FOVEA WEIGHTING OF MULTIVIEW COMPUTATIONAL DISPLAYS FOR ENHANCED USER EXPERIENCE

Fangzhou Luo *Xiaolin Wu*

McMaster University

ABSTRACT

A challenge for multiview displays is how to concurrently exhibit many different views on the same physical medium without sacrificing image quality for individual viewers. To address the above challenge, we propose a novel scheme of fovea weighting in the framework of the TPVM multiview computational display to enhance users' visual experiences. Underlying our new design is an optimization problem of nonnegative matrix factorization. This seemingly difficult problem turns out to be solvable by efficient algorithms. The effectiveness of the proposed multiview fovea weighting is validated by simulation results.

Index Terms— Computational displays, multiview displays, fovea vision, perceptual quality, optimization.

1. INTRODUCTION

Recent years have seen intensified research on and burgeoning commercial interests in a new generation of computational displays, driven by a wide range of virtual, augmented and mixed reality applications in diversified fields from man-machine interactions, medicine, entertainment, to automobile, etc [1]. Although computational displays fall into the domain of multidimensional signal processing, up to now the topic has mostly been studied by researchers in the fields of virtual/augmented reality (VR/AR) and computer graphics with task-specific motivations. The field of computational displays seems to attract very little attention of the image processing research community.

In this work we are interested in image processing methods to enhance functionalities and user experiences of multiview computational displays [2, 3]. The traditional stream multiview display methods, which were historically motivated by stereoscopic presentations, fall into two categories: spatial multiplexing (e.g., lenticular lenses and parallax barrier) and temporal multiplexing (e.g., active liquid crystal shutter glasses and polarization shutter glasses) [4, 5]. Recently, a new breed of multiview computational displays have emerged, including the tensor display [2] and the display of temporal psychovisual modulation (TPVM) [3]. These new computational display systems go beyond simple spatial or temporal multiplexing to generate multiple different views; instead, they decompose (encode) multiple concurrent views into a relatively small number of basis or tensor images. The resulting basis or tensor images are displayed at a high refresh rate exceeding the critical flicker frequency; different users perceive different images concurrently on the same display medium by properly fusing (decoding) these basis images. The assumption of computational displays is that the set of concurrent views being exhibited are highly correlated or permit sparse representations. In computational multiview display systems, the perceptual quality is primarily limited by the number of basis images, which is limited by the speeds of physical optoelectronic devices, i.e., the refresh rate of the display and spatial light modulators.

The main technical challenge for computational multiview displays is how to achieve high perceptual quality for all viewers while working with only a small number of basis images. We take on this problem in a perspective of subjective quality. In many applications of multiview displays, different users often have their own regions of interest at any given time. This allows us to exploit the well-known property of rapidly decreasing visual acuity from fovea to peripheral vision [6], and propose a spatially weighted optimization algorithm for multiview computational display. The proposed algorithm chooses the basis images and their fusion scheme in such a way that different concurrent views are exhibited at highest quality in viewers' focused regions, while allowing graceful image quality degradation in regions of peripheral vision.

The above scheme of fovea weighting can be integrated into both the tensor display and the TPVM display algorithms to improve subjective image quality. In what follows, we restrict our technical development to the TPVM-based multiview display systems, because the problem is considerably simpler for the TPVM display than for the tensor display, and because the TPVM displays can be made very large in size and hence more suitable for multiview applications.

The remainder of this paper is structured as follows. Section 2 outlines the principle and architecture of the TPVM multiview computational system. Section 3 presents the fovea weighting algorithm for optimizing multiuser

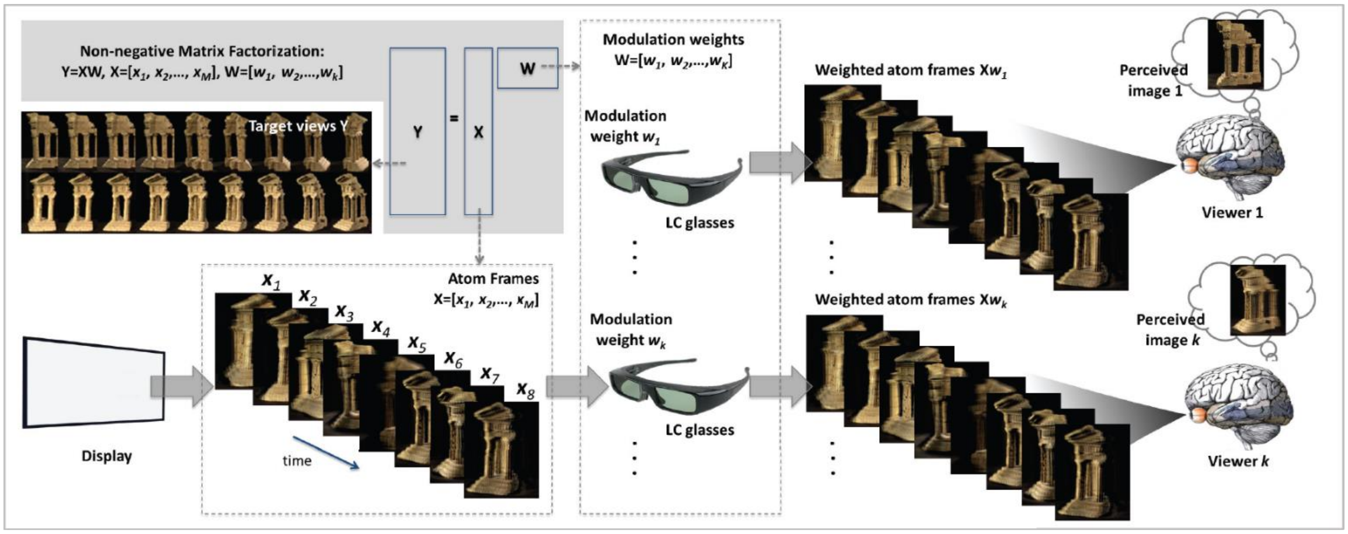


Figure 1: Image formation by TPVM.

perceptual quality. Section 4 reports experimental results, and Section 5 concludes.

2. TPVM MULTIVIEW DISPLAY

Before presenting our new fovea weighting scheme for multiview computational displays, it is necessary to briefly introduce the TPVM multiview display system, which is schematically illustrated in Figure 1. This computational display system can concurrently exhibit many different views of a scene (e.g., a virtual environment in VR) on a common display surface. These views are decomposed into a set of atom frames (basis images) to be displayed at a high frame rate exceeding the critical flicker frequency of 60 Hz for human eyes. Different users perceive their own head-tracking views through liquid crystal (LC) light modulation glasses. The LC glasses, synchronized with the high-speed display, can regulate how much of the light energy of each atom frame to pass through and reach retina, namely, perform amplitude modulation of atom frames, so that the human visual system (HVS) can fuse these modulated atom frames into desired images. The TPVM display system simplifies the end user device from head-mounted display (HMD) to light, simple LC glasses. As the glasses are see-through, the participants can conduct face-to-face communications or even body-to-body interactions, the reconstruction of co-presence in VE reduces to perceptual fusion of the virtual world and the participants' own physical proximity.

At the heart of TPVM is a problem of non-negative matrix factorization (NMF) [7]. Let $Y = (y_1, y_2, \ldots, y_K)$ be the K target images to be concurrently displayed to different viewers. The $N \times K$ matrix Y, where N is the number of pixels in each target image, needs to be decomposed into $Y = XW$, with the $N \times M$ matrix $X = (x_1, x_2, \ldots, x_M)$ being the set of basis images and the $M \times K$ matrix $W = (w_1, w_2, \ldots, w_K)$ being the K modulation coefficient vectors corresponding to the K target images. The resulting basis images $x_1, x_2, \ldots, x_M$ are cyclically displayed at a refresh rate above 60 M Hz, and the corresponding 2D optical signals are temporally modulated by active LC glasses according to weights $w_1, w_2, \ldots, w_K$. This optoelectronic display-glass coupling and the psychovisual temporal fusion mechanism of HVS jointly render the K concurrent target images $y_1, y_2, \ldots, y_K$ as different linear combinations of the $x_1, x_2, \ldots, x_M$ basis images. By now one can appreciate that in the TPVM paradigm, all heavy computations involved in multiview generation are delegated to a central server. End user devices become inexpensive, light LC glasses that are controlled by a modulation vector that only consumes a negligible bandwidth.

In practice, the image decomposition underlying TPVM has to respect a condition of non-negativity, because the light energy emitted by the display cannot be negative, and active LC glasses can only implement modulation weights between 0 and 1. Therefore, the introduced display system needs to solve the following problem of NMF

$$\min_{X,W} \|Y - XW\|_F^2 \, , subject\ to\ 0 \le X, W \le 1 \qquad (1)$$

3. FOVEA WEIGHTING OF TPVM

In this section, we discuss how to apply fovea weighting in the framework of TPVM; namely, formulate a weighted objective function and devise efficient algorithms to minimize it.

Assuming that the users of the TPVM multiview display system are all equipped with eye tracking devices of sufficiently high accuracy and response speed [8-10], thus the region of interests (ROI) for each viewer is known. For viewer k, the approximation error image $y_k - Xw_k$ is spatially weighted by a 2D Gaussian function that is aligned with the ROI of viewer k.

Let C be the K corresponding regions of interest of K target images. The $N \times K$ non-negative matrix C, which has the same size as Y, can be used as weight matrix inside the Frobenius norm. Now we need to solve the following optimization problem with ROI information:

$$\min_{X,W} \|C \circ (Y - XW)\|_F^2, subject\ to\ 0 \leq X, W \leq 1 \quad (2)$$

where $\circ$ is the Hadamard product. First, we suppose W is the only optimization variable in our problem, and the objective is to minimize

$$\begin{aligned} S(W) &= \sum_i \|c_i \circ (y_i - Xw_i)\|^2 \\ &= \sum_i \left(c_i \circ (y_i - Xw_i)\right)^T \left(c_i \circ (y_i - Xw_i)\right) \quad (3) \end{aligned}$$

where w_i refers to the ith column of W. Let diagonal matrix $D_i = diag(c_i)$, and the objective is to minimize

$$\begin{aligned} S(W) &= \sum_i (y_i^T D_i - w_i^T X^T D_i)(D_i y_i - D_i X w_i) \\ &= \sum_i (y_i^T D_i^2 y_i - w_i^T X^T D_i^2 y_i - y_i^T D_i^2 X w_i \\ &\quad + w_i^T X^T D_i^2 X w_i) \quad (4) \end{aligned}$$

Note that: $(w_i^T X^T D_i^2 y_i)^T = y_i^T D_i^2 X w_i$ is a scalar and equal to its own transpose, hence $w_i^T X^T D_i^2 y_i = y_i^T D_i^2 X w_i$ and the quantity to minimize becomes

$$S(W) = \sum_i (y_i^T D_i^2 y_i - 2 w_i^T X^T D_i^2 y_i + w_i^T X^T D_i^2 X w_i) \quad (5)$$

Differentiating this with respect to w_i and equating to zero to satisfy the first-order conditions gives

$$\begin{aligned} \frac{\partial S}{\partial w_i} &= -2X^T D_i^2 y_i + 2(X^T D_i^2 X) w_i \\ &= -2X^T diag^2(c_i) y_i + 2(X^T diag^2(c_i) X) w_i \\ &= 0 \quad (6) \end{aligned}$$

Suppose $X^T D_i^2 X$ is positive definite, then we can get solution for w_i

$$\begin{aligned} w_i &= (X^T D_i^2 X)^{-1} X^T D_i^2 y_i \\ &= (X^T diag^2(c_i) X)^{-1} X^T diag^2(c_i) y_i \quad (7) \end{aligned}$$

Suppose X is the only optimization variable in our problem, we can get the partial derivatives and the solution for x_i^T in a similar way.

$$\frac{\partial S}{\partial x_i^T} = -2y_i^T diag^2(c_i^T) W^T + 2x_i^T (W diag^2(c_i^T) W^T) \quad (8)$$

$$x_i^T = y_i^T diag^2(c_i^T) W^T (W diag^2(c_i^T) W^T)^{-1} \quad (9)$$

where x_i^T refers to the ith row of X.

It is worth noting that $diag(c_i)$ has a huge size in our problem. To accelerate the computation of W, we can replace the usage of diagonal matrix with Hadamard product and tiling of copies of c_i. For example, compute $[c_i\ c_i\ \ldots\ c_i] \circ X$ in the program instead of $diag(c_i)X$.

If the fast response of the TPVM multiview display system is paramount, such as in real time multiuser VR applications, then we suggest to use the projection to convex set (POCS) method [11] to solve the weighted NMF problem. In the POCS approach, the non-negativity constraints when alternatingly solving for W and X are relaxed; but after each iteration, the resulting matrices W and X are clipped back to the value range [0, 1]. This allows the weighted NMF problem to be solved very efficiently as a series of least-squares problems.

If a better display performance is needed, then we suggest to use the projected gradient descent method [12] to solve the weighted NMF problem, i.e., perform usual gradient update and then project back onto the convex feasible set. The computation of the gradient needed in this method is discussed in previous section. The projected gradient descent method can obtain a better convergence if a good starting value is chosen. So in practice, we can use the result of POCS method as starting value of the projected gradient descent method. However, it is easy to understand that convergence of gradient-based method can be very slow even learning rate adaptation or momentum technique is used.

Moreover, we find the third way to solve the weighted NMF problem. It is an improved version of POCS method, and has a much faster convergence compared with the projected gradient descent method. In each iteration of the POCS approach, we no longer consider the clipped values as variables but fixed constants. For example, in the POCS approach we solve for the jth column of W without the non-negativity constraints

$$w_j = \left(X^T diag^2(c_j) X\right)^{-1} X^T diag^2(c_j) y_j \quad (10)$$

But for some elements in w_j vector, the value will be clipped later, i.e., for i = 1 ... N (N ≤ M), $w_{a_i j} = b_i = 0\ or\ 1$, where a_i is the row number for the clipped value and b_i is a constant. Then we will have

$$y_j - \sum_{i=1}^{N} x_{a_i} w_{a_i j} = \sum_{i' \neq a_i} x_{i'} w_{i' j} \tag{11}$$

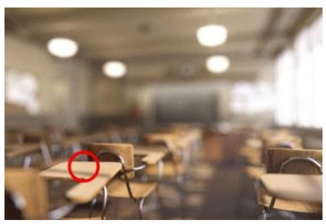
Now we can replace y_j with $y_j - \sum_{i=1}^{N} x_{a_i} w_{a_i j}$ and replace X with $(x_{i'})_{i' \neq a_i}$ in (10), then we can solve a similar problem only for $w_{i' j}$, i.e., non-clipped elements in w_j vector.

4. EXPERIMENTAL RESULTS

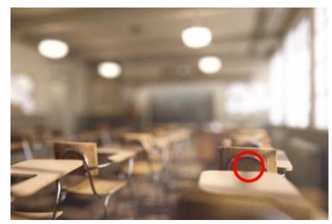
In the experiment setup, we try to accommodate the hardware limitation of the TPVM multiview display system in practice. Although M, the number of basis images, can be any positive integer theoretically, its feasible range is severely limited by the relatively low speed of active LC light modulation glasses. The speed of the off-the-shelf LC glasses of grey levels is difficult to exceed 180Hz. In other words, at the current state of the art, a practical TPVM multiview display system can only run two to three basis images. As such, as said in the very beginning of the paper, the main technical challenge facing TPVM is how to support as many viewers as possible while maintaining an acceptable perceptual quality for all of them, using only a small number of basis images. In the experiment reported below, we demonstrate that fovea weighting can achieve high perceptual quality for four different concurrent views with only two or three basis images, which is impossible with naïve temporal multiplexing as in the main stream stereoscopic displays.

The test images in our experiment are generated by Blender, an open source 3D creation suite. Listed in figures are the rendered images of the 3D demo model “Class Room”, which is available on the Blender website. In the experiment, four users are watching the classroom with four different perspectives and regions of interest. The red circles in Fig. 2 indicate the centers of regions of interest of different users. Fovea weighted PSNRs are reported in Figure 2.

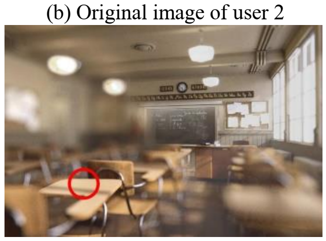

(a) Original image of user 1

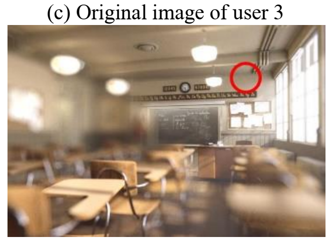

(b) Original image of user 2

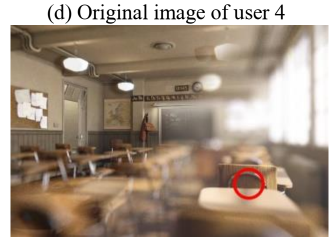

(c) Original image of user 3

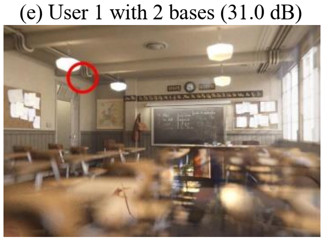

(d) Original image of user 4

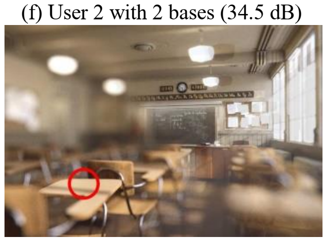

(e) User 1 with 2 bases (31.0 dB)

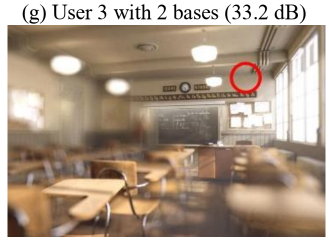

(f) User 2 with 2 bases (34.5 dB)


(g) User 3 with 2 bases (33.2 dB)


(h) User 4 with 2 bases (31.7 dB)


(i) User 1 with 3 bases (38.2 dB)


(j) User 2 with 3 bases (35.8 dB)


(k) User 3 with 3 bases (33.8 dB)

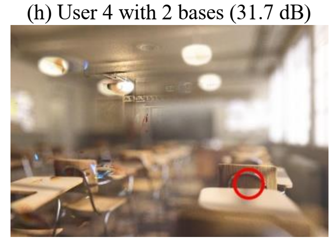

(l) User 4 with 3 bases (42.3 dB)

Figure 2: Results of fovea weighting of TPVM. The red circles indicate the centers of regions of interest of different users. The PSNRs reported from (e) to (l) are fovea weighted PSNR.

5. CONCLUSIONS

We introduce a new and effective strategy of multiview fovea weighting to improve subjective image quality of the TPVM multiview computational displays. This gives rise to an optimization problem of nonnegative matrix factorization. This seemingly difficult problem is shown to have efficient solutions. Empirical results demonstrate the advantages and usefulness of multiview fovea weighting for computational multiview displays.

6. REFERENCES

[1] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers & Graphics,* vol. 37, pp. 1012-1038, 2013.

[2] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," 2012.

[3] X. Wu and G. Zhai, "Temporal psychovisual modulation: A new paradigm of information display [exploratory dsp]," *IEEE Signal Processing Magazine,* vol. 30, pp. 136-141, 2013.

[4] N. A. Dodgson, "Autostereoscopic 3D displays," *Computer,* vol. 38, pp. 31-36, 2005.

[5] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proceedings of the IEEE,* vol. 99, pp. 540-555, 2011.

[6] F. W. Weymouth, "Visual sensory units and the minimal angle of resolution," *American journal of ophthalmology,* vol. 46, pp. 102-113, 1958.

[7] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature,* vol. 401, pp. 788-791, 1999.

[8] A. Duchowski, *Eye tracking methodology: Theory and practice* vol. 373: Springer Science & Business Media, 2007.

[9] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford, 2011.

[10] M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath, "Depth perception with gaze-contingent depth of field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 217-226.

[11] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM review,* vol. 38, pp. 367-426, 1996.

[12] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation,* vol. 19, pp. 2756-2779, 2007.