

Accelerating decomposition of light field video for compressive multi-layer display

Xuan Cao,^{1,3} Zheng Geng,^{1,4} Tuotuo Li,² Mei Zhang,¹ and Zhaoxing Zhang¹

¹State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

²Intel labs, 2200 Mission College Blvd, Santa Clara, 95054, USA

³caoxuan2012@ia.ac.cn

⁴zheng.geng@ia.ac.cn

Abstract: Compressive light field display based on multi-layer LCDs is becoming a popular solution for 3D display. Decomposing light field into layer images is the most challenging task. Iterative algorithm is an effective solver for this high-dimensional decomposition problem. Existing algorithms, however, iterate from random initial values. As such, significant computation time is required due to the deviation between random initial estimate and target values. Real-time 3D display at video rate is difficult based on existing algorithms. In this paper, we present a new algorithm to provide better initial values and accelerate decomposition of light field video. We utilize internal coherence of single light field frame to transfer the ignorance-to-target to a much lower resolution level. In addition, we explored external coherence for further accelerating light field video and achieved 5.91 times speed improvement. We built a prototype and developed parallel algorithm based on CUDA.

©2015 Optical Society of America

OCIS codes: (070.6120) Spatial light modulators; (100.6950) Tomographic image processing; (120.2040) Displays; (110.1758) Computational imaging.

References and links

1. J. Geng, "Three-dimensional display technologies," *Adv. Opt. Photonics* **5**(4), 456–535 (2013).
2. A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360° light field display," *ACM Trans. Graph.* **26**(3), 40 (2007).
3. A. Jones, M. Lang, G. Fyffe, X. M. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec, "Achieving eye contact in a one-to-many 3D video teleconferencing system," *ACM Trans. Graph.* **28**(3), 64 (2009).
4. J. Geng, "A volumetric 3D display based on a DLP projection engine," *Displays* **34**(1), 39–48 (2013).
5. C. Van Berkel, "Image preparation for 3D-LCD," *Proc. SPIE* **3639**, 84–91 (1999).
6. C. Van Berkel and J. A. Clarke, "Characterisation and optimisation of 3D-LCD module design," *Proc. SPIE* **3012**, 179–186 (1997).
7. Z. X. Zhang, Z. Geng, M. Zhang, and H. Dong, "An interactive multiview 3D display system," *Proc. SPIE* **8618**, 86180P (2013).
8. A. Jones, K. Nagano, J. Liu, J. Busch, X. M. Yu, M. Bolas, and P. Debevec, "Interpolating vertical parallax for an autostereoscopic three-dimensional projector array," *J. Electron. Imaging* **23**(1), 011005 (2014).
9. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models Visual Processing* (MIT Press, 1991), pp. 3–20.
10. M. Levoy and P. Hanrahan, "Light field rendering," *Proc. SIGGRAPH* (1996), pp. 31–42.
11. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.* **31**(4), 80 (2012).
12. M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Trans. Graph.* **33**(4), 58 (2014).
13. G. P. Bell, "Advanced metrics-based design methodology for multilayer 3D displays," *Proc. SPIE* **5443**, 239–248 (2004).
14. G. P. Bell, R. Craig, R. Paxton, G. Wong, and D. Galbraith, "Beyond flat panels: multi-layered displays with real depth," *SID Symposium Digest of Technical Papers* **39**(1), 352–355 (2008).
15. D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Trans. Graph.* **29**(6), 163 (2010).
16. G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: tomographic image synthesis for

- attenuation-based light field and high dynamic range displays,” *ACM Trans. Graph.* **30**(4), 95 (2011).
17. X. Cao, Z. Geng, M. Zhang, and X. Zhang, “Load-balancing multi-LCD light field display,” *Proc. SPIE* **9391**, 93910F (2015).
 18. G. Lippmann, “La photographie integrale,” *Comptes Rendus Acad. Sci.* **146**, 446–451 (1908).
 19. Y. Kim, J. Kim, J. M. Kang, J. H. Jung, H. Choi, and B. Lee, “Point light source integral imaging with improved resolution and viewing angle by the use of electrically movable pinhole array,” *Opt. Express* **15**(26), 18253–18267 (2007).
 20. T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Rev.* **51**(3), 455–500 (2009).
 21. V. Blondel, N. D. Ho, and P. Vandoren, “Weighted nonnegative matrix factorization and face feature extraction,” *Image Vis. Comput.* **2007** 1–17 (2008).
 22. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, “Real-time image generation for compressive light field displays,” *J. Phys. Conf. Ser.* **415**, 012045 (2013).
-

1. Introduction

Autostereoscopic display, as a glass-free 3D display technology, extended human’s visual perception significantly by emitting correct views toward to multiple directions [1]. To date, autostereoscopic display technologies project multiple views mainly depend on two strategies including 1) time or spatial multiplex and 2) array of projectors. As to the first strategy, time-multiplex inevitably utilizes mechanical motion [2–4]. Spatial-multiplex requires no mechanical motion but results in much lower image resolution, like lenticular lens display [5,6]. As to the second strategy, projector-array makes up the deficiency of spatial and angle resolution at price of linearly increased hardware cost and floor space [7,8].

The two schemes mentioned above have to trade off between quality of 3D display and hardware cost, hindering the commercialization of 3D display technology. The fundamental problem resulting in this trade-off is that the dimension of 3D content outdistances the display capability of single 2D display device (e.g. LCD or projector). The complete light transport in our three-dimensional world can be parameterized by 7D plenoptic function [9] which has higher dimension than 2D plane display device. In comparison to plenoptic function, 4D light field [10] is a subset of light transport, remaining beyond what can be projected by 2D display device.

Compressive light field display [11,12] technology mitigates the tradeoff between the spatial resolution and hardware cost as well as floor space. Multiple Layer Display (MLD) [13,14] breaks the barrier of plane display device, providing an effective and inexpensive solution for compressive light field display. For examples, dual LCD layers displayed a low-rank light field [15] and multiple layers projected a light field containing 7x7 viewpoints [16].

MLD operates as a spatial light modulator (SLM) by combining pixels in different layers to emit different light rays. In other words, comparing to time or spatial multiplex 3D display, MLD is based on pixel multiplex. The dimensions of light field (4D) are larger than dimensions of MLD (3D). Before displaying light field based on MLD, the target light field needs to be compressed (decomposed) into layer images. That is why we call this technology compressive light field display. The most challenging problem of compressive MLD is light field decomposition which renders massive light rays and compresses the 4D light field into multiple LCD layers. An effective decomposition solver plays a significant role for light field display. For example, non-negative tensor factorization is an effective and meaningful solver for light field decomposition. It starts iteration for optimization with initial estimates and converges to an optimal solution [11]. But the initial values are random noise which takes huge computation time to converge.

Our works touches on compressive light field display based on MLD. In this paper, we focus on accelerating light field video decomposition. In our experiments, we found that iteration from random initial values may take very long computation time due to utterly ignorance to the target values in the beginning. Our new algorithm explores the coherence in light field and starts iteration with better initial values. We construct a resolution pyramid to utilize the internal coherence and transfer the ignorance-to-target to a much lower resolution

level. We also explore the external coherence between adjacent light field frames to accelerate decomposition of light field video. Initial values figured out based on coherence are much closer to the target values than that of random noise. Inspired by [17], we further utilize spatial independence within light field to implement pixel-parallel decomposition based on GPU.

2. Related work

To date, LCD panel might be the most suitable electronic device for multiple layer display due to its transparency and pixel-independent controllability. By combining integral imaging technology [18], Kim replaced the static pinhole array with LCD panel and drove it as dynamic barrier improving image resolution and viewing angle [19]. Lanman and Wetzstein operated multiple LCD panels as tomographic light modulator to attenuate the light rays emitted from the backlight [15,16]. Lanman decomposed light field with 5×3 views, each with resolution of 840×525 , on an Intel Xeon 8-core 3.2 GHz processor with 8 GB of RAM. The optimization took approximately 10 seconds per iteration and at least 50 iterations are required for the PSNR to exceed 30 dB. Totally an average eight minutes was needed for single frame decomposition [15]. There is great room to develop live light field video display. Wetzstein further developed an iterative scheme based on non-negative tensor factorization (NTF) [[20,21] beginning with random initial values [11]. Non-negative tensor factorization provides physically meaningful solver for light field decomposition. But random initial values take more number of iterations to achieve convergence and may fall into a local-optimal solution. Wetzstein implemented GPU-based NTF solver by OpenGL and Cg displaying simple geometry model (e.g. teapot) interactively [11]. But it's not image-based light field decomposition. Cao developed a multi-zone and multi-layer joint optimization for light field decomposition which resolves the light field by spatially dividing the light field into multiple subzones [17]. Cao's multi-zone strategy accelerates light field decomposition but still starts iteration from random initial values. There is huge potential for accelerating light field video if closer-to-target initial values are utilized. Wetzstein implemented video-rate (24fps) light field decomposition but with lower spatial resolution and less viewpoints and layers ($3 \times 3 \times 320 \times 240$ with 2 layers) [22]. In this paper, we tried to accelerate image-based light field video decomposition on MLD without sacrifice of resolution of light field. This attempt may lead to a good direction for future researches to achieve video-rate live light field decomposition and 3D display.

3. Multiple layer display

Multiple layer display is not a novel concept, but becoming popular recently along with the development of compressive light field display [11,12], as shown in Fig. 1. Multiple layer LCD can be depicted as 3D cube matrix and light field is a 4D tensor. Compressive light field display is to project / reconstruct the 4D target light field from 3D LCD stacks. What we discuss in this paper is the inverse problem, how to decompose the target light field into LCD layer images? To put it simply, light field is depicted as a set of images captured from different viewpoints (e.g. Messerschmitt car with 7×7 viewpoints as shown in Fig. 1). There is horizontal and / or vertical parallax between different view images, as shown in Fig. 2. We define the target light field as $L_T(v_x, v_y, r_x, r_y)$ including $v_x \times v_y$ view images, each view image has resolution of $r_x \times r_y$. Totally $v_x \times v_y \times r_x \times r_y$ light rays need to be rendered for single light field frame.

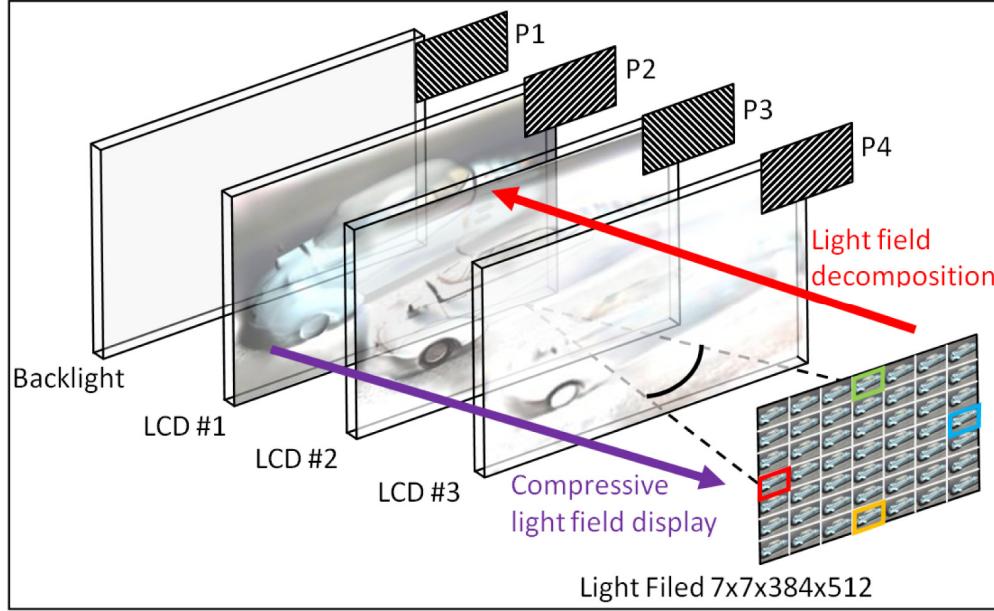


Fig. 1. Schematic diagram of compressive light field display based on multiple LCDs.

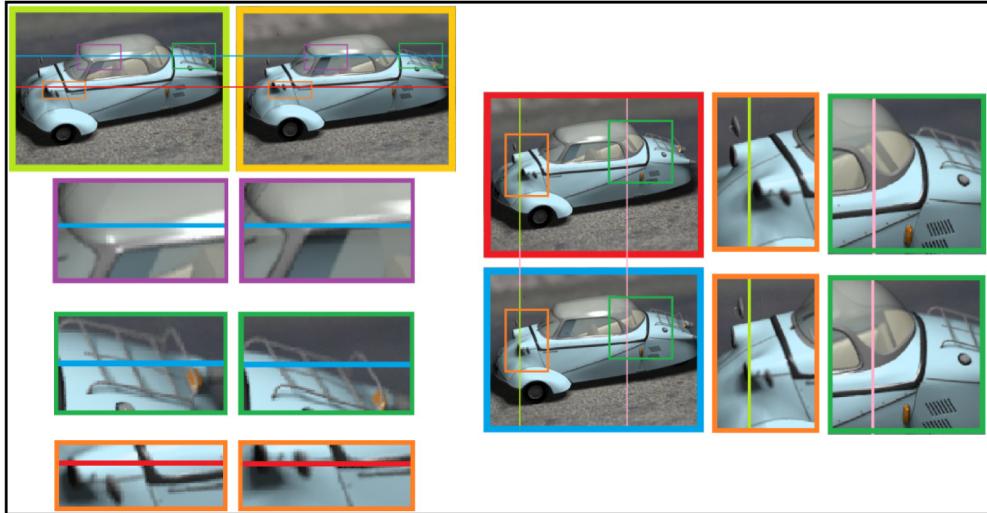


Fig. 2. Parallax of reconstructed light field by multiple layer images.

We evaluate the size of light field mainly by three factors: spatial resolution, viewing scope and angular resolution. Spatial resolution is the image resolution of single view. We depict viewing scope by horizontal and vertical angle of the frustum in which viewers can perceive correct 3D effect. For example, the horizontal angle of viewing scope is illustrated by the black dotted line in Fig. 1, and the angular vertex is located at the center of #1 LCD. Angular resolution is the number of viewpoints in fixed viewing scope. More viewpoints in fixed viewing scope mean higher angular resolution. Low angular resolution is the chief reason resulting in visual discontinuity and jump. In our case, there are 7 viewpoints within 10 degree angle in horizontal direction. The adjacent viewpoints have 1.67 degree interval. At a viewing distance of 2 meters, the spatial distance of adjacent viewpoints is about 29mm. When the number of viewpoints decreasing, the spatial distance of adjacent viewpoints increase. Then

people will perceive the visual discontinuity when they move heads. According to the definition of three factors evaluating the size of light field, the messerschmitt car shown in Fig. 1 has spatial resolution of 384x512, viewing scope of horizontal and vertical 10 degrees, and angular resolution of 0.7x0.7 views per degree.



Fig. 3. Layer images figured out by light field decomposition.

Before displaying light field, the most challenging problem is to decompose the light field into multiple layer images, as shown in Fig. 3. Decomposition of light field takes huge computation time. The computation cost increases linearly with the growing of light field size approximately. A larger size of light field (higher spatial/angle resolution and larger viewing scope) consumes huger computation cost. Displaying one light field frame (e.g. 7x7 views messerschmitt car in Fig. 1) needs one set of layer images (e.g. three LCD images in Fig. 3). Displaying a light field video requires a sequence of layer images sets.

Multiple LCD panels work as a spatial light modulator which attenuates the light rays emitted from backlight source, as shown in Fig. 4. On the assumption that the backlight has uniform illumination, the intensity of light rays is determined by the pixels on three LCD panels jointly, as expressed in Eq. (1).

$$\begin{cases} l_1 = \{a_2, b_1, c_1\} \\ l_2 = \{a_1, b_1, c_2\} \\ l_3 = \{a_2, b_2, c_3\} \end{cases} \quad (1)$$

where l_i is the i -th light ray, a_i, b_i, c_i are the pixels on three LCD panels along with the i -th light ray. The operator $\{a, b, c\}$ represents different rules of operation according to the setup of polarizing films between LCD panels. There are two operation fields including polarization field and production field. In polarization field, there are only two orthometric polarizing films mounted on back of #1 panel and front of #3 panel, as the P1 and P4 shown in Fig. 1. Simply $\{a, b, c\} = a + b + c$ in polarization field. In production field, there are not only the two orthometric polarizing films (P1 and P4 shown in Fig. 1) mentioned above, but also polarizing films between any two adjacent LCD panels (P2 and P3 shown in Fig. 1). And $\{a, b, c\} = a * b * c$ in production field. In other words, production field needs the four films. While polarization field needs only P1 and P4 films. It should be noted that the production field can be transformed to addition field by logarithmic operation. Then $\log(\{a, b, c\}) = \log(a * b * c) = \log(a) + \log(b) + \log(c)$. In this paper, we adopt production filed.

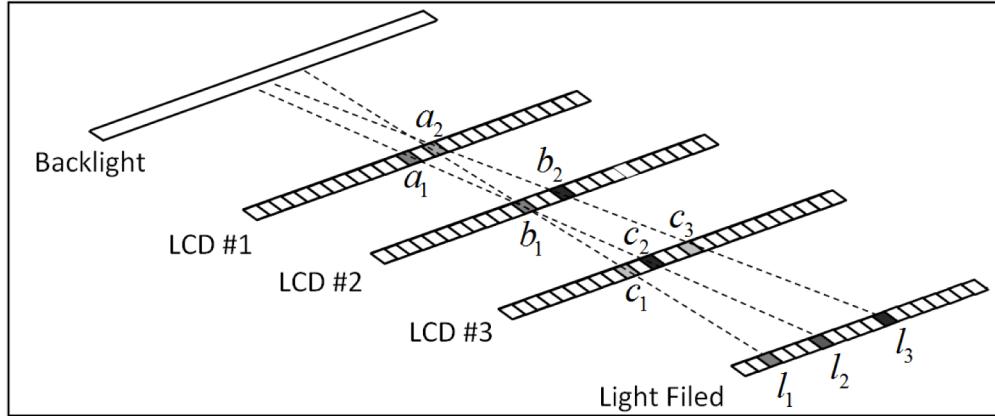


Fig. 4. Illustration of light rays attenuation in sub-dimension of light field.

Time-consuming decomposition prevents the light field display techniques from wide spread video-rate 3D display application. Cao accelerated decomposition of single light field frame by resolving the target light field [17]. On this basis, we further accelerate decomposition of light field video by exploring the internal coherence within single light field frame and the external coherence between adjacent light field frames. In the following sections, we discuss how to construct good initial values for decomposition and develop a pixel-parallel algorithm based on CUDA.

4. Coherence in light field decomposition

The high-dimensional light field decomposition problem can be cast as an over-determined equation set. According to the analysis in [17], the amount of light rays is generally much larger than the amount of controllable pixels on LCD panels. The multiple layer display system works in a state of serious over-load. In such case, there is not a unique nor exact analytical solution for this over-determined equation problem. Iterative algorithm provides an effective scheme to find a global optimal solution for this problem. But the quality of initial values greatly affects the convergence rate and reconstruction quality. Generally, iteration solver starts from random initial values because no knowledge about the target values is provided before iteration. Good initial values play a significant role in accelerating light field decomposition. Initial values which closer to the target result in faster convergence. In this section, we discuss how to get good initial values and accelerate decomposition of light field video by exploring the internal coherence within single light field and external coherence between adjacent light field frames.

4.1 Internal coherence within single light field frame

As mentioned above, the light field containing massive light rays requires huge computation cost. In theory, smaller light field size requires less computation time. We start light field decomposition in very small size using random noise as initial values. Considering the similarity between adjacent spatial light rays, we get better initial values for full-resolution light field through interpolation. We call this scheme as resolution-pyramid light field decomposition. Firstly, a light field resolution pyramid is constructed and the decomposition begins at the lowest resolution level, as shown in Fig. 5. For the lowest resolution level, we have to begin iteration with random initial values due to that we still know nothing about the target values. At the other levels, the initial values are figured out by interpolating the layer images at previous resolution level. By this way, we transfer the ignorance about the full-resolution light field to a very low resolution level.

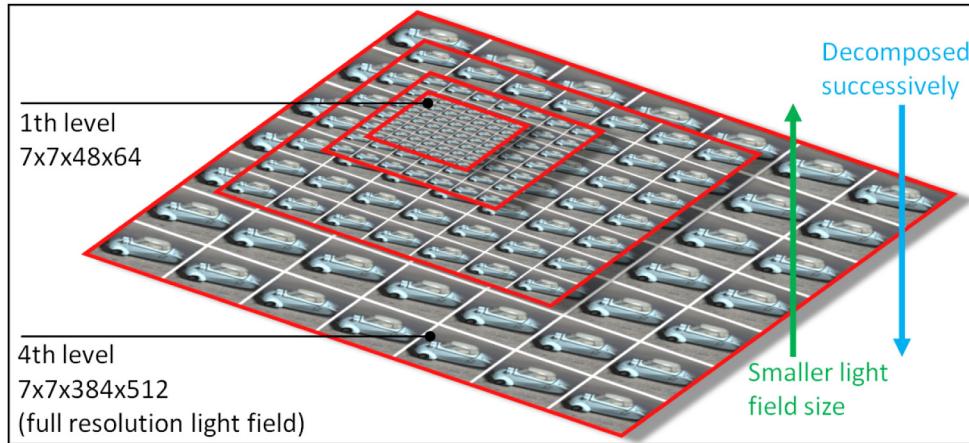


Fig. 5. Light field resolution pyramid for utilizing internal coherence.

How many levels should be constructed in the pyramid for achieving best performance? A low pyramid (e.g. 2 levels) results in that the resolution of top level is still very high. The ignorance about the target values fail to be resolved. A high pyramid (e.g. 10 levels) obtains a very small resolution at the top level and the ignorance about the target values is transferred to very small size. But more times of interpolations (e.g. 9) and decompositions (e.g. 10) are needed. We need to trade off between the resolution of top level and the number of levels. The basic rule of designing resolution pyramid is to make best of the GPU device(s). We decide the number of levels depends on the number of kernels in the GPU device(s). For example, the GPU device in our prototype is Nvidia Titan Black containing 15 multiprocessors and each multiprocessor has 192 CUDA cores. So there are totally 2880 kernel cores. The image resolution of light field is 384x512. The resolution of top level is 48x64 if the pyramid contains 4 levels. The number of light rays of the top level is 3072 which is close to the total kernel cores 2880. More levels (e.g. 5) or less levels (e.g. 3 levels) result in that light rays of top level is much smaller or larger than the number of GPU kernel cores.

Figures 6-8 illustrate light field decomposition at 1th, 2th and 4th resolution levels. As shown in Fig. 6, the initial values of 1th level are still random noise but in very small resolution. Excepting the 1th level, the initial values become meaningful and much closer to the convergence values, as shown in Fig. 7 and Fig. 8. Then only 2-3 iterations are needed to achieve satisfying layer images. The initial values of 2th level (in Fig. 7) are figured out by interpolating the layer images in 1th level (in Fig. 6). Similarly, as shown in Fig. 8, the initial values of 4th level (full-resolution level) are figured out by interpolating the layer images in 3th level. Here we apply simple and fast two-adjacent linear interpolation. Other interpolation algorithm remains a future research direction.

In short conclusion about resolution pyramid, each light field level is decomposed successively beginning with 1th level, as shown in Fig. 5. As to 1th level, the light field with size of $7 \times 7 \times 48 \times 64$ is decomposed into three layer images based on iteration solver, and the initial layer images are random cube noise of $3 \times 48 \times 64$. Then three 48×64 layer images are figured out, as shown in Fig. 6. For the 2th level, the light field with size of $7 \times 7 \times 96 \times 128$ is decomposed utilizing same iteration solver, but the initial layer images is not the random cube noise but interpolation of the three 48×64 layer images from 1th level. As such scheme, the layer images of full-resolution light filed can be figured out.

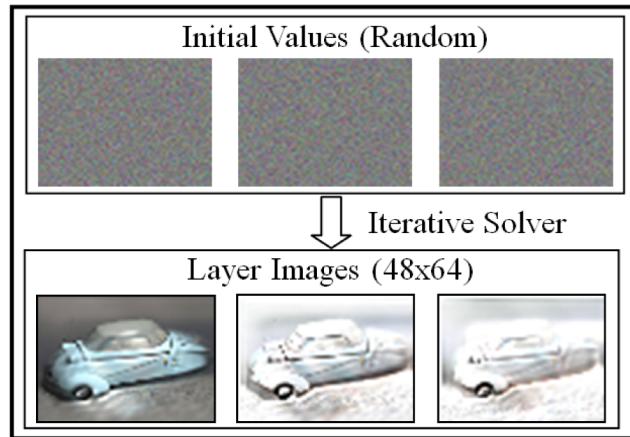


Fig. 6. Light field decomposition with random initial estimate at 1th pyramid level (7x7x48x64).

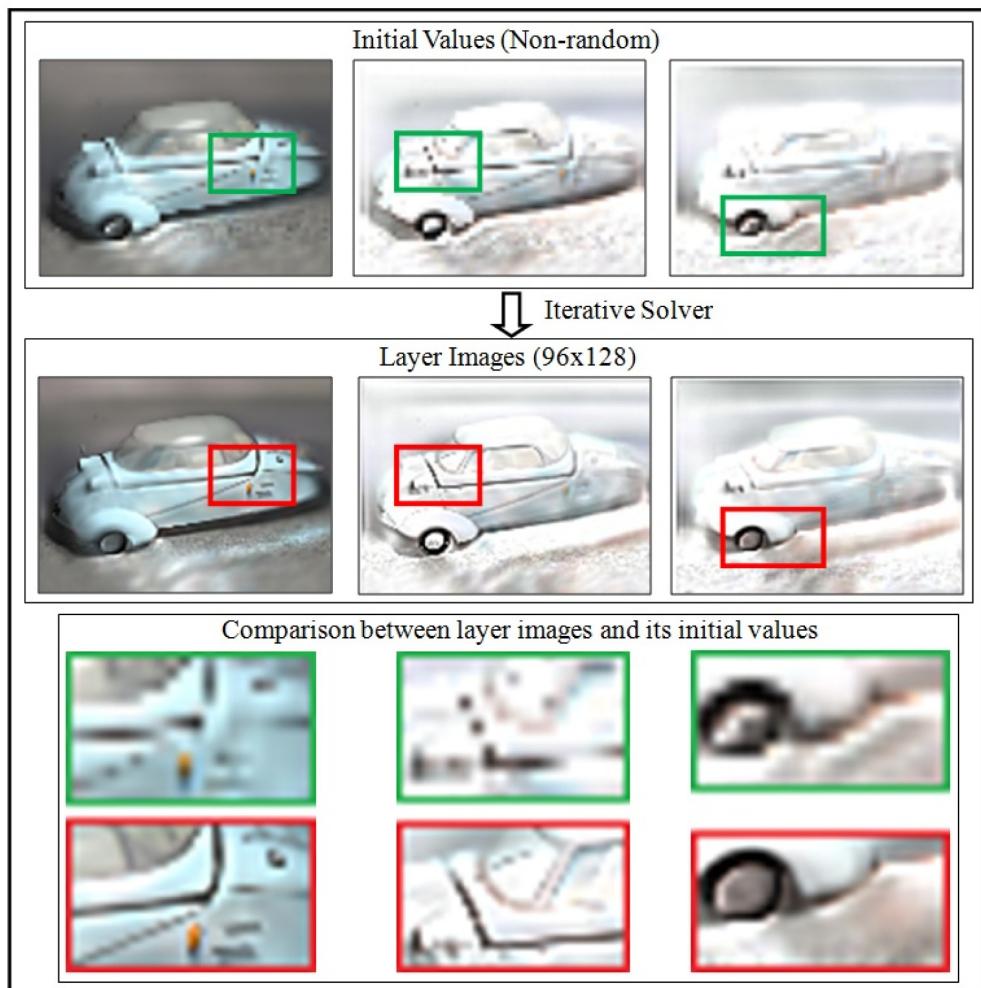


Fig. 7. Light field decomposition with optimal initialization at 2th pyramid level (7x7x96x128).

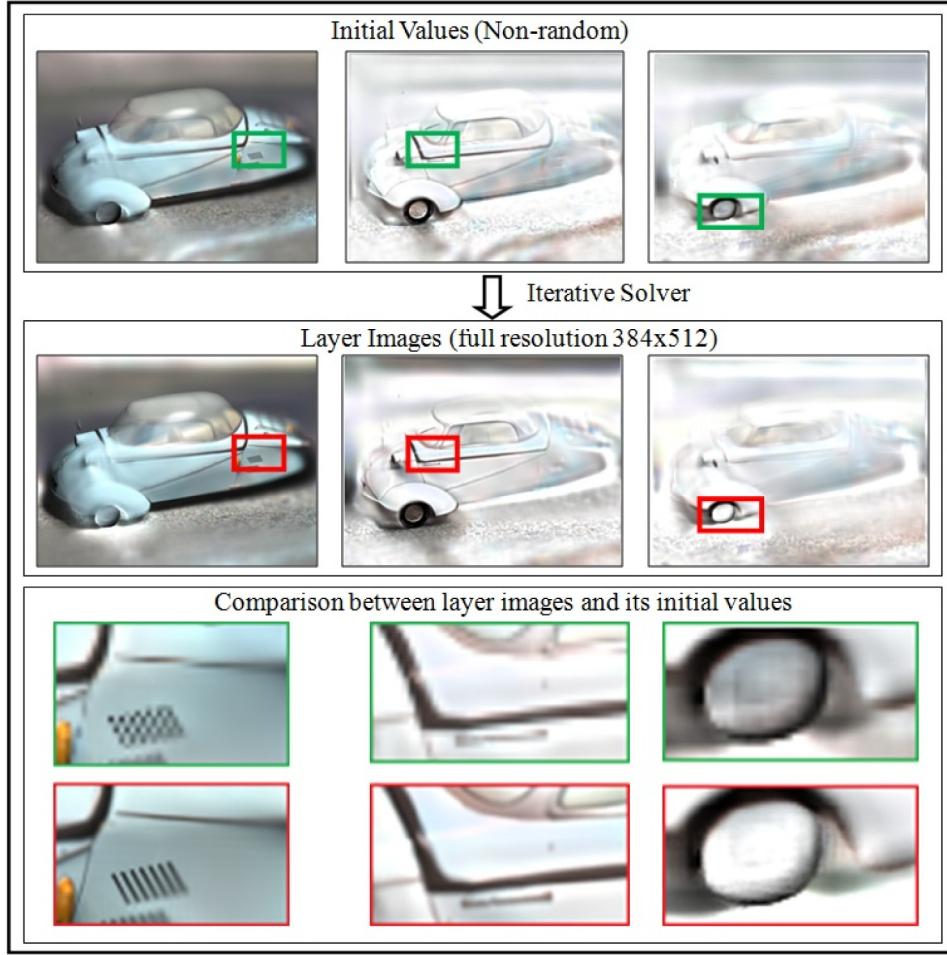


Fig. 8. Light field decomposition with optimal initialization at 4th pyramid level (7x7x384x512).

4.2 External coherence between adjacent light field frames

Beside the internal coherence between neighboring spatial light rays within single light field frame, we found that the time-adjacent light field frames are also coherent. In a light field video, current light field frame is similar with the previous light field frame. And the layer images of these two light field frames are also similar. The coherence between adjacent light field frames brings good and easy-to-obtain initial values for light field video decomposition. We use the layer images of previous light field frame as the initial values of current frame.

A light field video, similar with the traditional 2D video, has some frames containing big changed scenes. And these frames have little coherence with previous frame. We define these frames which are not coherent with previous frame as key-frames. Key-frames cannot utilize the layer images of previous frame as initial values due to their limited similarity. The initial values are figured out by utilizing internal coherence described in section 4.1.

To illustrate the decomposition of light field video clearly, we take an example as shown in Fig. 9. The main content is a bat waving wings, see details in Microsoft Direct X. We render the bat from 8x8 viewpoints, and each view image resolution is 384x512. As to the 1th light field frame, it has no previous frame providing initial values. So the 1th frame must be key-frame. And the first frame can only be decomposed utilizing resolution pyramid scheme. Starting from the 2th frame, the external coherence contributes to accelerating video decomposition, because

the layer images of first frame provide effective initial values. Accordingly, the n th frame utilizes layer images of $(n-1)$ th frame as initial values. The $(n-1)$ th frame provides much closer and better initial layer images for n th frame than random noise.

In a light field video, the scene may change hugely or the main object may move very fast. Such as the 57th frame, the bat draws back wings fast. If we decompose 57th frame based on the layers images of 56th frame, the iteration solver may not converge or achieve convergence very slowly. So the 57th frame is marked as another key-frame and decomposed like the 1th frame.

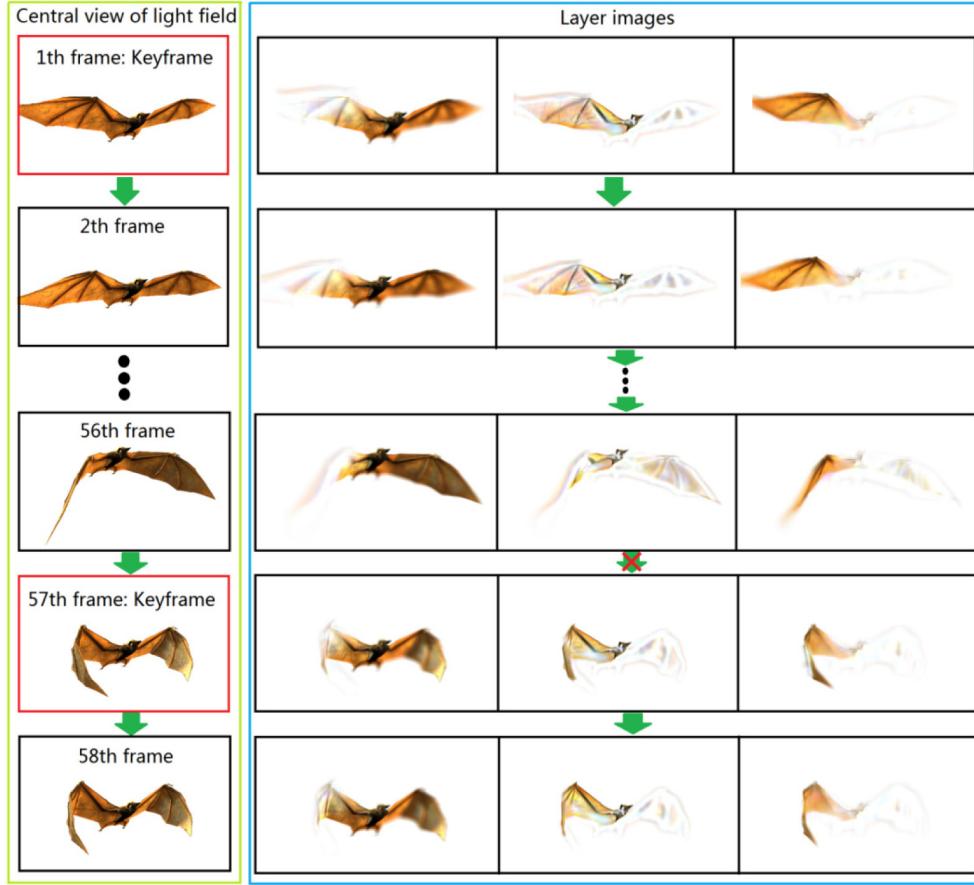


Fig. 9. Decomposition of light field video utilizing both internal and external coherence.

4.3 Assessment of computation cost

Our experiments show that the computation cost increases approximately linearly with the growing of light field size, as expressed in Eq. (2).

$$C \approx s \times I \times (v_x \times v_y \times r_x \times r_y) \quad (2)$$

where C is the computation cost for decomposing single light field frame; I denotes the number of iteration; Close-to-target initial values require much less number of iterations than that of random initial values. v_x , v_y , r_x and r_y have same definition with that in section 3; s is a scale factor representing computation cost consumed on single light ray in average.

We test our algorithm on five light field data from “camera culture research group” in media lab, MIT, including ‘Dice’, ‘Dragon’, ‘Happy_buddha’, ‘Messerschmitt’ and ‘Red_dragon’. All these light fields have same size including 7x7 viewpoints and image resolution of 384x512.

The resolution pyramid contains 4 levels as shown in the Table 1. The test is finished at Titan Black GPU card controlled by single host computer equipped with Xeon(R) W3565 and 14G RAM. The computation time and quality are shown in Table 1.

Table 1. Accelerating single light field frame by utilizing internal coherence.

	Res	LF	Dice	Dragon	Happy buddha	Messerschmitt	Red dragon	Average
Random initial values	384*512	Iteration	12	7	6	8	7	8
		Time(s)	2.683	1.567	1.343	1.788	1.574	1.791
		PSNR(db)	25.861	20.615	20.792	26.097	21.063	22.886
Resolution pyramid utilizing internal coherence	48*64	Iteration	3	3	3	3	3	3
		Time(s)	0.0523	0.0529	0.0515	0.0533	0.0521	0.0524
	96*128	Iteration	3	3	3	3	3	3
		Time(s)	0.0641	0.0647	0.0635	0.0607	0.0639	0.0634
	192*256	Iteration	3	3	3	3	3	3
		Time(s)	0.2079	0.2083	0.2058	0.2034	0.2064	0.2064
	384*512	Iteration	3	3	3	3	3	3
		Time(s)	0.6788	0.6698	0.6681	0.6677	0.6704	0.6710
	Total time		1.0031	0.9957	0.9889	0.9851	0.9928	0.9931
	PSNR(db)		25.868	20.624	20.802	26.105	21.041	22.888

As shown in Table 1, decomposition using random noise as initial values needs more number of iterations (8 in average) to reach satisfying quality. Our algorithm requires less number of iterations (3 in average) by utilizing the internal coherence. Although other resolution levels consume some computation time for decomposition and interpolation, the total time consumed in all resolution level is 44% less than that of random initial values.

Similarly, the external coherence makes much greater contribution to accelerating light field video decomposition. As shown in Fig. 10, a 24fps light field video with 8x8 viewpoints and resolution of 384x512, containing 480 frames, is decomposed separately by utilizing random initialization and external coherence. The red dotted line represents cost time of each frame's decomposition by utilizing the random initialization. The red full line represents the cost time of each frame's decomposition by utilizing the external coherence. At some key-frames, like the peaks of the red full line, the computation time increases due to deficiency of external coherence. Although the resolution pyramid scheme at key-frame takes more cost time, the total cost time consumed on decomposition of video is much less than that of iteration utilizing random initialization. It should be noted that the decomposition quality (PSNR in dB) depends on the content of light field. As the blue line shown in Fig. 10, some frames containing simple scene (e.g. spreading wings) achieve a little higher PSNR than frames containing complicated scene (e.g. huddled wings).

Our algorithm, iterating from previous frame, takes 2 iterations for each frame and achieves PSNR of 24.412 dB in average. Iteration solver beginning with random initialization takes 12 iterations to achieve PSNR with 24.399 dB in average. Iteration with random initialization achieves close PSNR (a little lower exactly) to that of ours at price of much larger computation time cost. Our computation time on decomposition of the 20s video is 4.705 minutes which is much less than the 27.82 minutes of random initialization, speeding up decomposition 5.91 times.

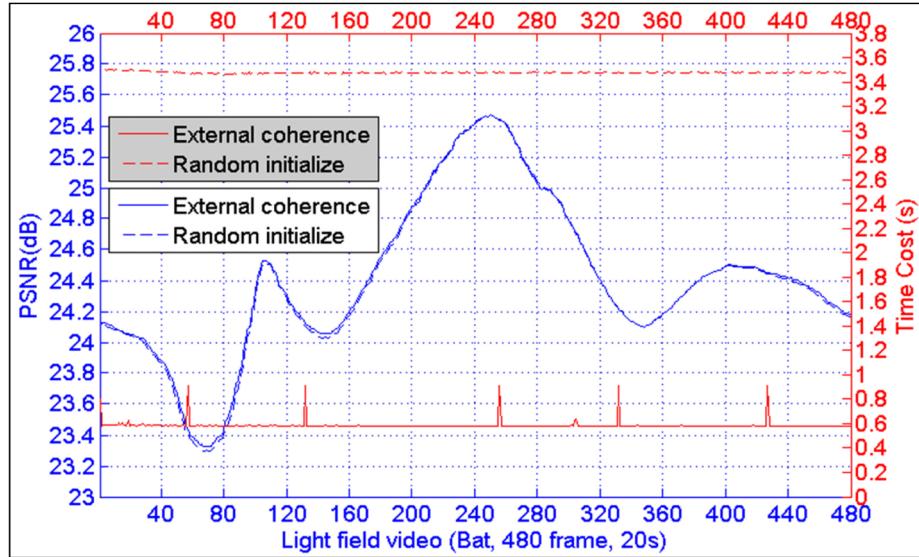


Fig. 10. Accelerating decomposition of light field video without quality deterioration.

4.4 Decomposition in parallel

CUDA developed by Nvidia extended display card to general parallel computation. In this section, we developed a pixel-parallel iteration algorithm to accelerate light field decomposition based on the CUDA technology.

Our pixel-parallel algorithm updates one layer at one time and the rest layers keep fixed before finishing the current layer update. All pixels in one layer are updated in parallel. As shown in Fig. 4, the pixel a_i is responsible to modulate multiple light rays. And the optimization purpose is to find a “best” value for pixel a_i forcing all light rays passed through pixel a_i to be close to the target light rays. The iterative update rules follow the Cao’s works [17]. Here, we repeat update rule of LCD #1 as expressed in Eq. (3). The update rule of LCD #2 and #3 are similar to that of LCD #1.

$$a' = a * \frac{\sum_{j=1}^V l_j * b_j * c_j}{\sum_{j=1}^V \hat{l}_j * b_j * c_j}, \text{ s.t. } 0 \leq a' \leq 1 \quad (3)$$

where a is old pixel value on LCD #1 and a' is updated pixel value; b_j and c_j are intersection pixels on LCD #2 and LCD #3 along j -th target light ray; V is the amount of views in target light field ($V = v_x * v_y$); \hat{l}_j is the j -th reconstructed light ray ($\hat{l}_j = a * b_j * c_j$); l_j is the j -th target light ray.

Advanced Nvidia Graphic Card (e.g. Titan Black in our system) supports up to 4-float vector operation. We vectorize the pixel data of three color channels and independently solve three color channels in parallel. To sum up, the SIMD (Same Instruction for Multiple Data) is applied in pixel-level and color channel-level. So our algorithm decomposes light field in pixel-parallel and color channel-parallel. The pseudo-code is shown in Table 2.

Table 2. The algorithm flow chart for light field decomposition in parallel.

if (true == keyframe _(i)) // Index “i” represents <i>i</i> -th light field frame	
{ // keyframe should utilize internal coherence	
For ithLevel = 1:1: (NumLevel-1) // construct a light field resolution pyramid	
{	
if (1 == ithLevel) // the top level must iterates from random initial values	
{	
A _(i, ithLevel) = random([0, 1]) ; // Set all pixels in all LCD layers	
B _(i, ithLevel) = random([0, 1]) ; // as random noise ranging from 0 to 1;	
C _(i, ithLevel) = random([0, 1]) ;	
}	
For ith = 1:1:NumIteration	
{ // Update pixels in three layers according to E.q. (2)	
A' _(i, ithLevel) = Update(A _(i, ithLevel)) ; A _(i, ithLevel) = A' _(i, ithLevel) ;	
B' _(i, ithLevel) = Update(B _(i, ithLevel)) ; B _(i, ithLevel) = B' _(i, ithLevel) ;	
C' _(i, ithLevel) = Update(C _(i, ithLevel)) ; C _(i, ithLevel) = C' _(i, ithLevel) ;	
}	
// get initial values of next-level by interpolating current layer images	
A _(i, ithLevel + 1) = Interpolation(A' _(i, ithLevel)) ;	
B _(i, ithLevel + 1) = Interpolation(B' _(i, ithLevel)) ;	
C _(i, ithLevel + 1) = Interpolation(C' _(i, ithLevel)) ;	
}	
A _(i) = A _(i, ithLevel + 1) ; // Set interpolation of penultimate-level layer images	
B _(i) = B _(i, ithLevel + 1) ; // as initial values of full-resolution level	
C _(i) = C _(i, ithLevel + 1) ;	
}	
else // if it's not key-frame, we get the initial values from previous frame directly	
{	
A _(i) = A _(i-1) ; // Set all LCD layers equal to layers image of previous frame	
B _(i) = B _(i-1) ;	
C _(i) = C _(i-1) ;	
}	
// Code above figured out the initial layer images of full-resolution light field by internal	
// or external coherence. Following code finish iteration beginning with such initialization.	
For ith = 1:1:NumIteration	
{	
For layer = {A _(i) , B _(i) , C _(i) } //All pixels in single layer are processed synchronously	
{ // One GPU kernel is responsible for one pixel in current LCD layer	
Sum = 0.0f;	
Sum' = 0.0f;	
For j = 1:1:V // V is the total amount of views	
{	
l _j = a*b _j *c _j ; // record the reconstruction of <i>j</i> -th light ray	
Sum = Sum + l _j * b _j *c _j ; // l _j is the <i>j</i> -th target light ray	
Sum' = Sum' + l' _j * b _j *c _j ;	
}	
a' = a*(Sum/ Sum') ; // update the pixel a	
a' = min(0, max(a', 1)) ; // restricted pixel a' ranging [0, 1]	
syncthreads() ; // wait for finishing all kernels	
}	
cudaDeviceSynchronize() ; // wait for finishing updating current layer	
}	

4.5 Analysis of iteration with optimal initial estimate

Light field video is a set of serial frames. Compressive light field video display is implemented based on decomposition of frame by frame. Our experiments have demonstrated that, no matter keyframe or non-keyframe, better initial estimates reduce computation time significantly. Here, we further analyze the performance of video decomposition by observing the iteration procedure at keyframe and non-keyframe.

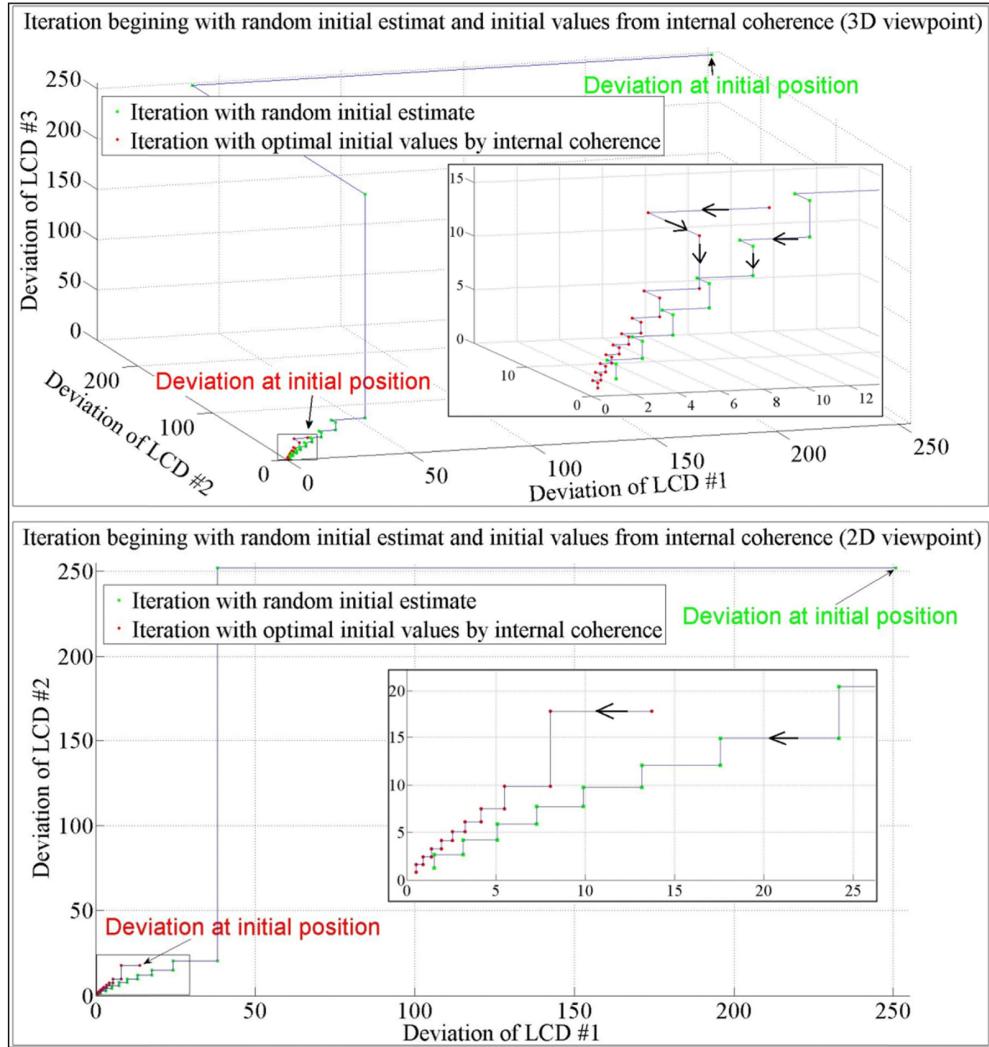


Fig. 11. Iteration with random estimate and initialization by internal coherence at keyframe (57th frame of “bat”).

The deviation of LCD image is quantified by Euclidean distance (error of L2 norm), as equation $\sqrt{\sum_{k=1}^N (P_{1k} - P_{2k})^2}$, where N is the total number of pixels on LCD panel, P_1 and P_2 denote pixels of current layer and target layer images accordingly. At key frame, as shown in

Fig. 11, we compare the iteration procedure beginning with random initial estimate and optimal initialization based on internal coherence. The three axes indicate the deviation of layer images of LCD #1, #2, #3. As shown in Fig. 11, the deviation of optimal initial values ($<13.7101, 17.7352, 11.6018>$) is much less than that of random initial estimate ($<250.7239, 251.9857, 248.9224>$). In other word, the optimal initial position is much closer to target than random initial position. From the 2D viewpoint (LCD #1 and LCD #2), we can see the deviation decrease along with iteration. To illustrate the details of iteration, the magnified regions are shown inside of diagram.

By same definition of coordinate system, the iteration procedure at non-keyframe is shown in Fig. 12. The deviation of optimal initialization based on external coherence is $<31.7313,$

$29.5772, 27.2280$. The deviation of random initial estimate is much larger up to $<250.3571, 252.1252, 249.0645>$. We also depict the decreasing of deviation from a 2D viewpoint (LCD #2 and LCD #3).

In conclusion, the internal and external coherence in light field video assign optimal starting position for iteration solver. Although the resolution-pyramid strategy and copies of previous frame consume additional time, the extra time is much less than that of additional iteration caused by random initial estimate.

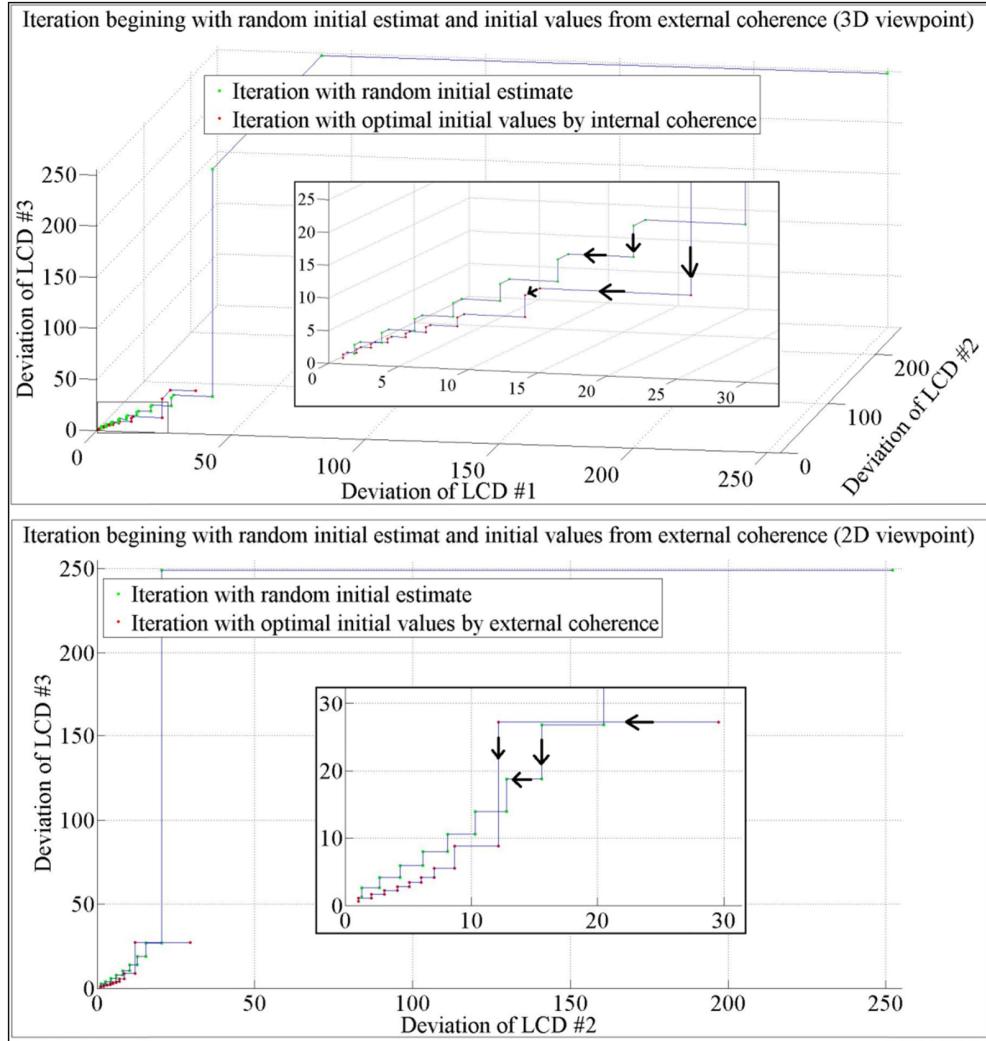


Fig. 12. Iteration with random estimate and initialization by external coherence at non-keyframe (58th frame of “bat”).

5. Implementation

We build a multi-layer light field display prototype, as shown in Figs. 13 and 14. The model of LCD panel is ASUS VG 278H which has resolution of 1920*1080 and refresh rate up to 144Hz. Two polarizing film are placed between adjacent LCD panels. Another two polarizing films are mounted outside of two LCD panels. The polarity of these four films keeps orthogonal with adjacent films. The three LCD panels are driven by DirectX and display layer images

synchronously. The host computer is Xeon(R) W3565 equipped with 14G RAM and the graphic card is Nvidia Titan Black with CUDA capability 3.5.

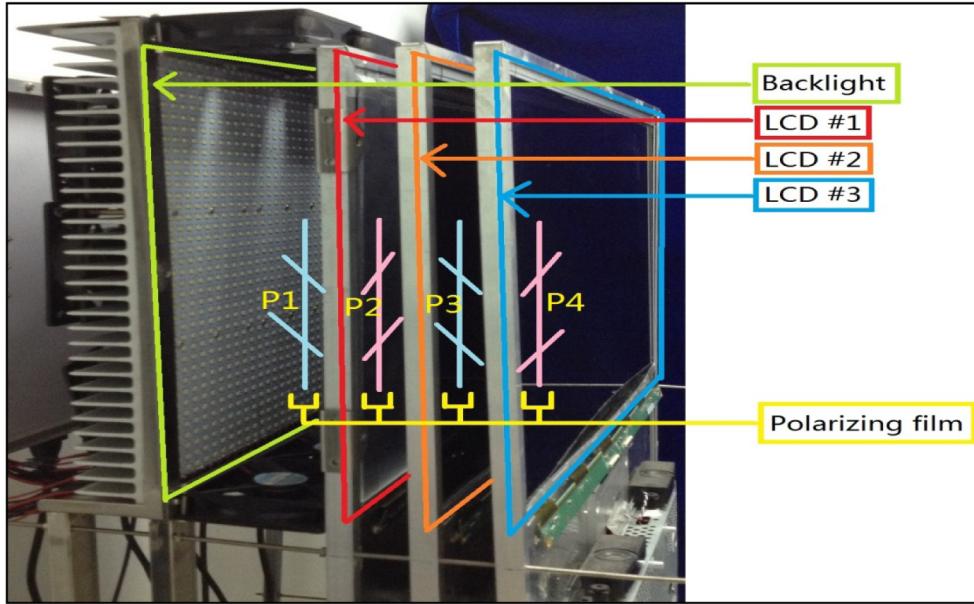


Fig. 13. Prototype of three layers compressive light field video display.



Fig. 14. Photograph of prototype.

6. Conclusion

Light field display, as a glass-free 3D display technology, is improving the comfort level of watching 3D media. Currently, several schemes have been developed for light field display. But there exist no commercial one can display live light field video. In this paper, we focus on accelerating light field video display based on multiple layers display. By utilizing internal and external coherence of light field, we speed up light field decomposition 5.91 times. We also build a prototype and develop a parallel algorithm based on CUDA technology. We hope the techniques discussed in this paper provide novel and useful insights for future live light field video display.

Acknowledgments

This work has been supported by the National High-tech R&D 863 Program of Institute of Automation, Chinese Academy of Sciences, grant 2012AA011903 and Peking University, grant 2015AA015905.