LATENT FACTOR MODELING OF PERCEIVED QUALITY FOR STEREOSCOPIC 3D VIDEO RECOMMENDATION

Balasubramanyam Appina[a], Mansi Sharma[b], Santosh Kumar[b],
Peter A. Kara[c], Aniko Simon[d], Mary Guindy[e, f]

[a]Department of Electronics and Communication Engineering,
Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, Chennai, India.
Email: appina@iiitdm.ac.in

[b]Department of Electrical Engineering,
Indian Institute of Technology Madras, Chennai, India.
Email: {mansisharma@ee, ee15b110@smail}.iitm.ac.in

[c]Department of Networked Systems and Services,
Budapest University of Technology and Economics, Budapest, Hungary.
Email: kara@hit.bme.hu

[c]Kingston University, London, UK
Email: p.kara@kingston.ac.uk

[d]Sigma Technology, Budapest, Hungary
Email: aniko.simon@sigmatechnology.se

[e]Holografika, Budapest, Hungary
Email: m.guindy@holografika.com

[f]Pazmany Peter Catholic University, Budapest, Hungary
Email: guindy.mary.mohsen.messak@ppke.hu

ABSTRACT

Throughout the recent decades, the movie industry has been continuously producing stereoscopic 3D contents on a wide scale of varying budget; while some of them are rather limited financially, others benefit from massive monetary reinforcements. It is also quite apparent how the related technologies have significantly improved over the years, resulting in better and better 3D visuals. However, at the time of this paper, it is still the sad truth that even those movies with virtually limitless financial support suffer from stereoscopic artefacts. Although certain scientific efforts have already addressed the objective determination and adequate classification of such issues, sufficiently integrating perceived quality into these models remains to be an open research question. This paper presents our work to reduce the gap between the proposed objective methodology and the human visual system. We introduce a new type of solution that relies on latent factor modeling. The proposed model is trained and tested on the benchmark Nama3ds1-cospad1 and LFOVIAS3DPh2 stereoscopic 3D video quality assessment datasets. The experiments highlight the practical efficiency and considerable performance of the resulting matrix-factorization-based recommendation system.

Index Terms— Stereoscopic 3D video, recommendation system, latent factor model, matrix factorization.

1. INTRODUCTION

At the time of this paper, there is a notable growth in both the demand and the supply of novel 3D video contents, and the stereoscopic 3D (S3D) format is rapidly emerging in both cinematic and streaming use cases. There are three popular approaches to generate S3D videos: (i) Scene acquisition using a stereo camera, (ii) 2D-to-3D video conversion, which means the creation of left- and right-eye views from the original source video, and (iii) rendering, which is the process of synthesizing views by means of 3D reconstruction or employing global 3D models and computer vision techniques [1–5].

Despite the advances in 3D technology, visual artefacts still frequently appear in S3D contents. A comprehensive study of visual artefacts in S3D contents was carried out at MSU Graphics & Media Lab, Moscow State University, under the VQMT3D project [6,7] in cooperation with IITP RAS. The research identified potential artefacts in several popular Hollywood movies, released in S3D format. The artefacts (e.g., disparity, scale, color, and sharpness mismatches, temporal asynchrony, cardboard and crosstalk effects etc.) are rather common within the 3DTV use case of S3D contents. The different types of artefacts affect S3D videos at various stages of the content delivery. The compression, blur and frame-freeze distortions degrade videos in the stages of format conversion, representation, coding and transmission [8]. Zeri and Livi [9] interviewed 854 people. They recognized frequent symptoms of eye strain, blurred vision and a burning sensation after watching 3D films in movie theaters. Even high-budget films, like *Pirates of the Caribbean*, *Dolphin Tale*, *The Three Musketeers*, or even *The Avengers* contain multiple scenes with geometric and color impairments, camera rotation differences, and vertical variation shifts between the left and right views. The study conducted by Barreda *et al.* [10] on 3D contents using psycho-physiological methods determined complex effects of visual discomfort regarding 3DTV, which may cause headache, nausea, fatigue, eye strain, etc. According to several publications, the compression artefacts and their variations (based on the depth range of 3D displays) may noticeably affect the perception of the viewers [11–14]. Furthermore, multiple studies conclude that discomfort is greater for certain types of distortions than for others when viewing S3D contents, and that the influence primarily originates from the content itself. [7, 15].

The most reliable way to reduce such distortions is to correct and enhance the quality of the S3D content during production. However, such correction processes are not cost-efficient, as they require immense efforts and they heavily rely on the achieved degree of automation. The algorithms for the automatic detection of the aforementioned artefacts and for quality assessment are already emerging [16, 17], yet the measuring frequency and the objective intensity of an artefact do not account for the actual level of discomfort suffered by the viewer. Therefore, it is critical to consider subjective perception ratings for artefacts. Such data can provide valuable feedback regarding the tolerance towards instances of visual degradation and may define perceptual thresholds.

In this paper, we propose a novel recommendation system for S3D movies. Based on the scientific literature available at the time of this paper, we conclude that the most significant component for designing a recommendation system for S3D movies is the proper characterization of perceptual discomfort with regards to particular distortion types. Despite the notable advances in objective quality assessment techniques for image and video contents [18–21], it is difficult to propagate the same achievement for S3D videos, as the automatic estimation of the relevant properties of issues regarding visual discomfort is quite far from being trivial. Yet one thing is certain: a well-designed recommendation system for such contents must take subjective ratings into account to serve its purpose in a user-centric and efficient manner.

Our recommendation system is built on the latent factor model that relies on the subjective ratings of the user. Given a set of reference-quality (i.e., completely free of visual distortions; also known as ground truth) and distorted S3D videos and their corresponding subjective ratings, our model utilizes matrix factorization maps on a set of latent features. The challenge of predicting perceptual quality ratings of S3D videos is formulated as a matrix completion problem for the content rating matrix. Our system rates the S3D videos in accordance with the user's level of discomfort. The recommendation mechanism can easily integrate within the matrix factorization methods of Netflix, which is the most important class of collaborative filtering approaches. The proposed system can be rather advantageous for efforts aiming to reduce the online flood of low-quality 3D contents by rating more consistently with quality. The results obtained by statistical analysis – conducted on the Nama3ds1-cospad1 [14] and the LFOVIAS3DPh2 [17] benchmark S3D datasets – demonstrate the prediction accuracy of our model.

The remainder of the paper is structured as follows: Section 2 reviews the related scientific literature. The mathematical model of our novel scientific contribution is provided in Section 3. The results are presented and discussed in Section 4. Section 5 concludes the paper and also points out the potential future continuations of our research efforts.

2. RELATED WORK

The algorithms used by Netflix for its recommendation system are comprehensively analyzed in the work of Fernández [22], which covers collaborative filtering, content-based filtering, model-based singular value decomposition (SVD), principal component analysis (PCA) and probabilistic matrix factorization techniques. The paper explains a movie recommendation mechanism built within Netflix on the matrix factorization (MF) [23] approach that learns the latent preferences of the users via ratings and provides a prediction of the miss-

ing ratings (i.e., it estimates the dot product of the latent factors) [22, 23].

Lu *et al.* [24] applied an MF model for computing the vector representations of words. The work demonstrates how a convolutional neural network (CNN) can be integrated into an MF model to produce interpretable recommendations. Lee *et al.* [25] developed a demo model that considers recently uploaded YouTube videos. Here the collaborative filtering approach is not applicable since it relies on the aggregation of user behavior. Instead, they approached the task of recommendation as content-based similarity learning, utilizing deep video embeddings. However, this approach is purely based on video content signals.

YouTube provides an absolutely immense collection of 2D videos. Since 2009, YouTube also supports two-channel S3D videos. The supported 3D formats include anaglyph videos in red / cyan, green / magenta and blue / yellow layouts, which are row / column interlaced for visualization. The work of Tsingalis *et al.* [26] presents a study on YouTube recommendation graphs of 2D and 3D videos. The authors investigated the statistical relevance and recommendation properties of social network sites such as Facebook, Twitter and Flickr. They also analyzed clustering methods to understand the existence of media content groups. The paper of Davidson *et al.* [27] discusses the recommendation system of YouTube. The study highlights that YouTube recommends personalized sets of videos to users based on their previous activity on the Internet. The publication of Covington *et al.* [28] addresses the substantial performance improvements brought by deep learning. A deep architecture – built on deep candidate generations – and a separate ranking model for recommending YouTube videos are proposed in the work.

Estrada and Simeone [29] developed a recommendation system for guiding physical object substitution in virtual reality (VR). The perception-based approach allows the individuals to watch the physical world whilst navigating the virtual environment through a video feed. Technically, in this framework, the user identifies the location of object placement in the surroundings.

The work of Niu *et al.* [30] presents a video recommendation system based on the affective analysis of the users. The proposed subjective model evaluates the feature of emotion fluctuation based on the Grey Relational Analysis (GRA). Certain video features are extracted and are mapped to the well-known Lovheim emotion-space, specifying prominent human feelings, patterns, attitudes and behaviour. These include anger, distress, surprise, fear, enjoyment, shame, interest and contempt. The GRA-based recommendation method is developed under the Fisher model to analyze the extracted emotions as factors.

The paper of Zhang *et al.* [31] introduces a recommendation system for a mobile augmented reality (AR) application, incorporating the user's preferences, location and temporal information in an aggregated random walk algorithm.

Their system predicts the user's preferences, modifying the graph edge weight and computing the rank score. Similarly, the solution of Shi *et al.* [32] predicts individual location recommendation, while the contribution of Chatzopoulos and Hui [33] anticipates object recommendation in mobile AR environments.

3. MATHEMATICAL MODELING OF THE S3D VIDEO RECOMMENDATION SYSTEM

In this section, the novel recommendation system for S3D videos is introduced. First of all, it is based on an MF model. The subjective perception of Viewers and S3D contents are mapped to a joint latent factor space. The row or column associated to a specific viewer or S3D content is referred as latent factor. In the mapped latent factor space, which has a dimensionality of f, the ratings are analyzed as inner products. Let us suppose that each S3D content i is associated with a latent vector $q_i^m \in R^f$, and each viewer u is associated with a latent vector $p_j^u \in R^f$. For a given S3D content i, the elements of q_i^m estimate the extent to which the content holds those factors, whether distorted with a particular artefact or free from that. For a given viewer u, the elements of p_j^u determine the extent of acceptance of the contents that are high on the corresponding factors, again, whether distorted with a particular artefact or not. The model approximates viewer u's rating of content i by measuring the resulting dot product, $\hat{r}_{ui} = q_i^{m^T} p_j^u$. The dot product captures the interconnection between viewer u and content i, that is, the viewer's overall tolerance regarding the particular distortion. Once the mapping is computed for each content and viewer to factor vectors $q_i^m, p_j^u \in R^f$, the proposed model determines the rating a viewer would give to any distorted S3D content by using $\hat{r}_{ui}$.

We avoided imputation in the model [34]. The observed ratings are modeled directly as suggested by multiple works of the scientific literature [23, 35], and avoided overfitting through regularization. On the set of the known matrix ratings, the regularized squared error is minimized to learn the factor vectors q_i^m, p_j^u as

$$\min_{\hat{p}, \hat{q}} \sum_{(u,i) \in \mathfrak{S}} (r_{ui} - q_i^{m^T} p_j^u)^2 + \lambda(||q_i^m||^2 + ||p_j^u||^2) \quad (1)$$

where, $\mathfrak{S}$ is the training set of (u, i) pairs for which r_{ui} is known.

To make the matrix factorization approach more effective for our application-specific requirements, we add biases in capturing the full ratings of the observed signals,

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^{m^T} p_j^u \quad (2)$$

The observed rating in (2) is broken down into its four components: global average (or mean), content bias, viewer

bias and viewer-content interaction. This allows each component to represent only the part of an observed signal relevant to it. The model is trained by minimizing the squared error function as

$$\min_{\hat{p},\hat{q},\hat{b}} \sum_{(u,i)\in\mathfrak{S}} (r_{ui} - \mu - b_i - b_u - q_i^{m^T} p_j^u)^2 +$$
$$+ \lambda(||q_i^m||^2 + ||p_j^u||^2 + b_i^2 + b_u^2) \quad (3)$$

The stochastic gradient descent algorithm [23, 36, 37] is used to optimize equation (3). For better accuracy in prediction, the algorithm loops through all ratings in the training data and estimates the model parameters. The system estimates $\hat{r}_{ui}$ for each given training case. The prediction error is determined as

$$E_{ui} = r_{ui} - \mu - b_i - b_u - q_i^{m^T} p_j^u \quad (4)$$

The parameters are updated as

$$b_i \leftarrow b_i + \varsigma(E_{ui} - \lambda b_i) \quad (5)$$

$$b_u \leftarrow b_u + \varsigma(E_{ui} - \lambda b_u) \quad (6)$$

$$q_i^m \leftarrow q_i^m + \rho(E_{ui}p_j^u - \lambda q_i^m) \quad (7)$$

$$p_j^u \leftarrow p_j^u + \rho(E_{ui}q_i^m - \lambda p_j^u) \quad (8)$$

where ρ and ς specify constant magnitudes that account proportion by which parameters are modified in the opposite direction of the gradient.

The objective of our matrix factorization model is to predict unknown S3D content ratings, from the trained model obtained by fitting the earlier observed ratings. We determine the regularization constant λ by cross-validation [38].

4. RESULTS AND DISCUSSION

The efficiency of the proposed S3D video recommendation model is evaluated on the Nama3ds1-cospad1 [14] and the LFOVIAS3DPh2 [17] S3D video datasets. The Nama3ds1-cospad1 dataset has 10 reference and 100 test sequences. The video sequences have a resolution of 1920×1080 and are available in .avi containers. The frame rate is 25 fps and the duration is either 13 or 16 seconds. The videos of the dataset are distorted by combinations of H.264, JP2K, scaling and (down-)sampling. These artefacts are applied symmetrically to both views of the S3D videos, and the corresponding Difference Mean Opinion Score (DMOS) scores are provided. The subjective assessment on perceptual quality was performed as single stimulus continuous quality evaluation (SSCQE) with hidden reference method. The test participants
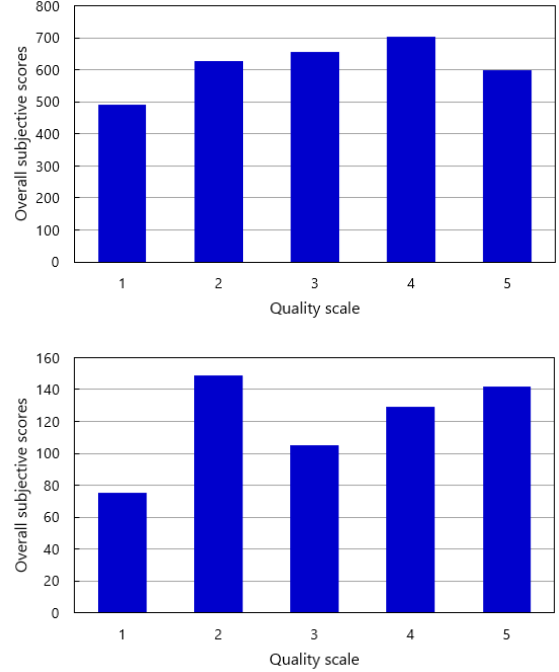

Fig. 1. Subjective score distribution of the datasets. The values of μ, m and σ denote the statistical measures (mean (μ), median (m) and standard deviation (σ)) of the subjective ratings. Top: Nama3ds1-cospad1 dataset [14]; $\mu = 3.09$, $m = 3$, $\sigma = 1.35$. Bottom: LFOVIAS3DPh2 dataset [17]; $\mu = 3.19$, $m = 3$, $\sigma = 1.37$.

used a 5-point Absolute Category Rating (ACR) to rate the perceived quality of S3D videos. A total of 28 subjects were involved in the study. The dataset contains each subjective rating, not just aggregated values. The LFOVIAS3DPh2 S3D video dataset has 12 reference sequences with a good variety of structure, texture, depth and temporal information. The videos have a resolution of 1920×1080 and a duration of 10 seconds with a frame rate of 25 fps. In total 288 test stimuli were created by applying H.264 and H.265 compressions, blur and frame freeze distortions. The dataset is a combination of symmetric and asymmetric S3D videos. The SSCQE method was used to perform the subjective tests, which were completed by 20 test participants. Similarly, each subjective quality rating is available, along with the DMOS values.

Figure 1 shows the distribution of the subjective quality scores for the Nama3ds1-cospad1 and the LFOVIAS3DPh2 datasets. From the plot, it is clear that both datasets are diverse in the range of perceived quality. Furthermore, the estimated statistics (μ, m and σ) of each dataset are consistent and follow the observed trends.

The left side of Figure 2 shows the first frame from left view of the 'Hall' S3D video sequence from the Nama3ds1-cospad1 dataset in reference quality, and the corresponding subjective assessment ratings and the ratings predicted by our novel model. The right side of the figure shows the same,
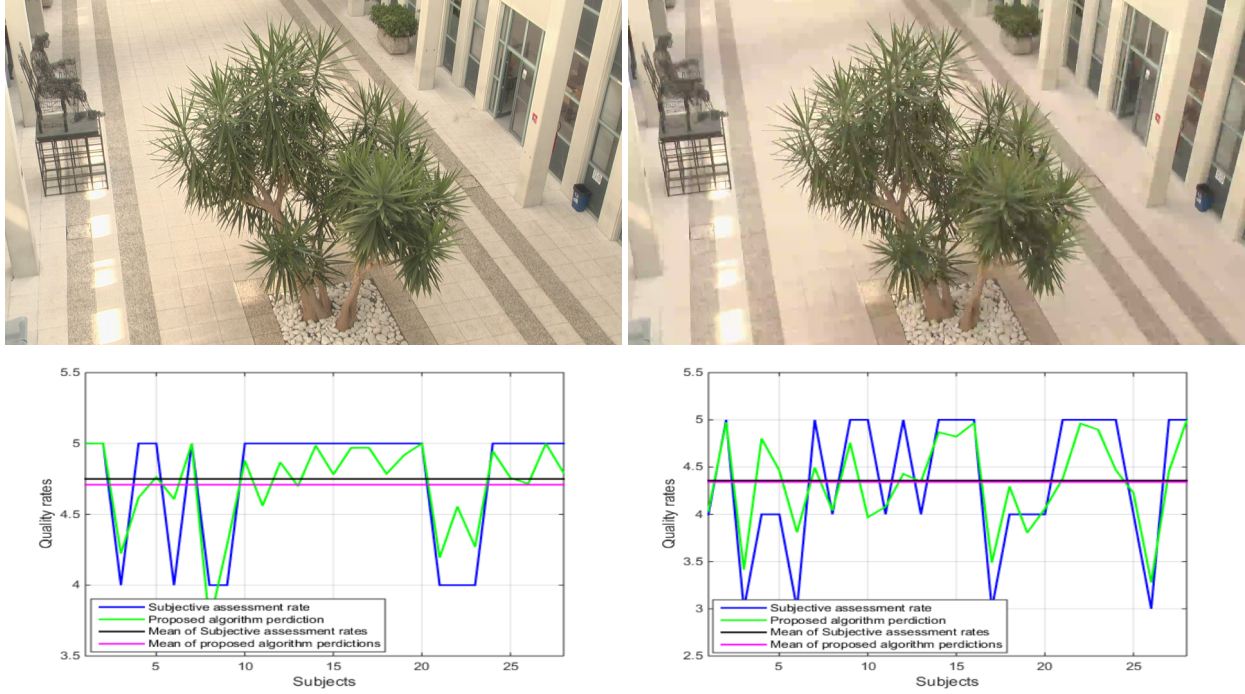
Fig. 2. Perceptual opinion score of each test participant and the corresponding predicted scores for the 'Hall' S3D video sequence from the Nama3ds1-cospad1 [14] dataset, and the assessed video frames. Left: Reference quality. Right: Distorted quality (H.264 compression).

Table 1. Performance of the proposed algorithm on the subjective scores of the Nama3ds1-cospad1 and the LFOVIAS3DPh2 datasets.

Score	Training Set			Testing Set		
	LCC	SROCC	RMSE	LCC	SROCC	RMSE
Nama3ds1-cospad1 [14]	0.8873	0.8858	0.6903	0.8753	0.8700	0.7527
LFOVIAS3DPh2 [17]	0.8966	0.8911	0.5809	0.8585	0.8288	0.6522

Table 2. Performance of the proposed algorithm on each distortion type of the LFOVIAS3DPh2 dataset.

Type	H.264	H.265	Blur	FF	Overall
LCC	0.910	0.904	0.877	0.908	0.858
SROCC	0.902	0.914	0.828	0.899	0.828
RMSE	0.357	0.384	0.423	0.371	0.652

but with H.264 compression (Quantization Parameter = 38). Based on these results, it is clear that the proposed algorithm accurately predicts the subjective quality ratings of the videos. It is also important to point out that the deviation between the average scores of subjective ratings and the predictions of the model is quite low.

The performance of the proposed algorithm is measured by using the Linear Correlation Coefficient (LCC), the Spearman Rank Order Correlation Coefficient (SROCC) and the Root Mean Square Error (RMSE). The LCC indicates the linear dependence between two quantities. The SROCC mea-

sures the monotonic relationship between two input sets. The RMSE measures the magnitude error between estimated scores and subjective scores. Higher LCC and SROCC values suggest good agreements between subjective and objective measures, and lower RMSE values signify more accurate prediction performance. For both databases, 80% of the scores provided by the test participants is used for algorithm training and the remaining samples are used for testing. These scores do not overlap, since using the same samples for both training and testing is not a valid option. We performed the random assignment for 100 trials of each epoch for statistical consistency and repeated it for 200 epochs. Finally, we calculated the mean of the LCC, the SROCC and the RMSE measures of all epochs to report the performance analysis. Table 1 shows the performance evaluation of the proposed algorithm on the training and test sets of the Nama3ds1-cospad1 and the LFOVIAS3DPh2 datasets. The results indicate that the proposed algorithm shows robust performance across both datasets.
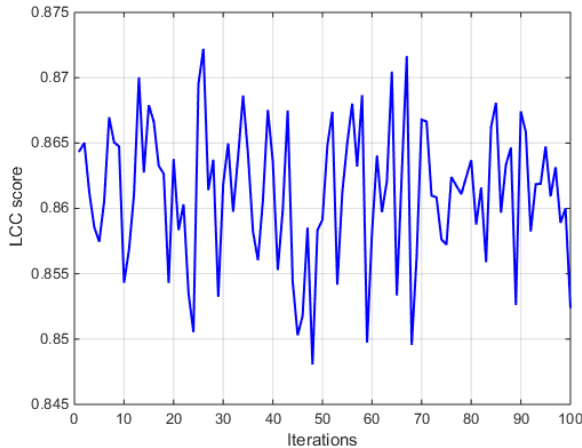
Fig. 3. Variation of the proposed algorithm's LCC score over 100 trails of an epoch. The standard deviation of the LCC score over 100 trails is 2×10^{-4}.
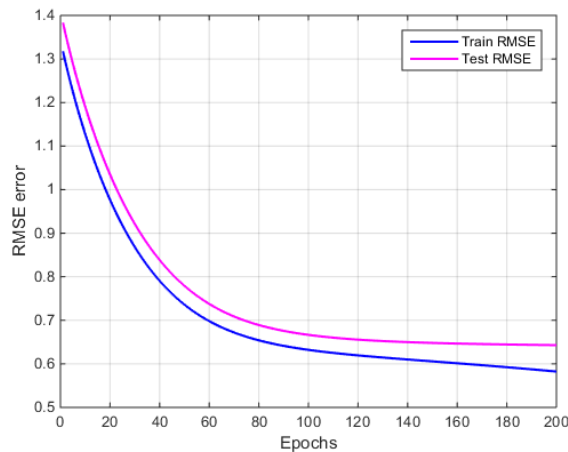

Fig. 4. RMSE error variation across epochs.

Figure 3 shows the LCC score variation of 100 iterations of an epoch. The figure reveals that the scores are consistent across all iterations, and furthermore, we experienced the lower standard deviation (2×10^{-4}) of 100 LCC scores. Figure 4 shows the average training and test RMSE measure variation over 200 epochs. The results indicate that both RMSE errors reduced with the number of epochs. Hence, these results further demonstrate the efficiency of the proposed algorithm. This is also aided by Table 2, which summarizes the performance values on each distortion type of the LFOVIAS3DPh2 dataset. The performance of the model is consistent across all distortion types of the dataset.

5. CONCLUSION

The paper presented a novel recommendation system for S3D movies. According to the best knowledge of the authors, this is the first implemented and evaluated objective quality assessment metric to consider the subjective ratings of 3DTV viewers for specific visual artefacts in order to predict personal preference and toleration. The spatial and temporal distortion types covered by this scientific contribution are JPEG, H.264 and H.265 compressions, resolution reduction, sampling, sharpening, blur and frame freeze. All of these distortions may adversely affect the S3D video signal at different stages of the content generation and delivery chain. The experimental results indicate that the proposed model based on matrix factorization provides enhanced accuracy for S3D video recommendation. Therefore, the model may support content providers such as Netflix, Amazon, TiVo, allowing them to efficiently minimize the flood of low-quality 3D contents.

In the future, we aim to further extend this recommendation system by considering the detailed analysis of commercial S3D movies. The model is to be improved by offering per-frame analysis of artefacts that may potentially cause visual discomfort. These effects include – but are not limited to – large horizontal disparity, vertical parallax, the cross-talk effect, the cardboard effect, stuck-to-background objects, stereo window violation, depth continuity, and many more. Existing metrics already report that such effects may result poor ratings. Combining objective and subjective scores shall help to further reduce the error rate while recommending new S3D movies.

Additionally, we shall perform affective analysis on the emotional reactions of 3DTV and VR viewers. We shall take into account both subjective scores and brain-activity measurements to understand the dependencies between the degree of viewer discomfort and the intensity of the distortions. This shall improve the classification of viewer's from different age groups by their susceptibility to artefacts and content types.

Moreover, we shall design new experiments to work on evaluation models like probabilistic matrix factorization (PMF) in order to further improve prediction accuracy. One specific experiment shall focus on the linear combinations of predictions of multiple PMF models with predictions of Restricted Boltzmann Machine (RBM) models. This approach could also significantly improve the accuracy of the performed predictions.

6. REFERENCES

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607–625, 2011.

[2] M. Sharma, S. Chaudhury, B. Lall, and M. Venkatesh, "A flexible architecture for multi-view 3DTV based on uncalibrated cameras," *Journal of Visual Communication and Image Representation*, vol. 25, no. 4, pp. 599–621, 2014.

[3] M. Sharma, "Uncalibrated camera based content generation for 3D multi-view displays," Ph.D. dissertation, 2017.

[4] M. Sharma, S. Chaudhury, and B. Lall, "A novel hybrid kinect-variety-based high-quality multiview rendering scheme for glass-free 3D displays," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2098–2117, 2016.

[5] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database," in *fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012, pp. 212–217.

[6] A. Antsiferova and D. Vatolin, "The influence of 3D video artifacts on discomfort of 302 viewers," in *International Conference on 3D Immersion*. IEEE, 2017, pp. 1–8.

[7] "MSU 3D-video quality analysis, video quality measurement tool 3D project, MSU graphics & media lab (video group), http://compression.ru/video/vqmt3d/."

[8] A. Gotchev, G. B. Akar, T. Capin, D. Strohmeier, and A. Boev, "Three-dimensional media for mobile devices," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 708–741, 2011.

[9] F. Zeri and S. Livi, "Visual discomfort while watching stereoscopic three-dimensional movies at the cinema," *ophthalmic and Physiological Optics*, vol. 35, no. 3, pp. 271–282, 2015.

[10] M. Barreda-Angeles, R. Pépion, E. Bosc, P. Le Callet, and A. Pereda-Banos, "How visual discomfort affects 3DTV viewers' emotional arousal," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE, 2014, pp. 1–4.

[11] B. Appina, K. Manasa, and S. S. Channappayya, "Subjective and objective study of the relation between 3D and 2D views based on depth and bitrate," *Electronic Imaging*, vol. 2017, no. 5, pp. 145–150, 2017.

[12] S. Jumisko-Pyykkö, T. Haustola, A. Boev, and A. Gotchev, "Subjective evaluation of mobile 3D video content: depth range versus compression artifacts," in *Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V*, vol. 7881. International Society for Optics and Photonics, 2011.

[13] G. Sanchez, J. Silveira, L. V. Agostini, and C. Marcon, "Performance analysis of depth intra-coding in 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2509–2520, 2018.

[14] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutierrez, and N. Garcia, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *Fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012, pp. 109–114.

[15] J. Li, M. Barkowsky, and P. Le Callet, "Visual discomfort in 3DTV: Definitions, causes, measurement, and modeling," in *Novel 3D Media Technologies*. Springer, 2015, pp. 185–209.

[16] A. Bokov, D. Vatolin, A. Zachesov, A. Belous, and M. Erofeev, "Automatic detection of artifacts in converted S3D video," in *Stereoscopic Displays and Applications XXV*, vol. 9011. International Society for Optics and Photonics, 2014, p. 901112.

[17] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya, and A. C. Bovik, "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5027–5040, 2019.

[18] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," in *Electronic Imaging*. International Society for Optics and Photonics, 2006.

[19] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[20] B. Appina, M. K., and S. S. Channappayya, "A full reference stereoscopic video quality assessment metric," in *International Conference on Acoustics, Speech and Signal Processing, IEEE*, March 2017, pp. 2012–2016.

[21] B. Appina and S. Channappayya, "Full-reference 3D video quality assessment using scene component statistical dependencies," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 823–827, June 2018.

[22] L. E. M. Fernández, "Recommendation system for netflix," *Faculty of Science Business Analytics, Vrije Universiteit Amsterdam*, 2018.

[23] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[24] Y. Lu, R. Dong, and B. Smyth, "Convolutional matrix factorization for recommendation explanation," in *Proceedings of International Conference on Intelligent User Interfaces Companion*, 2018, pp. 1–2.

[25] J. Lee, N. Kothari, and P. Natsev, "Content-based related video recommendations," *Advances in Neural Information Processing Systems Demonstration Track*, 2016.

[26] I. Tsingalis, I. Pipilis, and I. Pitas, "A statistical and clustering study on youtube 2D and 3D video recommendation graph," in *International Symposium on Communications, Control and Signal Processing*. IEEE, 2014, pp. 294–297.

[27] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston *et al.*, "The youtube video recommendation system," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 293–296.

[28] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in *Proceedings of ACM conference on recommender systems*, 2016, pp. 191–198.

[29] J. G. Estrada and A. L. Simeone, "Recommender system for physical object substitution in VR," in *IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 359–360.

[30] J. Niu, S. Wang, Y. Su, and S. Guo, "Temporal factor-aware video affective analysis and recommendation for cyber-based social media," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 3, pp. 412–424, 2017.

[31] Z. Zhang, S. Shang, S. R. Kulkarni, and P. Hui, "Improving augmented reality using recommender systems," in *Proceedings of ACM conference on Recommender systems*, 2013, pp. 173–176.

[32] Z. Shi, H. Wang, W. Wei, X. Zheng, M. Zhao, and J. Zhao, "A novel individual location recommendation system based on mobile augmented reality," in *International Conference on Identification, Information, and Knowledge in the Internet of Things*. IEEE, 2015, pp. 215–218.

[33] D. Chatzopoulos and P. Hui, "Readme: A real-time recommendation system for mobile augmented reality ecosystems," in *Proceedings of ACM international conference on Multimedia*, 2016, pp. 312–316.

[34] D. Bertsimas, C. Pawlowski, and Y. D. Zhuo, "From predictive methods to missing data imputation: an optimization approach," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 7133–7171, 2017.

[35] A. Paterek, "Improving regularized singular value decomposition for collaborative filtering," in *Proceedings of KDD cup and workshop*, vol. 2007, 2007, pp. 5–8.

[36] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 426–434.

[37] B. Jin and X. Lu, "On the regularizing property of stochastic gradient descent," *Inverse Problems*, vol. 35, no. 1, 2018.

[38] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in neural information processing systems*, 2008, pp. 1257–1264.