accuracy obtained with crowdsourcing evaluation. The one-way ANOVA results are shown on Figure 4, and led to $p = 0$ and $F = 8.32$. It shows that the face recognition accuracy corresponding to LDR images is statistically different with $p = 0$ from the accuracies corresponding to faces tone-mapped with five tone-mapping operators. However, the results obtained for different tone-mapping operators are not statistically different ($p \simeq 1$).

Table 1 and Figure 4 clearly demonstrate that using HDR imaging significantly increases recognition accuracy of human subjects, since the recognition accuracy for tone-mapped faces is at least 11% higher (i.e., for 'drago03') when compared to typical LDR faces. At the same time, different tone-mapping operators lead to similar recognition accuracy, which means that even the simplest *gamma* operator can be used in a practical scenario.

Table 2 shows the recognition accuracies of the three recognition algorithms for tone-mapped images when compared to *normal* LDR images. The table demonstrates that different tone-mapping operators affect three recognition algorithms differently. For instance, using *mai11* leads to the best the performance for all recognition algorithms, while *drago03* is more dependent on the choice of the recognition approach. Also, using HDR tone-mapping instead of LDR images improves the performance of LBPH recognition algorithm the most. It is important to note that in this study we are not concerned with the total recognition accuracy but are interested in the relative change in the accuracy when LDR faces are replaced by tone-mapped HDR faces. So, the overall low recognition scores, which are due to the use of somewhat simplistic recognition techniques, have little impact on the conclusions drawn.

To verify how the human recognition can be correlated with the used face recognition methods, the Pearson (measure of the model prediction accuracy) [5] and the Spearman (measure of the model prediction monotonicity) [22] correlations were computed between the raw voting data obtained in crowdsourcing evaluation and the results of face recognition algorithms. A non-linear regression was fitted to subjective and objective data restricted to be monotonic over its range, using the following regression equation:

$$\text{Acc}_p = b1 + \frac{b2}{1 + e^{(-b3 \times (RA - b4))}},$$

where Acc_p is the estimated recognition accuracy, RA is the recognition data of the method for which the estimation is computed, and $b1$, $b2$, $b3$ and $b4$ are the regression parameters, initialized with 0, 1, 0 and 1 respectively.

Figure 5 and Figure 7 show Pearson and Spearman correlations respectively between crowdsourcing data and face recognition algorithms. These figures show low correlation between human recognition and face recognition algorithms.

Pearson correlation (see Figure 5) show higher prediction accuracy in the case of LDR images compared to different TMOs. It means that the tested face recognition algorithms are not appropri-
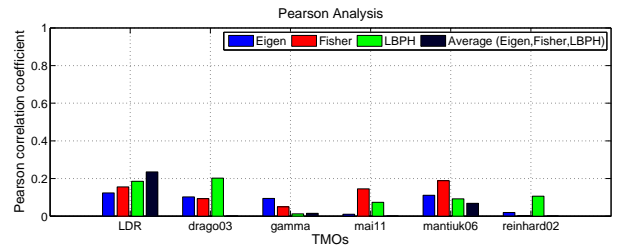

Figure 5: Pearson correlation between crowdsourcing recognition and face recognition algorithms for different TMOs.
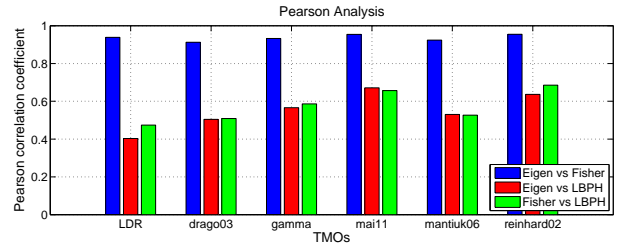

Figure 6: Pearson correlation between face recognition algorithms for different TMOs.

ate for images generated by TMOs. For several TMOs, the correlation bar in the figure is not perceptible, because the value is very close to zero, which means that the results of crowdsourcing and face recognition algorithms do not correlate at all. Spearman correlation (see Figure 7) also reveals very low prediction monotonicity in most of the cases. In contrast to Pearson case, Spearman correlation for different TMOs is often similar to the LDR case. However, the prediction monotonicity of the average value of face recognition algorithms' accuracy for different TMOs is not relevant, because there is no prediction accuracy.

Figure 6 and Figure 8 present Pearson and Spearman correlations between different face recognition algorithms. These figures demonstrate higher correlation between Eigen- and Fisher-based recognition algorithms and lesser correlation between them and LBPH. These results are expected, since Eigen- and Fisher-based recognition algorithms are more similar. Also, the figures show that different TMOs influence correlations differently.

6. CONCLUSION

In this paper, a dataset of HDR images with people in them was created, faces in these images were cropped, and five tone-mapping operators were applied to them to create five sets of faces to evaluate the change in the recognition ability of human subjects and computer vision compared with typical LDR faces. Crowdsourcing framework was setup to evaluate the recognition of human subjects and three face recognition algorithms were used to evaluate the recognition by a computer.

The results of the crowdsourcing-based and algorithm-based face recognition evaluation show interesting phenomena. Different tone-mapping operators affect three recognition algorithms differently but are very similar in the way they affect recognition by human observers. Also, face recognition results by humans show almost no correlation with recognition accuracy by machines.

In the future study, more advanced face recognition algorithms, i.e., based on deep learning networks, could be used, with additional pre-processing steps, such as contrast normalization and face alignment.
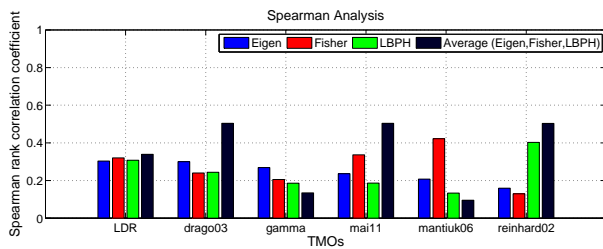
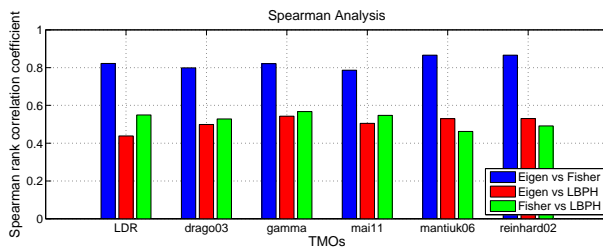Figure 7: Spearman correlation between crowdsourcing recognition and face recognition algorithms for different TMOs.

Figure 8: Spearman correlation between face recognition algorithms for different TMOs.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] B. Annighöfer, T. Tajbakhsh, and R.-R. Grigat. Prediction of results from subjective evaluation of real-time-capable tone-mapping operators applied to limited high-dynamic-range images. *Journal of Electronic Imaging*, 19(1):011015–011015–12, Jan. 2010.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22(3):419–426, Sept. 2003.

[5] J. Gibbons and S. Chakraborti. *Nonparametric Statistical Inference, Fourth Edition*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 2003.

[6] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia. Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. *IEEE Transactions on Multimedia*, PP(99):1–1, 2013.

[7] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz. Quantification of YouTube QoE via crowdsourcing. In *Symposium on Multimedia*, Dana Point, USA, Dec. 2011.

[8] I. Hupont, P. Lebreton, T. Mäki, E. Skodras, and M. Hirth. Is it possible to crowdsource emotions? In *International Conference on Communications and Electronics*, Da Nang, VN, July 2014.

[9] C. Keimel, J. Habigt, C. Horch, and K. Diepold.

[10] P. Korshunov, S. Cai, and T. Ebrahimi. Crowdsourcing approach for evaluation of privacy filters in video surveillance. In *Proceedings of the ACM Multimedia 2012 Workshop on Crowdsourcing for Multimedia*, CrowdMM'12, pages 35–40, Nara, Japan, Oct. 2012.

[11] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi. Crowdsourcing-based evaluation of privacy in HDR images. In *SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications*, Brussels, Belgium, Apr. 2014.

[12] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating HDR rendering algorithms. *ACM Trans. Appl. Percept.*, 4(2):9:1–9:27, July 2007.

[13] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH'05, pages 640–648, New York, NY, USA, Aug. 2005. ACM.

[14] Z. Mai, C. Doutre, P. Nasiopoulos, and R. Ward. Subjective evaluation of tone-mapping methods on 3D images. In *17th International Conference on Digital Signal Processing (DSP)*, pages 1–6, NJ, USA, July 2011.

[15] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel. Backward compatible high dynamic range MPEG video compression. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH'06, pages 713–723, New York, NY, USA, Aug. 2006. ACM.

[16] R. Mantiuk and H.-P. Seidel. Modeling a generic tone-mapping operator. *Computer Graphics Forum*, 27(2):699–708, 2008.

[17] K. Myszkowski, R. Mantiuk, and G. Krawczyk. High dynamic range video. *Synthesis Lectures on Computer Graphics and Animation*, 2(1):1–158, 2008.

[18] S. H. Park and E. D. Montag. Evaluating tone mapping algorithms for rendering non-pictorial (scientific) high-dynamic-range images. *J. Vis. Comun. Image Represent.*, 18(5):415–428, 2007.

[19] M. Pereira, J.-C. Moreno, H. Proença, and A. M. G. Pinheiro. Automatic face recognition in HDR imaging. *Proc. SPIE*, 9138:913804–913804–10, 2014.

[20] J. Redi, T. Hossfeld, P. Korshunov, F. Mazza, I. Povoa, and C. Keimel. Crowdsourcing-based multimedia subjective evaluations: a case study on image recognizability and aesthetic appeal. In *ACM CrowdMM 2013*, Barcelona, Spain, Oct. 2013.

[21] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267, July 2002.

[22] C. Spearman. Correlation calculated from faulty data. *British Journal of Psychology, 1904-1920*, 3(3):271–295, 1910.

[23] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591, 1991.

[24] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. volume 5666, pages 192–203. Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging, 2005.