REAL-TIME FPGA IMPLEMENTATION OF LINEAR BLENDING VISION RECONSTRUCTION ALGORITHM USING A SPHERICAL LIGHT FIELD CAMERA

Hossein Afshari, Abdulkadir Akin, Vladan Popovic, Alexandre Schmid and Yusuf Leblebici

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
first_name.last_name@epfl.ch

ABSTRACT

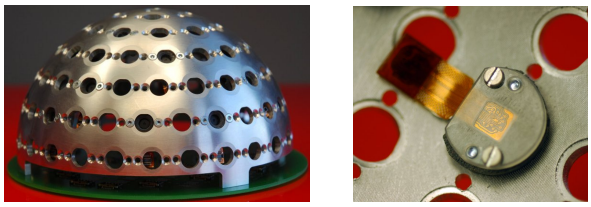
A custom spherical light-field camera used as a polydioptric system where imagers are distributed over a spherical geometry, each having its own vision of the surrounding and distinct focal plane. The spherical light-field camera is also an omnidirectional camera which records light information from any direction around its center. A novel linear blending technique is presented for vision reconstruction of a virtual observer located inside the spherical geometry of this camera. This blending technique improves the output quality of the reconstructed vision with respect to the ordinary stitching technique. A novel pixel gridding scheme is presented for rectangular displaying of the reconstructed vision induced from the spherical light field camera. This gridding technique preserve the correct size of objects when mapped on the spherical geometry of the Panoptic system. A hardware architecture based on FPGAs with the real-time implementation of the linear blending algorithm and the new pixel gridding scheme of the spherical light-field camera are presented along with imaging results.

Index Terms— Light Field Camera, Real-time, FPGA

1. INTRODUCTION

The current trend in constructing high-end computing systems consists of parallelizing large numbers of processors. A similar trend is observed in digital imaging where multiple camera inputs are utilized to obtain multiple images of a scene for enhancing the performance envelope of the image capture process. Increased image resolution [1] and dynamic range [2, 3] are examples of such enhancements. View interpolation for creating the illusion of a smoothly moving virtual camera in a scene is another use of multi-view systems [4].

Early systems for capturing multiple views were based on a single translating camera [5]. This idea was later extended to a dynamic scene by using a linear array of still cameras [6]. For capturing large data sets, researchers focused on arrays of video cameras. In addition to the synchronization of the cameras, enormous data rates present new challenges for the implementation of these systems. Primary camera array systems were built for recording, while later offline processing on PCs [4]. Other such systems [7], [8] were built with real-time processing capability for low resolution and low frame rates but still lacked synchronization. A general purpose camera array system for research was built at Stanford [9] with limited local processing at the camera level; nevertheless this system was aimed for recording large data and intensive offline processing as well. The developed camera array systems are bulky and not easy portable platforms; their control and operation depends on multi-computer setups; they are primarily intended for recording multiple views and not real-time embedded processing. In addition, image sensors on camera arrays are mounted on planar surfaces which prohibits them from covering the full view of their surrounding environment. Full view or panoramic imaging finds application in various areas such as navigation and robots, telepresence, remote monitoring and human motion. There are number of ways for acquiring omnidirectional images and applying them practically [10].



(a) (b)

Fig. 1. (a) Side view, (b) and internal view of the fabricated Panoptic camera with a hemispherical diameter of $2\,r_{\odot} = 129\mathrm{mm}$ and 104 camera positions.

A new approach for creating a multi-camera system distributed over a spherical geometry is presented [11, 12]. This new multi-camera system is referred to as the Panoptic camera. The Panoptic camera is an omnidirectional imager capable of recording light information from any direction around its center. It is also a polydioptric system where each CMOS facet has a distinct focal plane enabling the attribute of a

This research has been partly conducted with the support of the Swiss NSF under grant number 200021-125651. The authors gratefully acknowledge the support of XILINX, Inc., through the XILINX University Program.

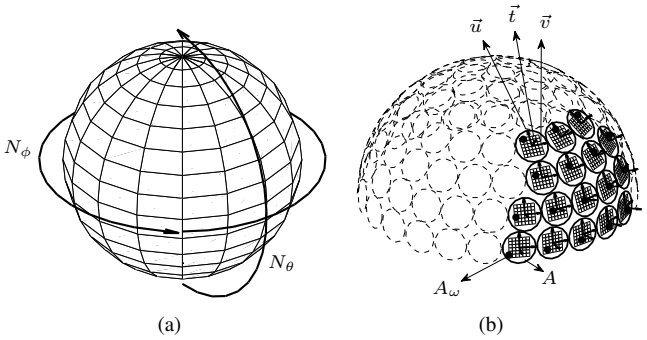


Fig. 2. (a) Pixelized sphere surface $\mathcal{S}_d$ with $N_\theta = 16$ latitude pixels and $N_\phi = 16$ longitude pixels (total of 256 pixels), (b) cameras contributing to pixel position ω along with their contributing positions on their respective image frames

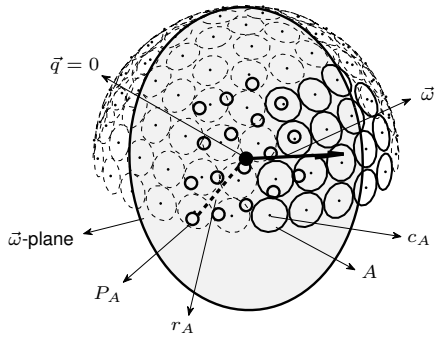


Fig. 3. Projections of camera centers contributing in direction $\vec{\omega}$ onto planar surface normal to $\vec{\omega}$.

multi-aperture camera. Fig. 1(a) depicts a custom made Panoptic camera. The internal view of this camera is shown in Fig. 1(b).

The omnidirectional vision reconstruction algorithm using the nearest neighbour technique and the coverage analysis of the Panoptic camera is discussed in detail in [11]. A novel linear interpolation technique which improves the output quality of the reconstructed image is presented in Section 2.1. A novel pixel gridding scheme for the reconstructed vision is presented in Section 2.2. This gridding scheme preserves the correct projection size of objects when mapped on sphere surface. Both techniques are developed in real-time for the presented hardaware platform in [11]. Imaging results and comparisons are presented in Section 4.

2. OMNIVISION RECONSTRUCTION ALGORITHM

The omnidirectional vision of a virtual observer located anywhere inside the hemisphere of the Panoptic structure can be reconstructed by combining the information collected by each camera. The method utilized for the reconstruction is the interpolation of light information in the light ray space domain (or light field [5]).

In this process, the omnidirectional view on a discretized sphere surface $\mathcal{S}_d$ of directions is estimated. The surface of this sphere is discretized into an equiangular grid with N_θ latitudes and N_ϕ longitudes pixels. Fig. 2(a) shows a pixelized sphere surface with sixteen pixels for N_θ and N_ϕ each. The direction of each pixel is identified by the unit vector $\vec{\omega} \in \mathcal{S}_d$ and relates to the spherical coordinates by $\omega = (\theta_\omega, \phi_\omega)$.

The construction of the virtual omnidirectional view $\mathcal{L}(\vec{q}, \vec{\omega}) \in \mathbb{R}$, where $\vec{q}$ is the observation point, is performed in two algorithmic steps. First, all the cameras having $\vec{\omega}$ in their angle-of-view are determined. Since $\vec{\omega}$ does not coincide with the exact pixel grid locations on the camera image frames, a first level of interpolation is conducted to extract the light intensity in that direction for each contributing camera. As a result, the first algorithmic step estimates the values $\mathcal{L}(c_i, \vec{\omega})$, where c_i is the radial vector directing at the center position of the i^{th} contributing camera's circular face. Fig. 2(b) shows an example of contributing cameras for a typical pixel direction $\vec{\omega}$ depicted in Fig. 3. The contributing position A_w of the camera A, providing $\mathcal{L}(c_A, \vec{\omega})$ is indicated in Fig. 2(b).

The second algorithmic step is performed in the space of light rays given by direction $\vec{\omega}$ and passing through the camera center positions. Under the assumption of constant light flux (CLF), the light intensity remains constant on the trajectory of any light ray. Following the CLF assumption, an orthographic plane is defined such that for a given direction $\vec{\omega}$, light ray intensity only varies in its respective orthographic plane. The orthographic plane is normal to $\vec{\omega}$. The orthographic plane pertaining to the direction $\vec{\omega}$ is indicated as the "$\vec{\omega}$-plane" in Fig. 3, and represented as gray-shaded circle (the boundary of the circle is drawn for clarity purposes). The light rays of direction $\vec{\omega}$ recorded by each contributing camera intersect the $\vec{\omega}$-plane in points that are the projections of the cameras focal points on this plane. The projected points of the contributing camera positions in $\vec{\omega}$ onto the $\vec{\omega}$-plane are highlighted by hollow points in Fig. 3. Each projected camera position P_{c_i} on the planar surface is assigned the intensity value $\mathcal{L}(c_i, \vec{\omega})$. As an example, the projected camera focal point of camera A onto the $\vec{\omega}$-plane (*i.e.*, P_A) in Fig. 3 is assigned the intensity value I_A. The virtual observer point inside the hemisphere (i.e. $\vec{q}$) is also projected on to the $\vec{\omega}$-plane. The light intensity value of the projected observer point (*i.e.*, $\mathcal{L}(\vec{q}, \vec{\omega})$) is estimated through the algorithmic aggregate of other intensity values, or a subset of them to extract one unique intensity value.

For example, each of the seventeen contributing camera positions shown with bold perimeter in Fig. 3 provides an intensity value which is observed into direction $\vec{\omega}$ for observer position $\vec{q} = 0$, which is the center of the sphere and indicated by a bold dot. A single intensity value is resolved among the contributing intensities through a two-dimensional interpolation on its respective $\vec{\omega}$-plane.

2.1. Linear Interpolation

In the nearest neighbour technique and considering the second algorithmic step, the light intensity of the virtual observer point per $\vec{\omega}$ direction is set as the light intensity of the best observing camera for each $\vec{\omega}$ direction. The nearest neighbour technique is expressed in (1) in mathematical terms:

$$\begin{aligned} j &= \text{argmin}_{i\in\mathcal{I}}(r_i) \\ \mathcal{L}(\vec{q},\vec{\omega}) &= \mathcal{L}(c_j,\vec{\omega}). \end{aligned} \tag{1}$$

Where $\mathcal{I} = \{i|\vec{\omega}\cdot\vec{t_i} \geq \cos(\frac{\alpha_i}{2})\}$ is the index set of contributing cameras for a pixel direction $\vec{\omega}$. A pixel direction $\vec{\omega}$ is assumed observable by camera c_i if the angle between its focal vector $\vec{t_i}$ and that of the pixel direction $\vec{\omega}$ is smaller then the half of the angle of view α_i of camera c_i [11].

The length r_i identifies the distance between the projected focal point of camera c_i and the projected virtual observer point on the $\vec{\omega}$-plane. The camera with the smallest r distance to the virtual observer projected point on the $\vec{\omega}$-plane is considered the best observing camera. For example, this distance is identified with r_A and depicted by a dashed line for the contributing camera A in Fig. 3.

The linear interpolation scheme incorporates all the contributing cameras intensity values through a linear combination. This is conducted by aggregating the weighted intensities of the contributing cameras. The weight of a contributing camera is the inverse function of the distance between its projected focal point and projected virtual observer point on the $\vec{\omega}$-plane (*i.e.*, r). The weights are also normalized to the sum of the inverse of all the contributing cameras distances.

The linear interpolation is expressed in (2) in mathematical terms.

$$\mathcal{L}(\vec{q},\vec{\omega}) = \frac{\sum_{i\in\mathcal{I}} \frac{1}{r_i}\mathcal{L}(c_i,\vec{\omega})}{\sum_{i\in\mathcal{I}} \frac{1}{r_i}}. \tag{2}$$

2.2. Constant Density Pixel Griding

The pixel directions $\vec{\omega}$ described in [11] and shown in Fig. 2(a) derive from an equi-angular segmentation of longitude and latitude coordinates of a unit sphere into N_ϕ and N_θ segments, respectively. This pixelization enables the rectangular presentation of the reconstructed image suitable for ordinary displays but results in a non-equal contribution for the cameras of the Panoptic system. The density of the pixel directions close to the poles of the sphere are higher compared to the equator of the sphere in the equi-angular pixelization scheme. Hence, cameras positioned closer to the poles of the spherical geometry contribute more pixels in comparison to other cameras of the system. The equi-angular pixel gridding derives mathematically from (3).

$$\begin{aligned} \phi_\omega(i) &= \tfrac{2\pi}{N_\phi} \times i, \quad 0 \leq i < N_\phi \\ \theta_\omega(j) &= \tfrac{\pi}{N_\theta} \times (j+\tfrac{1}{2}), \quad 0 \leq j < N_\theta. \end{aligned} \tag{3}$$

A constant density pixel gridding scheme resulting in an approximately even contribution of the cameras is devised for the Panoptic system by enforcing a constant number of pixel to area density ratio in both longitude and latitude coordinates of a unit sphere (4).

$$\begin{aligned} \frac{N_\theta \times i}{\int_0^{\phi_\omega(i)}\int_0^{\pi} \sin\theta \, d\theta \, d\phi} &= \frac{N_\phi \times N_\theta}{4\pi}, \\ \frac{N_\phi \times j}{\int_0^{2\pi}\int_0^{\theta_\omega(j)} \sin\theta \, d\theta \, d\phi} &= \frac{N_\phi \times N_\theta}{4\pi} \end{aligned} \tag{4}$$

The following pixel gridding scheme is derived by solving the two identities in (4).

$$\begin{aligned} \phi_\omega(i) &= \tfrac{2\pi}{N_\phi} \times i, \quad 0 \leq i < N_\phi \\ \theta_\omega(j) &= \arccos(1 - \tfrac{2j}{N_\theta}) + \theta_0, \quad 0 \leq j < N_\theta. \end{aligned} \tag{5}$$

The offset value θ_0 is added to the latitude gridding in (5) to avoid repetition of pixel direction for the $j = 0$ case.

3. FPGA IMPLEMENTATION

A custom FPGA board has been designed utilizing a XILINX Virtex5 XC5VLX50-1FF1153C FPGA in order to capture and process the video streams produced by the cameras in real-time. This board contains twenty camera interfaces. To support higher number of camera interfaces, multiple identical boards of the same kind are stacked. For scalability and extension purposes, the designed board also contains high speed LVDS serial links and extension connectors. The board is also equipped with a USB 2.0 device chipset for external access and high-speed data transfer. The FPGA board also contains two zero bus turn around (ZBT) SRAMs with 36 Mb capacity and an operating bandwidth of 166 MHz for each.

A scalable FPGA-based system is devised, using the designed FPGA board, to support the application development of the Panoptic camera. The devised system consists of four layers, i) layer A: imagers with programmable resolution, ii) layer B: concentrator FPGA boards, each handling local image processing over 20 imagers, iii) layer C: one central FPGA board for control, external access and last stage image processing, iv) layer D: A PC in charge of the applicative layer consisting of displaying the operation results transmitted from the central FPGA board. Fig.4(a) depicts the devised architecture for a typical Panoptic system.

3.1. Concentrator FPGA Architecture

The architecture of the FPGA is depicted in Fig. 4(b). The concentrator FPGA consists of five major blocks. The arrow lines depicted in Fig. 4(b) shows the flow of image data inside the concentrator FPGA. Image data streaming from the cameras enters the concentrator FPGA via the Camera input channel block. A time-multiplexing mechanism is implemented to store the incoming frame data from all the camera

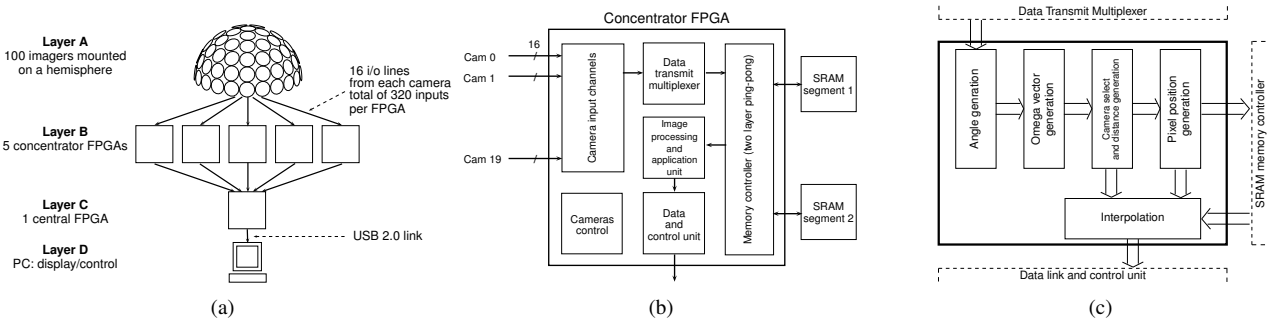


Fig. 4. (a) Architecture of the Panoptic system development platform, (b) architecture of a concentrator FPGA, (c) block diagram of the omnidirectional vision reconstruction unit inside the Image processing and application block.

modules into one of the single data port SRAMs. Hence the Data transmit multiplexer block time multiplexes the data received by the Camera input channel block and transfers it to the Memory controller block for storage in one of the SRAMs. The SRAMs are partitioned into twenty equal segments. The Memory controller block interfaces with two external SRAMs available on the board. The Memory controller block provides access for storing/retrieving the incoming/previous twenty image-frame data on the SRAMs. The SRAMs swap their role (*i.e.*, one being written to, and one being read) with the arrival of each new image frame from the cameras. The Image processing and application unit block is in charge of signal processing and basic functionalities such as single/dual video channel streaming, all channel capture and omnidirectional vision construction. The Image processing and application unit block accesses the SRAMs via the Memory controller block and transfers the processed/image data to the Data link and control unit block. The Data link and control unit block provides transmission capability over external interfaces available on the board such as high speed LVDS serial links or USB 2.0 link. The Cameras control block is in charge of programming and synchronizing the cameras connected to the FPGA board.

3.2. Hardware Implementation

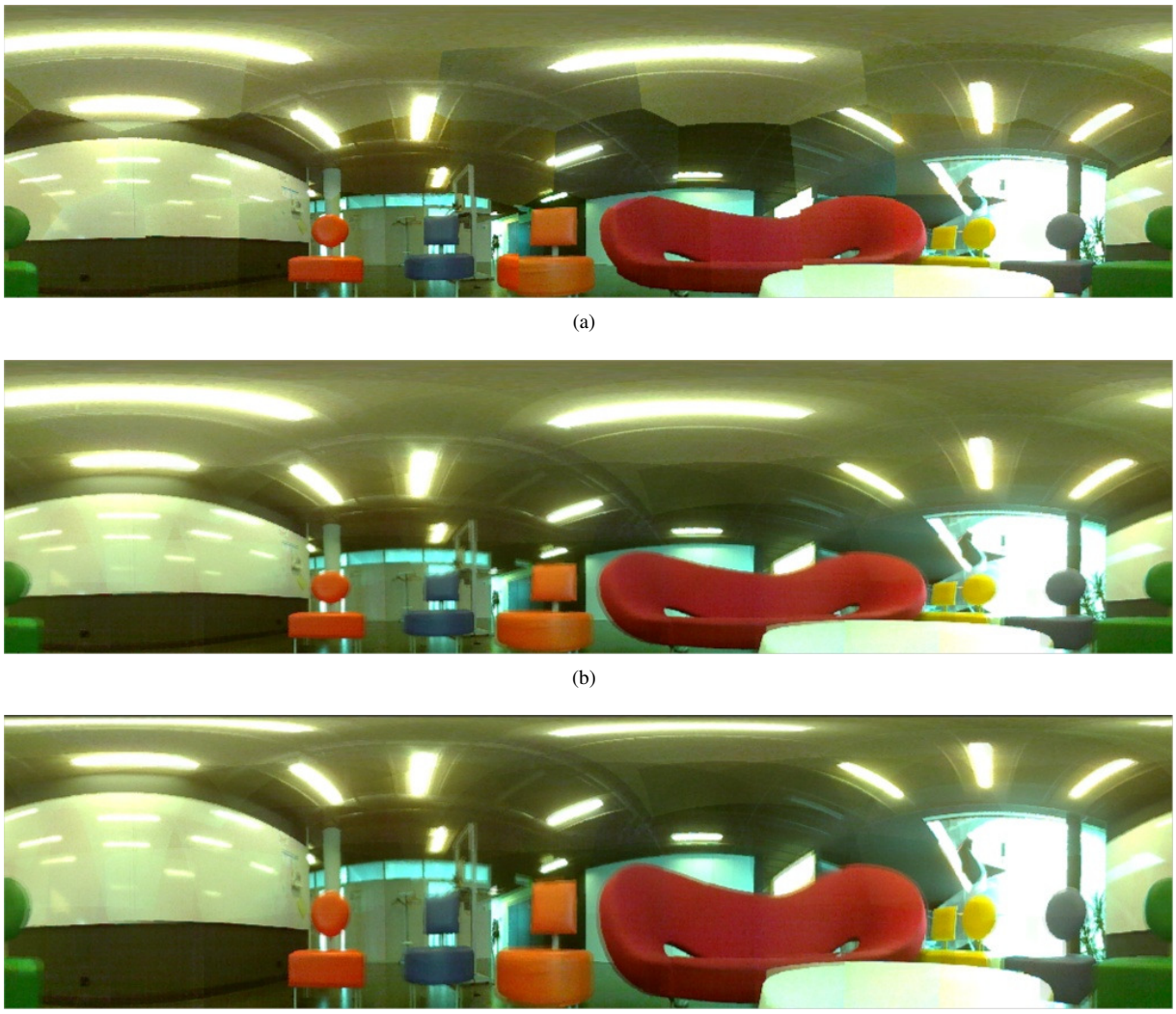
The omnidirectional vision reconstruction algorithm unit is implemented inside the image processing and application unit block. The block diagram of this unit is shown in Fig.4(c). This signal processing entity comprises five modules. The Angle generation module generates the spherical coordinates (*i.e.*, $(\theta_\omega, \phi_\omega)$), of the directions (*i.e.*, $\vec{\omega}$) of intrest for construction. The span and resolution of the output vision is selectable within this module. The equi-angular or constant density grid can be selected for the generation of spherical coordinates (*i.e.*, $(\theta_\omega, \phi_\omega)$). The maximum reconstruction resolution is 32M pixels. High resolutions are achieved trading off frame per seconds (*i.e.*, speed). The Omega vector generation module calculates the unit radial vector pertaining to the spherical positions $(\theta_\omega, \phi_\omega)$ received from the Angle generation module. The Camera select and distance generation module identifies which cameras contribute (*i.e.*, observe) to the construction of the $\vec{\omega}$. The index of the contributing cameras are passed to the Pixel position generation module. Concurrently, this module computes the weight (*i.e.*, coefficient) of the contribution for the participating cameras in direction $\vec{\omega}$. This weight is the reciprocal function of the distance between the projected focal point and the projected virtual observer point on the $\vec{\omega}$-plane for each contributing camera. The weights are used in the interpolation process of the second algorithmic step mentioned in Section 2. The Pixel position module retrieves the pixel light intensity value of the contributing cameras upon receiving their index and true pixel position from the Camera select and distance generation module. The index is used to access the correct segments where the image frame of a camera is stored and the true pixel position is used to access the target address within each memory segment inside the SRAMs. The Interpolation module receives the pixel light intensity values from all the contributing cameras along with their contributing weights and estimates a single light intensity value for each $\vec{\omega}$ direction.

3.3. Central FPGA and scalability

Each concentrator FPGA board can interface with twenty cameras. In order to support higher number of cameras and increase the output throughput of the Panoptic camera, multiple concentrator FPGA boards must be incorporated. Hence, the omnidirectional construction workload is distributed and operate in parallel on all concentrator FPGA boards. A central FPGA is required to receive the output data from the concentrator FPGA boards and transfer the final result to a PC for display.

4. RESULTS

The Panoptic system with thirty embedded cameras presented in [11] is used for real-time image extraction and evaluation.



(a)

(b)

(c)

Fig. 5. A seating place in Swiss Federal Institute of Technology in Lausanne (EPFL, BC01). Panoramic construction with pixel resolution $N_\phi \times N_\theta = 1024 \times 256$ (a) using nearest neighbour technique and equi-angular pixelization, (b) using linear interpolation technique and equi-angular pixelization, (c) using linear interpolation technique and constant density pixelization.

The camera module utilized in the prototypes is PIXELPLUS PO4010N single-chip common intermediate format (CIF, 352×288). The thirty camera system contains two concentrator FPGA boards and one central FPGA [11]. The FPGA device utilization summary of a single concentrator board is stated in Table 1. The implemented architecture manipulates an output pixel every 20 clock cycles. The operating frequency of the implemented design in the FPGA is 133 MHz. Hence the video streaming rate for each concentrator FPGA board is 6.25M pixels per second. The power consumption of each FPGA board in operation is measured at 5 Watts. Three captured snapshots of the same scene from the real-time output (*i.e.*, 25 frame per second) of the Panoptic device with thirty embedded cameras is shown in Fig. 5. The horizontal and vertical directions in the shown panoramic constructions correspond to ϕ and θ spherical coordinates, respectively. Fig. 5(a) corresponds to the panoramic scene constructed for a virtual observer located at the center of the sphere using the nearest neighbour technique. No gain compensation and radiometric calibration has been used for the cameras. Hence the boundaries between the cameras are apparent and high intensity changes are visible in Fig. 5(a). The linear interpolation technique improves the color intensity variations as observed in Fig. 5(b) and provides a scene with less sharp color transitions. The linear interpolation techniques also result in ghosting effects for the objects that are closer to the Panoptic system. The ghosting effect is a known phenomenon in linear blending techniques. In Fig. 5(a) and

Table 1. FPGA device utilization summary

Resources	Used	Available	Utilization
Occupied Slices	3861	7200	53%
Slice Registers	10186	28800	35%
BlockRAM/FIFO	20	48	41%
DSP48Es	44	48	91%

(b) equi-angular pixelization has been utilized. In Fig. 5(c), the constant density pixel gridding is used. As observed in Fig. 5(c) the constant density gridding emphasizes the objects closer to equator of the sphere of the Panoptic system by enlarging them whereas the equi-angular grid emphasizes the top view of Panoptic system. The cameras of the Panoptic system have been calibrated for their true geometrical position in the world space and the extraction of their intrinsic paremeters.

5. CONCLUSION

The linear interpolation technique for the omnidirectional vision reconstruction of the Panoptic camera is presented. A novel pixelization gridding scheme is presented which enables an equal contribution scheme for the cameras of the Panoptic system. The architecture of an FPGA based system for the real-time deployment of the linear interpolation algorithm is shown. Sample pictures of the real-time output of a prototype Panoptic device are demonstrated. The introduced linear interpolation decreases the high light intensity variations in the reconstructed image but cause ghosting effects for the objects close to the Panoptic system. Multi-band blending technique is considered for the next real-time application deployment of the Panoptic device. Future work related to the Panoptic device focuses on distributed and parallel implementation of the Panoptic system in an interconnected network of cameras arrangement.

6. REFERENCES

[1] R. Szeliski, "Image mosaicing for tele-reality applications," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, Dec 1994, pp. 44–53.

[2] Mann, Picard, S. Mann, and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proceedings of IS&T*, 1995, pp. 442–448.

[3] Paul E. Debevec and Jitendra Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97*, 1997, pp. 369–378.

[4] Peter Rander, P. J. Narayanan, and Takeo Kanade, "Virtualized reality: constructing time-varying virtual worlds from real world events.," in *IEEE Visualization*, 1997, pp. 277–284.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH '96*, 1996, pp. 31–42.

[6] D. Taylor, "Virtual camera movement: The way of the future?," *Management Science*, vol. 7, pp. 93–100, 1996.

[7] Cha Zhang and Tsuhan Chen, "A self-reconfigurable camera array," in *ACM SIGGRAPH 2004 Sketches*, 2004, p. 151.

[8] Jason C. Yang, Matthew Everett, Chris Buehler, and Leonard McMillan, "A real-time distributed light field camera," in *Proceedings of the 13th Eurographics workshop on Rendering*, 2002, pp. 77–86.

[9] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, pp. 765–776, July 2005.

[10] Y. Yagi, "Omni directional sensing and its applications," *IEICE Transactions INF & SYST*, vol. E82-D, no. 3, 1999.

[11] Hossein Afshari, Laurent Jacques, Luigi Bagnato, Alexandre Schmid, Pierre Vandergheynst, and Yusuf Leblebici, "The panoptic camera: A plenoptic sensor with real-time omnidirectional capability," *Journal of Signal Processing Systems*, pp. 1–24, http://dx.doi.org/10.1007/s11265-012-0668-4.

[12] H. Afshari, L. Jacques, L. Bagnato, A. Schmid, P. Vandergheynst, and Y. Leblebici, "Hardware implementation of an omnidirectional camerawith real-time 3d imaging capability," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.