Depth extraction with offset pixels

W. J. Yun, Y. G. Kim, Y. M. Lee, J. Y. Lim, H. J. Kim, M. U. Khan, S. Chang, H. S. Park, and C. M. Kyung

1The School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
2The Center for Integrated Smart Sensors, Daejeon, South Korea
3QiSens, South Korea
4The Division of Electrical Engineering, Kongju National University, Cheonan, South Korea

*wjyun@duo.kaist.ac.kr

Abstract: Numerous depth extraction techniques have been proposed in the past. However, the utility of these techniques is limited as they typically require multiple imaging units, bulky platforms for computation, cannot achieve high speed and are computationally expensive. To counter the above challenges, a sensor with Offset Pixel Apertures (OPA) has been recently proposed. However, a working system for depth extraction with the OPA sensor has not been discussed. In this paper, we propose the first such system for depth extraction using the OPA sensor. We also propose a dedicated hardware implementation for the proposed system, named as the Depth Map Processor (DMP). The DMP can provide depth at 30 frames per second at 1920 × 1080 resolution with 31 disparity levels. Furthermore, the proposed DMP has low power consumption as for the aforementioned speed and resolution it only requires 290.76 mW. The proposed system makes it an ideal choice for depth extraction systems in constrained environments.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (100.2000) Digital image processing; (100.6890) Three-dimensional image processing; (110.6880) Three-dimensional image acquisition

References and links

#330162 https://doi.org/10.1364/OE.26.015825
Journal © 2018 Received 25 Apr 2018; revised 25 May 2018; accepted 27 May 2018; published 7 Jun 2018
1. Introduction

Depth sensing with a compact system opens a wide variety of exciting applications, especially in the era of hand-held devices. Numerous applications such as facial expression recognition, hand gesture recognition, image-based scanning and augmented reality require small and fast depth extraction imaging systems. Porting such applications to mobile devices presents a great challenge due to power, size and speed constraints. A system that can estimate color as well as depth of a scene and is portable on mobile devices will allow development of numerous innovative and life-changing applications.

Numerous depth extraction schemes have been developed in the past; however, their feasibility especially for mobile devices is questionable. Structured light-based methods [1] illuminate the scene with a light pattern and analyze the deformation in the light pattern to estimate depth. Depthsensing systems based on time-of-flight (ToF) sensors [2] consider the time taken by a light source to be reflected back to estimate the depth of the scene. Both these schemes require active illuminants; thus, these schemes do not work in outdoor conditions. Furthermore, generally the depth extracted by ToF sensors is blurred with over-smoothed object boundaries. Stereo matching, which has its roots in binocular vision of animals, has been the most thoroughly investigated depth extraction scheme in literature [3]. However, stereo-based schemes require at least two camera sensors to observe the given scene; thereby, these schemes require excessive resources. Perhaps, it is due to the above reasons that depth-sensing cameras have not penetrated through the commercial markets despite rigorous efforts.

To limit the resources used for depth extraction, depth extraction systems with a single sensor have been proposed recently. These systems estimate depth with a single shot. [4] proposes Dual Aperture (DA) camera which is a system based on two co-centric apertures with different diameters and spectral characteristics. They allow IR to pass through a smaller aperture and RGB through a larger aperture. The difference of blur of the IR and RGB images is used to estimate depth of the scene. In [5], Offset Apertures (OA) for the IR and RGB images are proposed, where the disparity of the IR and RGB images is used to estimate depth. Although the IR image allows well-aligned RGB images, the IR signal can corrupt the RGB image. Furthermore, the computational requirement of [4] is very high as it requires estimating the blur difference through convolution with point spread functions of various sizes.

Recently, the concept of Offset Pixel Aperture (OPA) was presented in [6] and its manufacturing was discussed in [7]. Pixels are offset by having the metal layer opening at different locations for pairs of pixels. OPA retains the advantage of OA as it is based on disparity which can be more accurately estimated compared to the de-focus blur. Moreover, OPA does not use the IR signal; thereby, it avoids the RGB signals being corrupted by the IR signal. Avoiding cross-spectral disparity estimation has yet another advantage in that improved depth quality is observed. Another key advantage that the OPA camera carries compared to stereo cameras is that image rectification is not required as the pixels are well aligned. This not only removes the computations but also reduces the amount of memory required to perform rectification.

Despite all its advantages, there are numerous challenges to adopt the OPA sensor for real
life applications. These include lack of green pixels, severe defocus, shading and noise artifacts. Perhaps, it is due to these reasons that the results of [6] and [7] are only limited computer simulations and a point light source, respectively. In this paper, we present a compact depth-extraction system for the OPA sensor for practical utility. The key contributions of this paper are as follows.

- We present the first demonstration of color image and depth extraction with the OPA principle. We propose a configuration of algorithms that can deal with the unique challenges presented by the OPA sensor. The performance of our method is validated through experimental results.

- We propose a dedicated implementation of the depth extraction system. For this, we make rigorous efforts towards minimizing the hardware costs of the dedicated implementation. As a result, the proposed system is capable of depth extraction in real time with WUXGA-resolution videos.

OPA carries numerous advantages compared to competing technologies. Generally, using disparity provides better estimates of depth of a scene compared to using blur. Having multiple sensors and/or rectification increase cost. Also, IR corrupts the RGB image quality. OPA uses disparity for depth estimation, requires a single sensor, does not require rectification and does not use IR, making it a better candidate for depth estimation as summarized in Table I.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth principle</td>
<td>Accuracy</td>
<td>Disparity</td>
<td>Blur</td>
<td>Disparity</td>
<td>Disparity</td>
</tr>
<tr>
<td>No. of sensors</td>
<td>Cost</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rectification needed</td>
<td>Cost</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>IR used</td>
<td>Image quality</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

The rest of this paper is organized as follows. The OPA camera is discussed in Section 2. The depth extraction method with all the constituent stages using OPA is described in Section 3. Section 4 describes the proposed hardware implementation for fast depth extraction. The system evaluation through experiments is discussed in Section 5.

2. Offset pixel aperture camera

Unlike preceding single-shot depth estimation systems, the OPA camera has apertures offset at the pixel level. In detail, for every row the centers of the odd pixels apertures are offset from the centers of the even pixel apertures. This offset is across the pixels of the same channel only. This generates a depth-dependent disparity across pixels, which indicates the depth. A graphical illustration of the OPA camera is shown in Fig. 1. The pixel apertures are at different locations compared to the pixel centers. The OPA sensor is fabricated using 110 nm, 2 poly and 4 metal layers CMOS Image Sensor (CIS) technology. In the conventional image sensor, the metal layers are commonly used for wiring transistors and do not cover the photodiode for sensing light. In contrast, the OPA sensor utilizes the first metal layer to form the offset aperture by covering the photodiodes. The pixel pitch is 2.8 um, the size of the pixel aperture is 1.3 um × 2.8 um and the pixel apertures are offset by 0.75 um from the pixel centers. In this work, we used a lens system with a focal length of 6 mm and an F-number of 1.4.

Offsetting the pixel apertures cannot guarantee depth extraction, rather a modified color filter array (CFA) is also required for the OPA camera. Although it is possible to use the common Bayer CFA with the OPA camera such that the offset apertures cover pixels from the same color channel,
the resulting depth quality will be strongly dependent on the scene characteristics. For example, if the offset apertures are placed over the green pixel then a scene lacking in green spectrum will have poor depth quality. To circumvent this problem, white pixels, with a uniform spectrum across the visible color channels and the near infra-red, are used with the offset apertures. The wide spectrum of the white pixel compensates for the reduced SNR resulting from the smaller aperture of the white pixel. Specifically, the offset pixel aperture size is 46% of the normal aperture but the spectrum of the white pixel is three times wider than that of the normal pixel. Thus, approximately 1.4 dB increase in SNR is observed. The sensor array used in the OPA camera is shown in Fig. 2. The white pixels of the odd and the even rows are offset from each other. This offset results in a disparity used for depth estimation. In the rest of the paper, we term the white channel image obtained by skipping the even rows as the left white image and the white channel image obtained by skipping the odd rows as the right white image. Because the green color pixel is absent, the green color is estimated by using the white, the red and the blue colors. The green color is obtained by subtracting the neighboring red and blue pixel values from the white pixel value as in [8].

An illustration of the relationship between depth and disparity is shown in Fig. 3. It is seen that
when the object is close to the OPA camera, the disparity is low. However, as the object is moved farther away from the OPA camera, the disparity increases across the left and right white images. As disparity is proportional to a blur, the far object has larger disparity and is more blurred as illustrated in Fig. 3.

![Fig. 3. A graphical illustration of the depth-disparity relationship with the OPA camera.](image)

### 3. Depth extraction using OPA camera

The end-to-end depth extraction process for the OPA camera can be divided into three stages: pre-processing, main processing and post-processing (see Fig. 4). Pre-processing prepares the input image for depth extraction while post-processing is used to remove errors from the obtained depth map. It should be noted that all the processes are chosen so that the hardware complexity remains as low as possible. Efforts to reduce hardware complexity at higher levels of abstraction, such as the algorithmic level, are significantly more influential compared to efforts at lower levels of abstraction [9].

![Fig. 4. Depth extraction process for OPA camera.](image)

#### 3.1. Pre-processing

In OPA, the disparity which is used for depth extraction in inevitably accompanied by a proportional blur. Some measures are, therefore, needed to suppress the blur while retaining the disparity information. For accurate matching across the left and right white images, textures are enhanced by extracting the gradients of the images. Gradient images exaggerate the intensity changes, thus, small textures are more prominent resulting in better matching. Gradient images are extracted as

\[ I'(x, y) = I(x + 1, y) - I(x - 1, y), \]  

(1)
where \( I(x, y) \) and \( I'(x, y) \) are the pixel intensity and the one-dimensional gradient at the position \((x, y)\), respectively.

Offset apertures in the OPA camera can cause illumination difference in the left and right white images. De-centered apertures of OPA camera cause aperture shading effect on the images, leading to a spatially-varying scale difference between left and right white images. To compensate for the scale difference, a local normalization is performed with respect to a local neighborhood \( \Omega \) of \( N \times N \) pixels.

\[
I_N(x, y) = \frac{I'(x, y) - \mu_\Omega(x, y)}{\sigma_\Omega(x, y)},
\]

(2)

where \( \mu_\Omega(x, y) \) and \( \sigma_\Omega(x, y) \) are the mean and standard deviation of pixels in the \( N \times N \) neighborhood centered at \((x, y)\).

Noise in the image can also severely effect patch-matching. There have been numerous noise reduction schemes proposed in the literature. However, to achieve high efficiency, we use simple mean-filtering for noise reduction.

### 3.2. Main processing

To obtain the depth map of the scene, pixel-to-pixel correspondence across the left and the right white images are required. It is known that matching single pixels is quite erratic. Therefore, pixel correspondence is obtained by not only considering the intensity at a given pixel but also the neighborhood of each pixel. Using Sum of Absolute Difference (SAD) the cost for every pixel is given by

\[
C(x, y, d) = \sum_{x,y \in \Pi} | I_L(x, y) - I_R(x + d, y) |,
\]

(3)

where \( I_L \) is the left white image, \( I_R \) is the right white image, and \( \Pi \) is the window centered at \((x, y)\). It should be noted that using SAD does not compensate for scale differences across the left and the right white images. However, the scale difference has already been compensated for in the pre-processing stage. This is a great advantage computationally. Local scale invariant schemes such as Normalized Cross Correlation (NCC) and [10] are computationally much more complex compared to SAD. For every pixel at every disparity level, NCC requires \( N \times N \) multiplications and additions whereas the proposed method does not require any multiplications.

Cost aggregation methods improve the cost at a given pixel location based on the neighboring pixels. Methods that consider the whole image for cost aggregation, also called global methods, tend to be robust against noise but over smooth the depth map. On the contrary, local methods respect object boundaries in the depth map but are prone to errors. To obtain a better compromise, the Semi-global matching (SGM) scheme has been proposed [11]. The aggregated cost a given pixel is obtained by aggregating costs from equiangular pixel paths that converge at the given pixel. Generally, eight paths are used.

Using eight paths for cost aggregation can be very costly in terms of hardware. Pixel intensities are obtained from the camera in row-by-row fashion. Similarly, the cost at each disparity is also obtained row-by-row. To implement the eight-path cost aggregation, we need to store the cost at all disparities for the whole image as seen in the Fig. 5(a). This requires a huge amount of memory. Our aim is to estimate the aggregated cost in an online fashion, i.e., compute the aggregated cost with the available costs only. Therefore, we aggregate costs along the rows of the image as shown in the Fig. 5(b). Mathematically,
\[ A_r(x, y, d) = C(x, y, d) + \min \left\{ \begin{array}{l} A_r(x - 1, y, d) \\ A_r(x - 1, y, d - 1) + P_1 \\ A_r(x - 1, y, d + 1) + P_1 \\ \min_i A_r(x - 1, y, i) + P_2 \end{array} \right\} - \min_k A_r(x - 1, y, k), \quad (4) \]

where the last term is included for normalization. \( A_r(x, y, d) \) and \( C(x, y, d) \) are the aggregated cost and the cost at the current pixel \((x, y)\) and disparity \(d\), respectively, and \(P_1\) and \(P_2\) are penalty terms. The penalty \(P_1\) is typically a constant and the penalty \(P_2\) is inversely related to the gradient value. This is similar to the approach of [12] where aggregation is performed over every row, however, we use a different set of rules for aggregation compared to [12]. By applying the above equation, the disparity is smoothed based on the color as well as depth similarity of the neighboring pixels.

![Fig. 5. The paths for cost aggregation where (a) shows the typical eight-path approach and (b) the approach used in this work.](image)

After cost aggregation, winner-takes-all (WTA) is performed for disparity estimation at a pixel. Though interpolation schemes are available for choosing the best disparity, we use WTA as it requires relatively fewer computational resources.

3.3. Post-processing

It is a general observation that the sudden changes in the depth map should occur at pixels only where there are sudden changes in the intensity. This observation has been used by numerous methods to remove errors from the depth map such as weighted median [13], joint-bilateral [14] and joint-guided filtering [15]. However, these approaches are too cumbersome for an efficient implementation. Here, we propose the following method for removing errors from the depth map.

\[ D(x, y) = \frac{\sum_{i,j \in \psi} D_i(i, j) \mathbb{1}\{|I(i, j) - I(x, y)| < S\sqrt{I(x, y)}\}}{\sum_{i,j \in \psi} \mathbb{1}\{|I(i, j) - I(x, y)| < S\sqrt{I(x, y)}\}}, \quad (5) \]

where \(D_i\) is the disparity before post-processing, \(S\) is a threshold parameter, \(\psi\) denotes an window centered at \((x, y)\) and \(\mathbb{1}\{.\}\) is the indicator function which returns 1 if the condition in the parenthesis is true. In effect, only inliers are averaged in the window, where the inliers are determined by a threshold set by the shot noise assumption over the image. Under the shot noise assumption, the noise is proportional to the root of the intensity. Thus, neighboring pixels the intensity difference of which with the center pixel is greater than the threshold are considered as outliers.
4. Dedicated implementation

Although the OPA camera is aimed towards low power and small area, it may lose its purpose if the depth extraction process consumes significant computational power or area. Also, the disparity estimation process is not fast enough for real-time depth extraction in practical camera applications. For acceleration and low power consumption, all depth extraction processes are implemented on a dedicated hardware platform, which is termed as the depth map processor (DMP). DMP has a core block for depth extraction (pre-processing, main processing and post-processing), and peripheral blocks for configuration and transmission (I2C slave, SPI slave, and depth map transceiver) as shown in Fig. 6. All these blocks are integrated on a single chip. The interface block is for configuring the core blocks as well as external communication, for which we use an I2C slave, SPI slave and a depth map transceiver. DMP can be reconfigured through I2C and SPI slave as per the application. Input image size, disparity range, and parameters of the cost aggregation and the depth noise reduction are reconfigured by setting the internal registers through serial interfaces. Also, any functional block can be skipped by changing the configuration. Depth map generated by the core block is transmitted through the depth transceiver. Note that the DMP is fully pipelined to achieve high frame rate and does not include a frame buffer.

![Fig. 6. Structure of DMP.](image)

For applying local patch-based processes, such as SAD, a window of pixels in the image needs to be stored. To process all the windows in an image, we use a FIFO which to store the inputs of the image. Suppose the window is sized $M \times M$ and the image has a width of $W$ then for window-based operations, we require to store $M - 1$ rows of length $W$. If $B$ is the bit-width then the memory required for moving-windows-based operations is given by

$$\text{Memory} = (M - 1)WB.$$  \hfill (6)

DMP can handle images of width up to 2048 pixels. However, smaller-sized memories are required in some computations as the left and right white images are half the size of the input image. Input bit width of DMP is 10. DMP uses 12 scanline buffer memories. The size of each of these scanline buffer memories is shown in Table 2.

The channel splitter constructs the four channel images from the raw image. During channel splitting, a scanline buffer is used because two lines are needed to extract four channel components for the current pixel. The local normalization in the preprocessing block is performed for both the left and right white images. The implementation of Eq. (2) is shown in Fig. 7. Before the local normalization, the data bit width is reduced from 10 bits to 8 bits by truncation of LSB 2 bits for
Table 2. Size of scanline buffer memories

<table>
<thead>
<tr>
<th>Block</th>
<th>Size [KBytes]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-processing</td>
<td></td>
</tr>
<tr>
<td>Channel splitter</td>
<td>$5 \times 1$</td>
</tr>
<tr>
<td>Mean filter</td>
<td>$4 \times 4$</td>
</tr>
<tr>
<td>Local normalization</td>
<td>$8 \times 2$</td>
</tr>
<tr>
<td>Stereo matching unit</td>
<td>$32 \times 2$</td>
</tr>
<tr>
<td>Post-processing</td>
<td></td>
</tr>
<tr>
<td>Depth noise reduction</td>
<td>$8 \times 1, 5 \times 1$</td>
</tr>
<tr>
<td>Synchronizer</td>
<td>$44 \times 1$</td>
</tr>
<tr>
<td>Total</td>
<td>158</td>
</tr>
</tbody>
</table>

reduction of hardware cost. To reduce data bit-width while minimizing the loss of information, gamma correction performs a non-linear correction on the intensity values so the low intensity pixels are not completely vanished by bit-truncation. For referring to the neighboring pixels, a scanline memory is used. The size of neighborhood $\Omega$ is $9 \times 9$, the channel image width is 1024, and the bit width is 8. Thus, the size of the scanline memory is 8 KBytes, also seen from Eq. (6). The mean of pixel intensities in the window $\mu_{\Omega}$ is the output of the upper block and the standard deviation $\sigma_{\Omega}$ is the output of the lower block. The normalized intensity is the output by the divider.

![Diagram](image)

**Fig. 7.** Hardware block diagram of the local normalization, which is the implementation result of Eq. (2).

DMP uses SAD for cost generation. SADs for all 31 disparity levels are performed in parallel as shown in Fig. 8. Each SAD unit is implemented by a subtractor and a summation of the absolute values. Two 32 KBytes scanline memories are used to store the pixels of the windows from the left and right white images, where the size of the window $\Pi$ is $65 \times 33$. To store a window with disparity $d$ in $I_R$, the pixel values are stored in a shift register. This results in the SAD for 31 disparity levels.

Costs for each disparity level are aggregated in parallel as shown in Fig. 9. On the RHS of Eq. (4), the summations of the second term, $P_1$, and $P_2$ are implemented in the each SGM unit as well as the subtraction of the last term. The second term on the RHS of Eq. (4) is calculated by the comparator in the each SGM unit. The minimal aggregated cost which is the last term on the RHS of Eq. (4) is the output of the left upper comparator. In Eq. (4), the aggregated cost of the previous pixel $(x - 1, y)$ is used to aggregate cost of the current pixel $(x, y)$. Thus, the output of the cost aggregation block, which is the aggregated cost of the current pixel $(x, y)$, is fed to the logic blocks for aggregation at the next pixel $(x + 1, y)$.

The proposed post-processing scheme of Eq. (5) is implemented as shown in Fig. 10. The image bit width is 8, the depth map bit-width is 5 and the size of the window $\psi$ is $9 \times 9$. Therefore, 8 KBytes and 5 KBytes of scanline memories are used for buffering the reference image and the
5. Experimental results

In this section we present the evaluation of the proposed system. More specifically, we present the quantitative and qualitative results with the depth extraction system. The ASIC implementation results of the dedicated hardware are discussed. Some applications of the proposed system other than depth extraction such as night vision and 3D reconstruction are also presented.

5.1. Quantitative analysis

For quantitative evaluation, we placed a flat and textured surface at varying distances from the camera and estimated its depth. More specifically, the flat surface is placed at distances ranging from 20 cm to 200 cm with a 3 cm interval. For comparison with the existing single lens based depth sensing cameras such as OA [5] and Pixel Aperture camera (PA) [16], we performed the same experiment with both the OA and PA cameras. PA has two co-centric pixel apertures with...
Fig. 10. Hardware block diagram of the depth noise reduction, which is the implementation result of Eq. (5).

different diameters by metal layer opening. The difference of blur of the images obtained from the different size of the apertures is used to estimate depth of the scene. The experimental results of the estimated depth against the distance is shown in Fig. 11. The results show that the proposed system has much longer depth range and much lower depth variance than both the OA and the PA.

For more detailed analysis of the result, we introduce a confidence interval which shows the reliability of estimated values based on an assumption that the samples follow the Gaussian distribution. In the experimental result, the confidence interval is defined as distinguishable distance interval with 95 % probability and it can be seen as the reliability of the estimated depth.

\[
Q(x) = \left( x - \frac{1.96 \times \sigma(x)}{\Delta d} x + \frac{1.96 \times \sigma(x)}{\Delta d} \right),
\]

where \( \sigma(x) \) is the standard deviation of the estimated depth at distance \( x \) and \( d \) is estimated depth. The lower confidence interval, the better it is. For each distance of 20 cm, 100 cm, and 200 cm, the proposed has a confidence interval of (20-0.14,20+0.14) cm, (100-3.4,100+3.4) cm, and (200-12.2,200+12) cm, respectively. PA has a confidence interval of (20-0.65,20+0.65) cm, (100-13,100+13) cm, and (200-57.2,200+57) cm for same distances. The available depth extraction range of the OA is limited to 50 cm as shown in Fig. 11(b), because the slope of the graph is close to zero at more than 50 cm. Thus, the confidence interval of the OA is infinite from more than 50 cm since the denominator on the Eq. (7) is close to zero. For the distance of 20 cm, the OA has the confidence interval of (20-1.7,20+1.7) cm. For all distances, the proposed has the lower confidence interval.

5.2. Color quality of OPA camera

Since the green pixel is missing in the OPA CFA, the green color is first generated by a linear combination of the white, red and blue pixel values as follows.

\[
g(x, y) = c_1w(x, y) + c_2r(x, y) + c_3b(x, y) + c_4
\]

where \( g(x, y), w(x, y), r(x, y), \) and \( b(x, y) \) are the green, white, red, and blue pixel intensities at pixel position \((x, y)\), respectively. \( c_1, c_2, c_3, \) and \( c_4 \) are coefficients. For the green color restoration, we use the standard 24-color classic X-rite color chart with 8-bit encoding and the ground truth of the color chart. The linear regression with least squares is used to obtain the coefficients. After the green color restoration, demosaicing is performed to get the RGB image with full resolution. Bicubic interpolation is adopted for the demosaicing. The results are shown in Fig. 12. The result of captured color chart has an RMS error of 11.64 compared to the ground truth color chart.
Fig. 11. Plot of the average estimated depth of a flat surface against its distance with (a) the proposed, (b) the OA, and (c) the PA cameras. The bars show the variance in disparity.

Fig. 12. Color quality evaluation of OPA camera where (a) shows a synthetic ground truth color chart, (b) is the captured color chart by OPA camera, reference image taken by a mobile camera is shown in (c), and (d) is captured by OPA camera for qualitative analysis.
5.3. Hardware implementation

In this subsection the ASIC implementation results of the DMP proposed in Section 4 are presented. The DMP has been implemented on a 0.11 um CIS process and is shown in Fig. 13. A special block for self-check has been also included in the DMP, which passes test patterns to the DMP chip for evaluation. These test patterns are stored in a ROM. Our efforts to simplify the depth extraction algorithm pays dividends as the chip has a core area of 52.63 mm$^2$ with a gate count of 1.156 M and 219.4 KBytes of memory. The gate count for each block is shown separately in Table 3. It is seen that most of the computational resources are required by the stereo matching unit. It is not surprising as the stereo matching unit has to compare image patches at numerous disparity levels. Note that the proposed DMP chip can handle images of widths up to 2048 pixels and a maximum disparity of 31.

![Fig. 13. Micro-photograph of the DMP chip and details of its implementation](image)

Table 3. Gate Count of DMP Sub-Blocks

<table>
<thead>
<tr>
<th>Block</th>
<th>Gate count[K]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth transceiver</td>
<td>1.115</td>
</tr>
<tr>
<td>12C slave</td>
<td>5.677</td>
</tr>
<tr>
<td>SPI slave</td>
<td>6.153</td>
</tr>
<tr>
<td>TPG, Synchronizer, Debouncer</td>
<td>40.152</td>
</tr>
<tr>
<td>Pre-processing</td>
<td></td>
</tr>
<tr>
<td>Channel splitter</td>
<td>36.31</td>
</tr>
<tr>
<td>Mean filter</td>
<td>109.349</td>
</tr>
<tr>
<td>Local normalization</td>
<td>89.616</td>
</tr>
<tr>
<td>Stereo matching unit</td>
<td></td>
</tr>
<tr>
<td>SAD</td>
<td>728.681</td>
</tr>
<tr>
<td>SGM</td>
<td>42.169</td>
</tr>
<tr>
<td>WTA</td>
<td>28.518</td>
</tr>
<tr>
<td>Post-processing</td>
<td></td>
</tr>
<tr>
<td>Depth noise reduction</td>
<td>68.260</td>
</tr>
<tr>
<td>Total</td>
<td>1156</td>
</tr>
</tbody>
</table>

The proposed hardware implementation shows high performance in speed and power. The maximum clock frequency is 80 MHz. In other words, the system can process 80 mega pixels per
second. The proposed DMP is capable of processing 30 frames per second (fps) of Full HD (1920 × 1080) resolution while consuming 290.76 mW. The power consumption is almost negligible to modern systems with high speed capacity, for example, GPUs. For extremely power constrained environments, the DMP can process 800 × 600 resolution frames at 20 fps while consuming 51.99 mW only. The power consumption results against the clock frequency are shown in Fig. 14. The complete hardware solution including the DMP chip, the ASIC board, the OPA sensor and the OPA camera evaluation board is shown in Fig. 15. The output of the DMP chip is shown in Fig. 16. The size of the depth map is reduced by the matching window size of the SAD and the size of the search range. Suppose the size of the left and the right white images is $W \times H$, the size of the matching window of the SAD is $M \times N$, and the search range is $R$, then the size of the resulting depth map is given by \((W - (M - 1) - (R - 1)) \times (H - (N - 1))\). Here $W = 771$, $H = 551$, $M = 65$, $N = 33$, and $R = 31$. Thus, the resulting depth map is $677 \times 519$. This image boundary removal also compensates for radial distortion. The objects in the foreground can be clearly distinguished from the background, showing the effectiveness of the proposed system. Also, a demonstration of the real-time depth extraction with the proposed system is provided as a supplementary material (see Visualization 1).

![Fig. 14. Power consumption of the DMP with operating frequency.](image)

### 5.4. Applications

The depth extraction system proposed in this work can be used in numerous application. Numerous methods cannot be implemented in constrained environments such as mobile applications due to their slow speed, high power consumption or requirement for bulky platform. The proposed system, on the other hand, has high speed, low power consumption and requires less area; therefore, it is ideal for mobile applications. All applications that require depth can utilize the proposed system.

3D reconstruction is a classic problem of generating a 3-dimensional representation of a given scene. 3D reconstruction typically requires multiple images of the same scene. In fact, techniques based on structure from motion or bundle adjustment require a huge number of images for accurate 3D reconstruction. These schemes are computationally very expensive and can take long durations to generate the 3D reconstructed images. However, with the proposed system, 3D reconstruction can be performed with a single shot of the scene. In Fig. 17, a 3D reconstruction
Fig. 15. The components of the implemented system. (a) DMP chip on ASIC board, (b) the OPA sensor, and (C) the OPA camera evaluation board.

Fig. 16. Scenes observed by the OPA camera and their depths estimated by the DMP. Gray scale bar shows disparity. The operating distance of the scenes is from 20 cm to 120 cm.
of the human face using the proposed system is shown. Since a normal human face has no depth discontinuities, we apply sub-pixel disparity estimation [17] and more aggressive smoothing compared to the post-processing approach mentioned earlier. We perform mean filtering to the depth map followed by texture mapping for visualization. The reconstructed image clearly maintains the 3D structure of the face. This can be used to counter the efforts made to counterfeit face, where faces printed on an image are used to fool the face detection system. The proposed system can be used to identify if the face is 2D (fake) or 3D (real).

Fig. 17. 3D reconstruction of a human face using the proposed OPA system.

An interesting application of the proposed system is its capability of night vision. Not only a given scene can be observed without visible light but also the depth of the scene can be estimated. This is due to the use of white pixels for depth extraction. The white pixels have a wide spectrum, absorbing near infra-red as well. Thus, the proposed system not only works in visible light but also in darkness under an IR illuminant if the IR cut filter is detached. The results are shown in Fig. 18.

6. Conclusion

In this paper, we present the first working system for depth extraction with the proposed offset pixel aperture (OPA) scheme. A fast depth-extraction system for the OPA sensor has been proposed. The complete system is presented in detail. Also, an ASIC implementation of the system is presented in the paper. It is seen that the proposed system can extract depth with
Fig. 18. (a) The observed scene with a normal camera under IR illumination, (b) the same scene captured by OPA camera under IR illumination, and (c) its depth by the proposed system.

high speed, low power and a small area. This makes the proposed system suitable for many applications, especially embedded systems such as mobile applications where the speed, power and area constraints are the most strictest. The proposed system is capable of estimating depth of a scene in real-time at Full HD resolution.

**Funding**

Center of Integrated Smart Sensors within the Ministry of Science, ICT and Future Planning (CISS-2013073718).