# Time-consistent Depth Estimation for 2D-to-3D Conversion System Using Color Histogram and Variance Map

Seungjun Shin[1], Kyeongbo Kong[2], Sung In Cho[3] and Woo-Jin Song[4]
[1,2,4]Department of Electrical Engineering, POSTECH
77 Cheongam-Ro, Nam-Gu, Pohang, Gyeongbuk, Republic of Korea
[3]LCD Division, LG Display Inc.
245 LG-ro, Wollong-myeon, Paju-si, Gyeonggi-do, Republic of Korea
E-mail : {[1]ymah_89, [2]kkb4723, [4]wjsong}@postech.ac.kr, [3]csi2267@lgdisplay.com

**Abstract**: We present a time-consistent depth estimation algorithm for 2D-to-3D video conversion. The algorithm exploits variance of image to identify the foreground, and uses modified scene change detection method that exploits a color histogram. In simulations the proposed algorithm did not commit depth-reversal or saliency-instability.

*Keywords*—**Saliency, Depth-reversal, Estimation instability**

## 1. Introduction

The major impediment to development of the 3D display market is lack of video that has 3D content. Therefore, 2D-to-3D conversion systems are required, but development of such systems is a difficult challenge. Depth can be estimated using saliency, which denotes the region to which the human visual system pays most attention. Methods that use saliency generally assign the closest depth to the most-salient points.

Saliency-based depth estimation [1] uses color saliency to assign a depth map; this method is efficient, but has two major problems. 1) It uses a strong assumption that the object occupies a smaller portion of the scene than does the background; this assumption is frequently violated and when it is, the depth map can be reversed. 2) The saliencies of the colors are mutually dependent, so when the proportion of a scene that is a given color changes, the saliences of this color and of all other colors are affected; the result may be a rapid change in the depth map. In this paper, we propose a saliency-based depth estimation for 2D-to-3D conversion algorithm that does not have these problems.

(a) Zoom-in image (b) Depth map of (a) using saliency-based depth estimation [1]

Figure 1. Depth-reversal problem

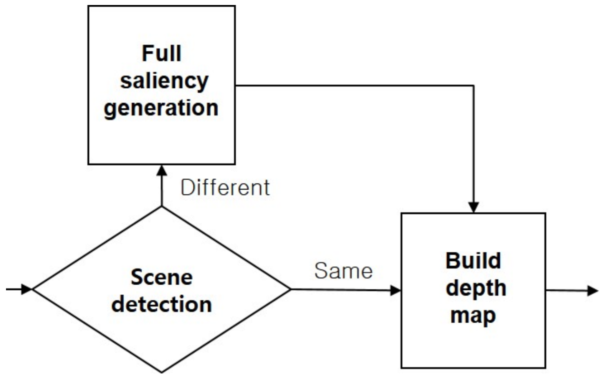



Figure 2. Instability system diagram of saliency-based depth estimation [1]

## 2. Conventional algorithm [1]

### 2.1 Saliency generation

The saliency of color $c_l$ is defined as

$$S\left(c_l\right)=\sum_{j=1}^{n} f_j \cdot \left(D\left(c_l, c_j\right)\right) \tag{1}$$

where $D\left(c_l, c_j\right)$ is the color distance between $c_l$ and $c_j$, $n$ is the number of different pixel colors, and $f_j$ is the frequency of pixel color $c_j$ in the image [2].

The calculation of (1) is computationally too expensive for real-time application. To reduce the complexity, a histogram is generated by quantizing all colors into $12^3$ kinds, and saliency is calculated only for the most frequent colors (i.e., which occupy the top $95\%$ of the histogram). Each infrequent color is considered to be the same as the closet frequent color. This process can result in similar colors being quantized to different values; to reduce this quantization artifact the color space is smoothed as shown by

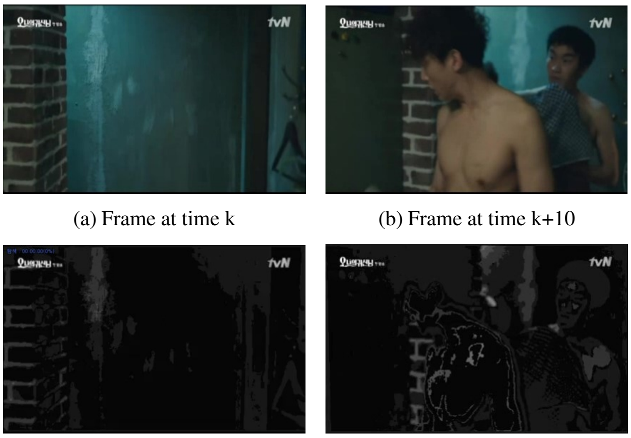
$$S^{'}\left(c\right)=\frac{1}{\left(m-1\right)}\sum_{i=1}^{n}\left(T-D\left(c, c_i\right)\right) S\left(c_i\right)\Big) \tag{2}$$

where $\sum_{i=1}^{m}\left(D\left(c, c_i\right)\right)$ is the sum of distances between color $c$ and its $m$ nearest neighbors.

The final depth map is defined as

$$Depth\left(c_l\right)=\frac{S\left(c_l\right)-S_{\min}}{S_{\max}-S_{\min}} \cdot 255 \tag{3}$$

where $S_{\min}$ and $S_{\max}$ are the minimum and maximum saliency values of all colors in the image.



(a) Frame at time k (b) Frame at time k+10

(c) Depth map of (a) using saliency based algorithm (d) Depth map of (b) using saliency-based algorithm

Figure 3. The solution [1] to the instability problem

### 2.2 Depth-reversal

Saliency based on a color histogram used as depth entails a strong assumption that the object is smaller than the background. This assumption is frequently violated; when this happens, the color of the object has low saliency, so it is assigned a low depth value, and the depth map inverses. This phenomenon is called the 'depth-reversal problem'. As shown in figure 1, the assumption that the salient object (head) is smaller than the background is violated (Fig. 1a). The head is assigned low saliency and is therefore given a distant depth value (dark) (Fig. 1b).

### 2.3 Instability problem

Equation (1) indicates that if the proportion of an image that is a given color changes, the saliency of all other colors also changes. As a consequence, the saliency of an object can change in consecutive frames (the instability problem). To eliminate this instability, [1] proposed a scene-detection system in which the saliency is not updated until the scene changes. The system check two simplified criteria. Color type change is defined as

$$ratio_n^{cl} = \frac{n\left(\{c_l\}_n \setminus \{c_l\}_{n-1}\right)}{n\left(\{c_l\}_{n-1}\right)} \tag{4}$$

where $n\left(\{c_l\}_{n-1}\right)$ is the count of frequent colors in frame $(n-1)$, $\setminus$ is the set relative complement operation. Color histogram change is given by

$$ratio_n^{hist} = \sum_{i=1}^{n_\cap} \left|h\left(c_i\right)_n - h\left(c_i\right)_{n-1}\right| \tag{5}$$

where $n_\cap$ is the count of common frequent colors in frame $n$ and $(n-1)$, and $h\left(c_i\right)_n$ is the bin count of color $c_l$ in frame $n$. $ratio_k^{cl}$ is examined first, and $ratio_k^{hist}$ is tested only when $ratio_k^{cl}$ is small. Either large $ratio_k^{cl}$ or $ratio_k^{hist}$ indicates a different scene. The algorithm update a saliency value if and only if a new scene is detected. This approach stabilizes the depth map effectively, but cannot assign a saliency value to a new frequent color (Fig. 2d) while the scene does not change (Fig. 2b).

## 3. Proposed algorithm

To solve these problems, we propose the following process.

### 3.1 Saliency generation

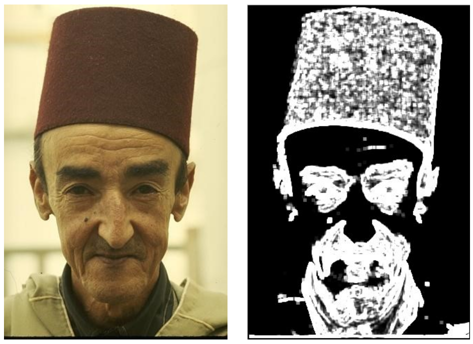
We adopt the saliency (1).

### 3.2 Local feature calculation

Commonly, the salient object has more details and boundaries than does the background, so the object has higher variance than does the background (Fig. 4). Therefore we use the variance map to identify the object. The variance map is defined as

$$var\left(x\right) = u_k\left(x^2\right) - \left(u_k\left(x\right)\right)^2 \tag{6}$$

where $x$ is the elements of the pixel domain, $u_k\left(\cdot\right)$ is the local mean with size-$k$ window.



(a) Original image (b) Variance map of (a)

Figure 4. Variance map as local feature of object

(a) NCC calculation using depth map (b) NCC calculation using complement depth map

Figure 5. Checking for reversal

### 3.3 Checking for reversal

To check whether depth reversal occurs, we use normalized cross-covariance

$$NCC = \sum_x \frac{\left(d\left(x\right) - \mu^d\right)\left(var\left(x\right) - \mu^{var}\right)}{N\sigma_{var}\sigma_d} \tag{7}$$

where $x$ is the elements of the pixel domain, $d\left(\cdot\right)$ is the depth map of pixel domain, $\mu^d$ is the mean of $d$, $\mu^{var}$ is the mean of $var$, $\sigma_d$ is the standard deviation of $d$, $\sigma_{var}$ is standard

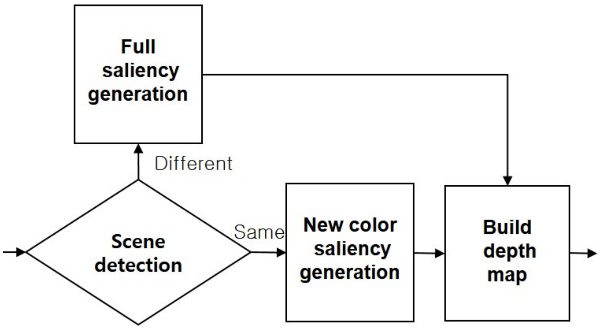



Figure 6. Instability system diagram of the proposed algorithm

deviation of $var$, and $N$ is the number of pixels. We also define a complement depth map

$$Depth'(c_l) = 255 - Depth(c_l), \tag{8}$$

and compare it to the depth map by $NCC$. If $NCC$ is higher in the complement depth map than in the depth map, then depth reversal has occurred. In this case, $Depth'$ is selected as the final depth map instead of $Depth$.

### 3.4 Instability removal

We modify the scene change detection system [1] to solve this problem. Firstly, if the scene does not change, the system updates saliency only if it detects a new frequent color that was not present in the previous frame; this process may cause flicker in the image, because the normalization value ($S_{\max}$, $S_{\min}$) changes. To prevent this flicker, the system saves the normalization value of the first until the scene changes. By maintaining the normalization value and updating only the saliency of a new frequent color, the algorithm prevents instability without failing to consider that color.

## 4. Simulation results

In this section, we present simulation results to evaluate the numerical performance and visual quality. We compare the proposed algorithm with saliency-based algorithm [1] RGBD data sequences; Ballet $(1024 \times 768)$, beer garden $(1920 \times 1080)$, break dancing $(1024 \times 768)$, book arrival $(1024 \times 768)$, tanks $(400 \times 300)$, book $(400 \times 300)$. By comparing normalized cross-covariance(NCC) with ground-truth depth map, the performance of these algorithms is evaluated. Table 1 shows the average NCC values for several sequences. From Table 1, we can see that the proposed algorithm had slightly higher estimation accuracy than saliency-based algorithm [1]. We can also see that the proposed algorithm prevented depth-reversal (Fig. 7), and eliminated instability without failing to consider a new frequent color (Fig. 8).

## 5. Conclusion

We presented a time-consistent saliency-based depth estimation algorithm. To prevent depth-reverse problem, we use the variance map as local feature of object, and check whether depth reversal occurs by using normalized cross-covariance.

(a) Original image (b) Depth map of (a) using saliency-based algorithm (c) Depth map of (a) using the proposed algorithm

(d) Original image (e) Depth map of (d) using saliency-based algorithm (f) Depth map of (d) using the proposed algorithm

(g) Original image (h) Depth map of (g) using saliency-based algorithm (i) Depth map of (g) using the proposed algorithm

(j) Original image (k) Depth map of (j) using saliency-based algorithm (l) Depth map of (j) using the proposed algorithm

Figure 7. The solution to the depth-reversal problem

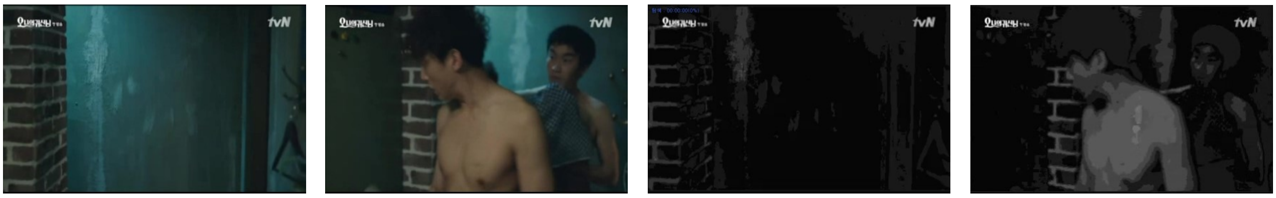
We also improve instability removal process of conventional algorithm. As a result, the proposed algorithm did not commit depth-reversal or instability problem, and achieve higher NCC value. Fig. 9 shows system diagram of the proposed algorithm.

## 6. Acknowledgement

This work was supported by LG Display under LGD-Postech research Program and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) (IITP-2016-H8601-16-1005) supervised by the IITP(Institute for Information & communications Technology Promotion).

Table 1. Normalized cross-covariance with ground truth

| **Algorithm** | **Image** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Ballet** | **Beer garden** | **Book** | **Book arrival** | **Break dancing** | **Tank** | **Avg.** |
| **Proposed** | 0.2698 | 0.5172 | 0.3678 | **0.7857** | **0.7734** | 0.2728 | 0.4978 |
| **Saliency-based** | 0.2698 | 0.5172 | 0.3678 | 0.0593 | 0.4630 | 0.2728 | 0.3250 |



(a) Frame at time k

(b) Frame at time k+10

(c) Depth map of (a) using the proposed algorithm

(d) Depth map of (b) using the proposed algorithm

Figure 8. Instability removal of consecutive frames

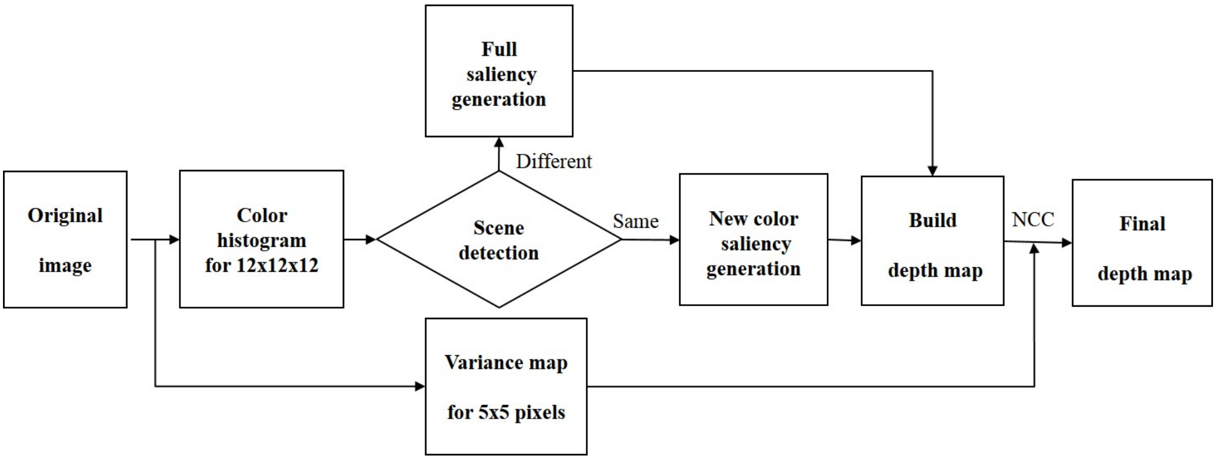



Figure 9. System diagram of the proposed algorithm

## References

[1] Zhen Zhang, Shouyi Yin, Leibo Liu and Shaojun Wei, "A real-time time-consistent 2D-to-3D video conversion system using color histogram," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 4, pp. 524-530, Nov. 2015.

[2] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr and S. M. Hu, "Global Contrast Based Salient Region Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, March 1 2015.