

Research Article

Merging Static and Dynamic Depth Cues with Optical-Flow Recovery for Creating Stereo Videos

Fang-Hsuan Cheng¹ and Tze-Yun Sung²

¹ Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu City 300-12, Taiwan

² Department of Electronics Engineering, Chung Hua University, Hsinchu City 300-12, Taiwan

Correspondence should be addressed to Tze-Yun Sung; bobsung@chu.edu.tw

Received 20 February 2013; Revised 18 April 2013; Accepted 27 April 2013

Academic Editor: Shengyong Chen

Copyright © 2013 F.-H. Cheng and T.-Y. Sung. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A method for estimating the depth information of a general monocular image sequence and then creating a 3D stereo video is proposed. Distinguishing between foreground and background is possible without additional information, and then foreground pixels are moved to create the binocular image. The proposed depth estimation method is based on coarse-to-fine strategy. By applying the CID method in the spatial domain, the sharpness and the contrast of an image can be improved by the distance of the region based on its color. Then a coarse depth map of the image can be generated. An optical-flow method based on temporal information is then used to search and compare the block motion status between previous and current frames, and then the distance of the block can be estimated according to the amount of block motion. Finally, the static and motion depth information is integrated to create the fine depth map. By shifting foreground pixels based on the depth information, a binocular image pair can be created. A sense of 3D stereo can be obtained without glasses by an autostereoscopic 3D display.

1. Introduction

For more than 100 years, the concept of depth in 3D images and videos has existed. Wheatstone [1] first created a stereoscopic picture pair by using the binocular parallax theory. Wheatstone makes the first prism stereoscope based on visual parallax theory. David Brewster (1781–1868) [2] uses two lenses to build a prism stereoscope. Both of them used two cameras. A picture would be taken for both the right and the left eye. The end result is a 3D effect when looking through prism stereoscope. George Swan Nottag, a London merchant, established a company for 3D glasses in 1845. More than one million 3D glasses and stereoscopic images were sold within four years. Brewster developed lenticular stereoscopic in 1858. Anderton [3] proposed a method to make 3D projectors by using polarized light in 1895. The inventor of TV, John Baird, showed 3D picture on his TV in 1942. Half a century later, the Japanese company SONY was trying to sell their 3D TV. At the same time, NHK was trying to provide a 3D TV service. Anaglyph 3D movies were very popular in 1950. The popularity of the 3D application caused the development of new technology which has been making rapid progress.

As the technology progresses, vivid 3D stereoscopic vision is drawing more and more attention. Since 3D stereoscopic movies and films are becoming more and more popular, 3D stereo application will lead to the evolution of the next generation TV system [4–8]. The 3D stereoscopic display technology has developed from red-blue glasses in the early days to 3D LCD display [9] without glasses. All these stereo equipments allow people to perceive a 3D stereo effect by feeding parallax images into the left and right eyes. Therefore, how to derive the binocular image pairing from a 2D image sequence has become a focus in research.

When looking at an object, the differences between the left and right eye are referred to as parallax. The distance in depth between objects is picked up because of the parallax of images. A pair of images which has parallax information can be used to give the perception of depth between objects, just like when the human brain perceives something. Using this theorem, an image which has parallax information can be created. This then gives the impression of 3D stereo when using a lenticular auto-stereoscopic 3D display [9].

It is very easy to generate a stereo image by using two cameras shooting the same image simultaneously [10]. The image

pair can be fed into the left and right eyes and a stereo image effect forms in the mind of the viewer. In order to make the 3D stereo image more readily available, a 3D stereo image sequence can be derived from a 2D image sequence. The most important 3D stereo display technology can be classified into two categories. One is obtaining the depth information from a single 2D image sequence [11, 12] and the other is deriving a binocular image pair from the image together with its depth information [13, 14]. In this study, the computed image depth (CID) method was used to estimate the coarse depth of a single image, and then the binocular image pair can be created by an image shifting operation based on the temporal motion information estimated with optical-flow method.

2. Related Works

Much attention has been paid to 3D stereo technology in recent years. 3D stereo products are becoming more and more popular. Getting the correct depth map is significant. The most popular and easy way to get 3D stereo images is to use two cameras set in a horizontal line simulating the human eye. The stereo image made with two cameras is the most reliable and direct way when compared to others because it does not need a complex calculation. All that is needed is controlling the cameras and making sure that the object and the cameras are set on the points of an isosceles triangle. The images made by this way already include the depth information. This method is already being used to make 3D stereo movies. Since two cameras are used to get the stereo images, this method cannot be used to get the depth information from the image or video of a single camera.

There are several ways to get the depth information from the image or video of a single camera. The utilization of database is one of them [15–17]. The depth map in the database is checked after the object is identified. The depth information of this method will be correct, if the match between the object and the database is correct. Using the vanishing point and the vanishing line to estimate the depth is another way [18]. When the edges of the scenery converge at the furthest point, the depth information can be defined by the point and the line. These are called the vanishing point and the vanishing line. Another way to estimate the depth information is using computed image depth (CID) method [19]. This method extracts information such as color, shadow, layers, clearness, contrast, object size, object overlap, and camera focus from the image to estimate image depth. The area of the image captured by the camera with correct focus will be clear in high contrast. In this study, the coarse depth information was first obtained by using CID method and then combined with the optical-flow method to estimate the fine depth information for the image sequence.

3. Coarse Depth from Single Image

Depth information can be determined from the arrangement of objects and the scene in the image shot by cameras. 2D display equipments such as the TV and movies are the most common display equipments recording only 2D information. However, there are some depth clues contained in the 2D

image. By analyzing the entire depth cue, the depth of objects in the scene can be estimated.

3.1. Computed Image Depth (CID). The computed image depth (CID) method is often used to estimate the depth information in a static image. The method utilizes image-specific information such as color, shadow, layer, clarity, contrast, object size, object overlap, and camera focus to analyze depth structure of the image properties. Although the accurate distance of objects in an image is difficult to derive from only a 2D image, the relative distance can be estimated from image clues, and coarse depth information can be generated. The image is divided into several image blocks, and the distance of the image block is estimated from the image clarity and contrast.

3.1.1. Clarity. In general, the clarity of an object provides an important clue of the distance from the viewer. This depends on several factors such as camera focus and illumination. The clearer the image object is, the nearer it is to the viewer. The image clarity can be computed by a *Laplacian* 3×3 matrix operation as in (1). The higher the magnitude of the *Laplacian* value is, the clearer the image block and the nearer it is. Then the image block can be classified into three distance levels according to clarity with (2),

$$L = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}, \quad (1)$$

$$f_s = \begin{cases} 1 \text{ (Far)}, & 0 \leq S < \frac{1}{3}S_{\max}, \\ 2 \text{ (Middle)}, & \frac{1}{3}S_{\max} \leq S \leq \frac{2}{3}S_{\max}, \\ 3 \text{ (Near)}, & \frac{2}{3}S_{\max} < S \leq S_{\max}, \end{cases} \quad (2)$$

where f_s denotes the depth value, S is the clarity of image block, and S_{\max} is the maximum clarity of all image blocks.

3.1.2. Contrast. When the statistical luminance distribution of image is centralized, the information content contained in the image is low. The image contrast measured by the amount of information gives an indication of how far the image object is from the viewer. The higher the contrast of the image object is, the nearer it is to the viewer. The image contrast can be computed by calculating the variance of image blocks as in (3), and the distance of each image block from the viewer can be estimated. Then the image block can be classified into three distance levels according to contrast with (4). Consider the following:

$$C_m(x, y) = \sqrt{\frac{1}{9} \left(\sum_{k=1}^3 \sum_{l=1}^3 (I(x+k-2, y+l-2) - \mu(x, y))^2 \right)}, \quad (3)$$

TABLE 1: Relationship between depth and distance.

Value	Distance
6	Near
4~5	Middle
2~3	Far

where $\mu(x, y) = (1/9)(\sum_{k=1}^3 \sum_{l=1}^3 I(x+k-2, y+l-2))$,

$$f_c = \begin{cases} 1 \text{ (Far)}, & 0 \leq C < \frac{1}{3}C_{\max}, \\ 2 \text{ (Middle)}, & \frac{1}{3}C_{\max} \leq C \leq \frac{2}{3}C_{\max}, \\ 3 \text{ (Near)}, & \frac{2}{3}C_{\max} < C \leq C_{\max}, \end{cases} \quad (4)$$

where f_c denotes the depth value, C is the contrast of image block, and C_{\max} is the maximum contrast of all image blocks.

3.1.3. Color. The depth estimated from the clearness and contrast is not so accurate that some errors will occur. Therefore, other information is added to make sure that the depth estimated with the CID method is closer to the true depth when combining clarity and contrast. Because the color of an image provides abundant information, this information can be used to correct errors in depth estimation. First, the color space is transformed from RGB to YCbCr, and then the objects are segmented by using Cb and Cr channels. Errors in the estimation of depth can be corrected by color segmentation, because the same object will have similar color at the same depth. The image blocks with the similar colors are merged from top to bottom until there are no new blocks to be merged. Then adjacent colors are merged if they are similar. Besides, general color concepts such as blue sky or green grassland are also used to correct errors in the estimation of depth.

3.2. Coarse Depth Map Generation. By applying the CID method, a coarse depth map can be roughly estimated from a single image. The overall operation is described as follows. First, the image is divided into many blocks (16×16 pixels), and the clarity and contrast of each block are calculated. Then each image block is classified according to three distance levels by the summation of the depth values calculated from clarity and contrast. Figure 1 is an example of estimated depth by combining the clarity and contrast. The relationship between depth and distance is shown in Table 1. Finally, the error in distance level is updated using color segmentation and general color as shown in Figure 2.

4. Fine Depth from Image Sequence

Image sequence contains more depth information than a single image [20]. Assuming that an object moves from left to right in image, the object with large motion distance appears to be closer than the object with small motion distance. This concept can be applied to image sequence as shown in

Figure 3. The object's motion distance between previous and current frame is used to estimate depth. The motion direction and luminance change of corresponding pixels between previous and current frame can be defined as optical flow [21, 22] or image flow.

4.1. Image Flow. Image flow is defined as the movement of pixels in the image plane when either the object or camera is moving. The difference between previous and current frame can be computed to generate instantaneous speed and motion. Figure 4 shows a point P with coordinate (X, Y, Z) project onto image plane to get pixel p with coordinate (x, y) . When P is moving with instantaneous speed \vec{S} , we can get the image flow $\vec{V} = (u, v) = (\Delta x/\Delta t, \Delta y/\Delta t)$ of pixel p in image plane.

4.2. Optical Flow. Optical flow is defined as the luminance variations of pixel due to illumination change or object movement. A static object projected on the image plane has no image-flow vector, but a nonzero optical-flow vector is obtained if illumination is changed. When an object is moving in image sequence under stable illumination, the image flow can be regarded as the optical flow.

Assume that an object is moving slowly while the illumination of the environment does not change, and then the luminance of pixels will be unchanged after being projected on the image plane. So the conservation equation preface of the luminance can be expressed as below:

$$E(x + \Delta x, y + \Delta y, t + \Delta t) = E(x, y, t), \quad (5)$$

where $E(x, y, t)$ denotes the luminance of pixel p of (x, y) on the image plane, Δx and Δy are displacements that p moves in the image plane after Δt time. Neglecting high-order items of the previous equation, we obtain

$$E(x, y, t) + E_x \Delta x + E_y \Delta y + E_t \Delta t = E(x, y, t), \quad (6)$$

where E_x , E_y , and E_t denote the luminance of pixel $p(x, y)$ with deviation to x , y direction and time t . By eliminating the term $E(x, y, t)$, we obtain

$$E_x \Delta x + E_y \Delta y + E_t \Delta t = 0. \quad (7)$$

Then the following luminance change equation is obtained by dividing it by Δt ,

$$E_x u + E_y v + E_t = 0, \quad (8)$$

where $u = \Delta x/\Delta t$ and $v = \Delta y/\Delta t$.

Horn and Schunck [23] proposed a first-order differential calculation method for E_x , E_y , and E_t without recursive calculations as follows:

$$E_x \approx \frac{1}{4} (E_{i,j+1,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i+1,j,k+1}),$$

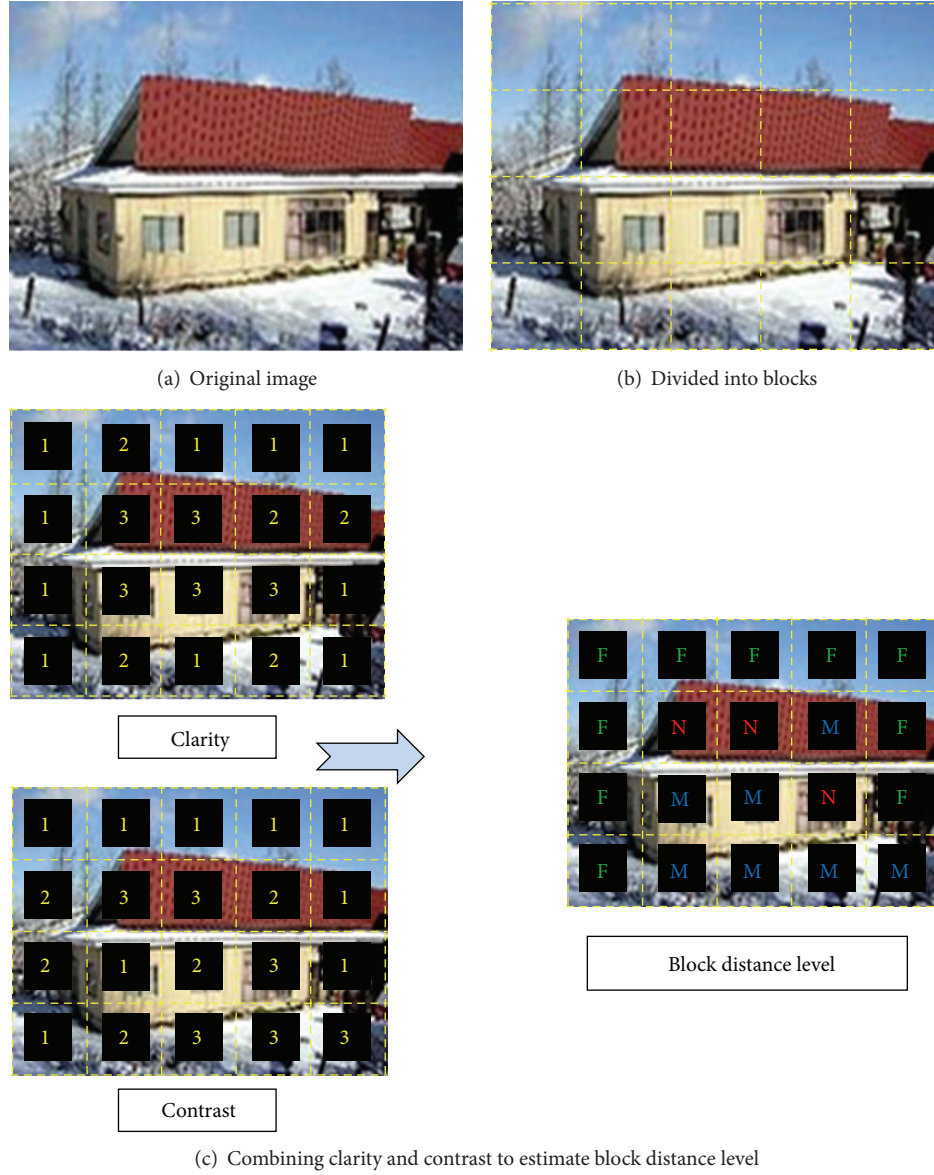


FIGURE 1: Distance of image block by calculating depth value based on clarity and contrast.

$$\begin{aligned}
 E_y &\approx \frac{1}{4} (E_{i,j+1,k} - E_{i+1,j+1,k} + E_{i,j,k} - E_{i+1,j,k} \\
 &\quad + E_{i,j+1,k+1} - E_{i+1,j+1,k+1} + E_{i,j,k+1} - E_{i+1,j,k+1}), \\
 E_t &\approx \frac{1}{4} (E_{i,j,k+1} - E_{i,j,k} + E_{i+1,j,k+1} - E_{i+1,j,k} \\
 &\quad + E_{i,j+1,k+1} - E_{i,j+1,k} + E_{i+1,j+1,k+1} - E_{i+1,j+1,k}),
 \end{aligned} \tag{9}$$

where E_x and E_y are the first-order partial differential of the x and y directions. The gradient ∇ is the sum of the absolute values of E_x and E_y . The difference image $D(x, y)$ is the absolute value of E_t , where E_t is the difference between two

consecutive images. Figures 5(a) and 5(b) show the gradient and difference image calculated by

$$\begin{aligned}
 \nabla &= |E_x| + |E_y|, \\
 D(x, y) &= |E_t(x, y)|.
 \end{aligned} \tag{10}$$

Because the optical-flow difference image as shown in Figure 5(b) is calculated from both the previous and the current frame, it can be enhanced for further segmentation. In order to enhance the difference image, the motion history image (MHI) is calculated by adding a half of the motion history image previously made to the difference image, see (11). Figure 5(c) shows the MHI image obtained by (11). Finally, the search region of object image can be defined by

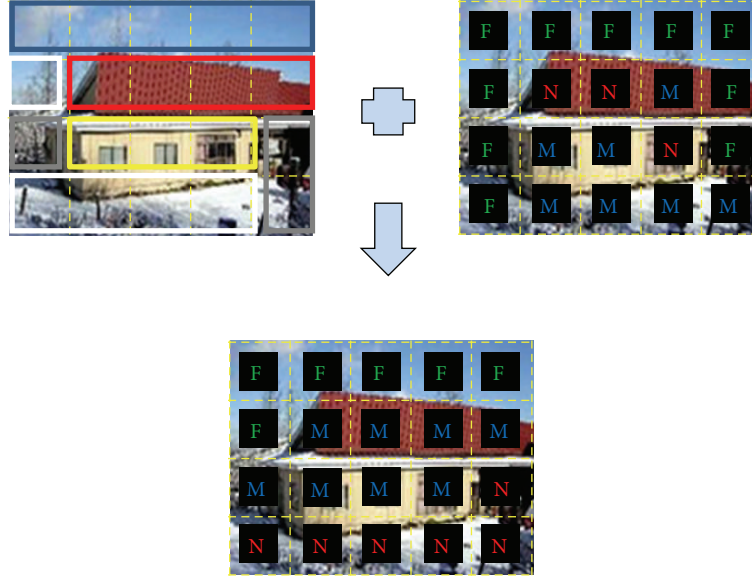


FIGURE 2: Distance error updated using segmentation and general color.

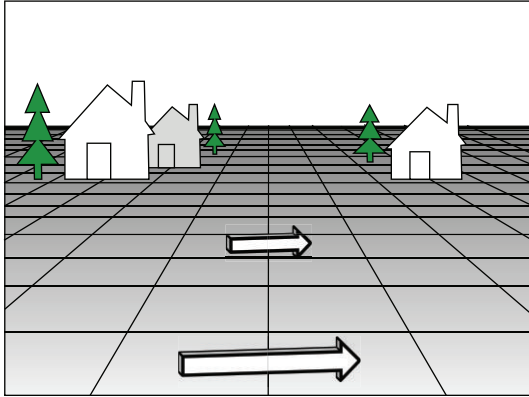


FIGURE 3: Relationship between distance and motion.

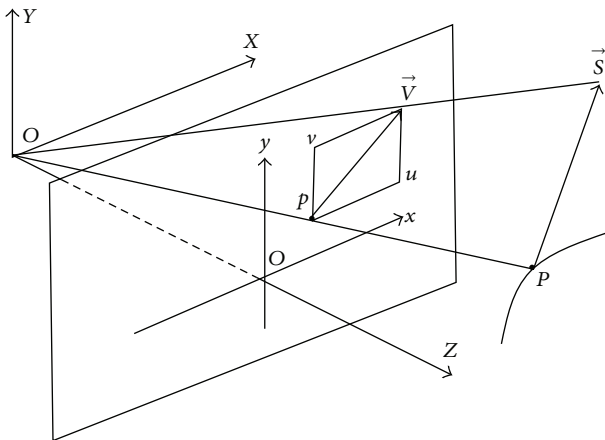


FIGURE 4: Illustration of image flow.

AND operation of the MHI and the gradient image as shown in Figure 5(d). Consider the following:

$$M(x, y) = D(x, y) + \frac{M(x, y)}{2}. \quad (11)$$

4.3. Object Segmentation. After determining the search region of an object, the binary image of an object can first be obtained by a simple threshold method; see Figure 6(a). Because the coarse depth estimated with the CID method is calculated on a block basis, the block image of an object can be obtained from a binary image as shown in Figure 6(b). The calculation of block distance level might be wrong in CID due to the object's texture. The block distance should be the same as or similar to one adjacent to it because they belong to the same object. So distance level in the CID method has to be corrected by using object segmentation. Because the same object usually has the same optical-flow value, the distance level can be updated by using the highest frequency in the optical-flow blocks. Figure 7 shows the experimental results after the distance level has been corrected according to object segmentation by the optical-flow method.

4.4. Background Segmentation. Background is the deepest area in the image. If a segment of the background area can be effectively detected, then the depth information can be evaluated more accurately. A true and more complete depth map can be acquired with accurate background information.

First, the RGB color information is down sampled to 3 bins per each channel, see (12). Then RGB color space is down sampled to 27 bins. Figure 8(a) shows the experimental results after down sampling. Because the background area usually appears on the top of the image, for example, the sky, the color histogram for top of the image can be recorded,

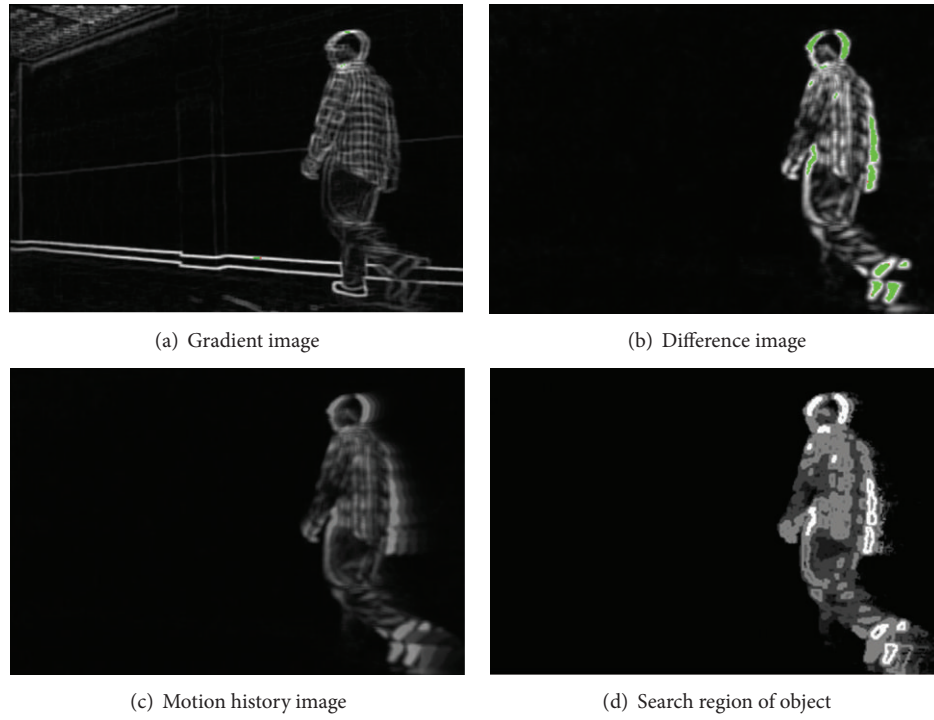


FIGURE 5: Moving object detection with optical-flow method.

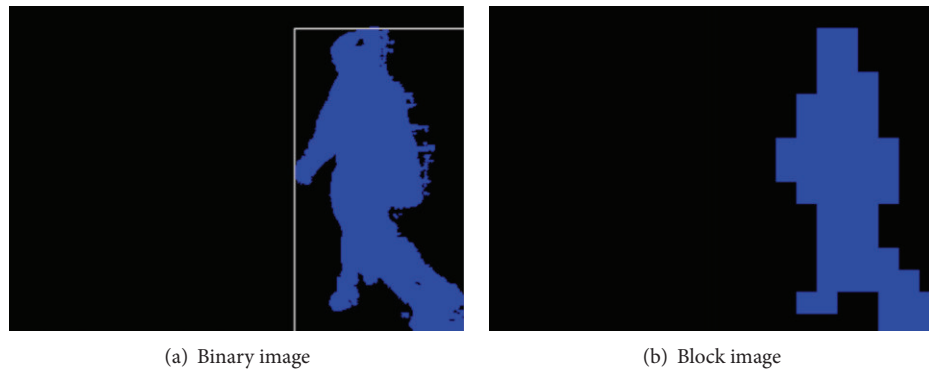


FIGURE 6: Object segmentation with the optical-flow method.

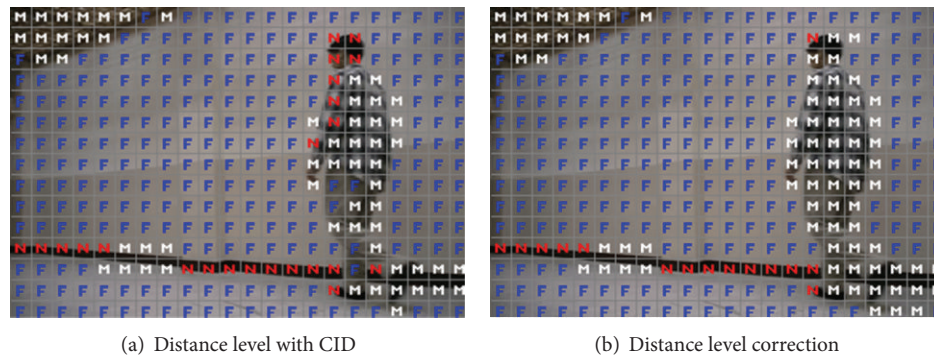


FIGURE 7: Distance level correction according to object segmentation using the optical flow method.

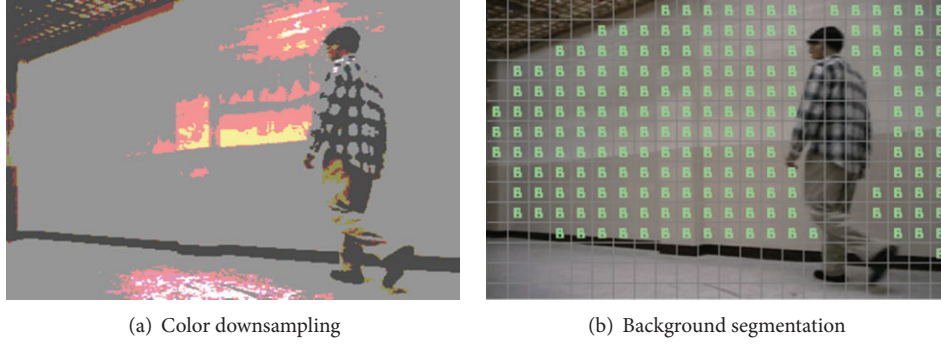


FIGURE 8: Illustration of background segmentation.

and then the color with the highest frequency can be set as the background color. Then the same colors are merged from top to bottom until there are no new blocks to be merged. The colors adjacent to each other are checked and merge if they have the same color information. Figure 8(b) shows experimental results after background segmentation

$$R = \begin{cases} 75, & 0 < R \leq 75 \\ 150, & 75 < R \leq 150 \\ 255, & 150 < R \leq 255 \end{cases} \quad \text{same for } G \text{ and } B \text{ channel.} \quad (12)$$

After background segmentation, the background is set to the deepest area in the image when creating the 3D stereo image.

4.5. Binocular Image from Optical-Flow Vector. Considering a point in spatial space, one can see it as point A with the left eye but as point B with the right eye due to the different viewing angles. In Figure 9, if the point is moving in space, then point A seems to be moving to C using the left eye and B moving to D using the right eye. While A is moving to C through B, A and B will have the time difference Δt and the same with C and D. If point A is at time t , and C at time $t + 1$, then B and D will be at time $t + \Delta t$ and $t + \Delta t + 1$. Projecting A and C on the image plane, pixels a and c are obtained in the left eye image. By moving pixel a to b and c to d in the left eye image, pixel b' and d' are obtained in the right-eye image. With this procedure, a binocular image can be created.

Two frames are used to substantiate this idea. First, the previous frame is used as the left-eye image, and then the optical-flow vector of the same point in the current frame is calculated. The shift of the pixel can be defined from the optical-flow vector according to the depth ratio. Then the right-eye image can be created according to the pixel shift in the current frame. When the motion is not large enough, the pixel shift calculated by the depth ratio will be small. This causes an unacceptable stereo effect. In order to improve the stereo effect, the pixel shift is subtracted from the previous frame to create the left-eye frame, and the pixel shift from the current frame is added to create a double depth distance in the right-eye frame as shown in Figure 10.

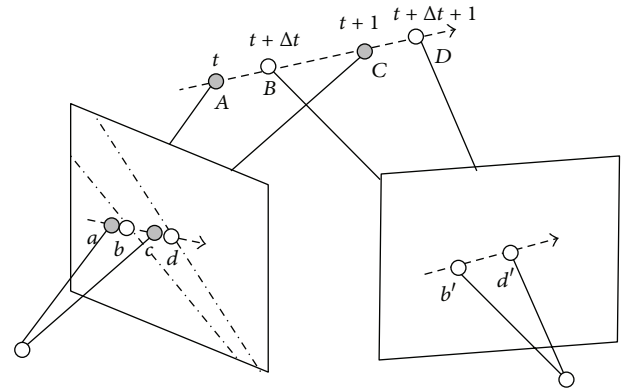


FIGURE 9: Demonstration of time difference of moving point.

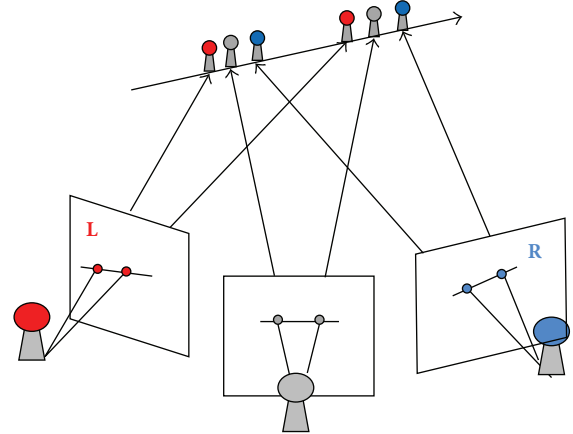


FIGURE 10: Illustration of binocular image pair creation.

The optical-flow method cannot fully define the distance between objects. So the CID method and the optical-flow method are combined to improve the depth estimation and recovery. The pixel shift value Δs is calculated as in Table 2. Because of the horizontal parallax, the 3D stereo effect can only be perceived with a horizontal pixel shift based on the depth information, so the vertical parallax is fixed to $(y_1 - y_2)/4$.

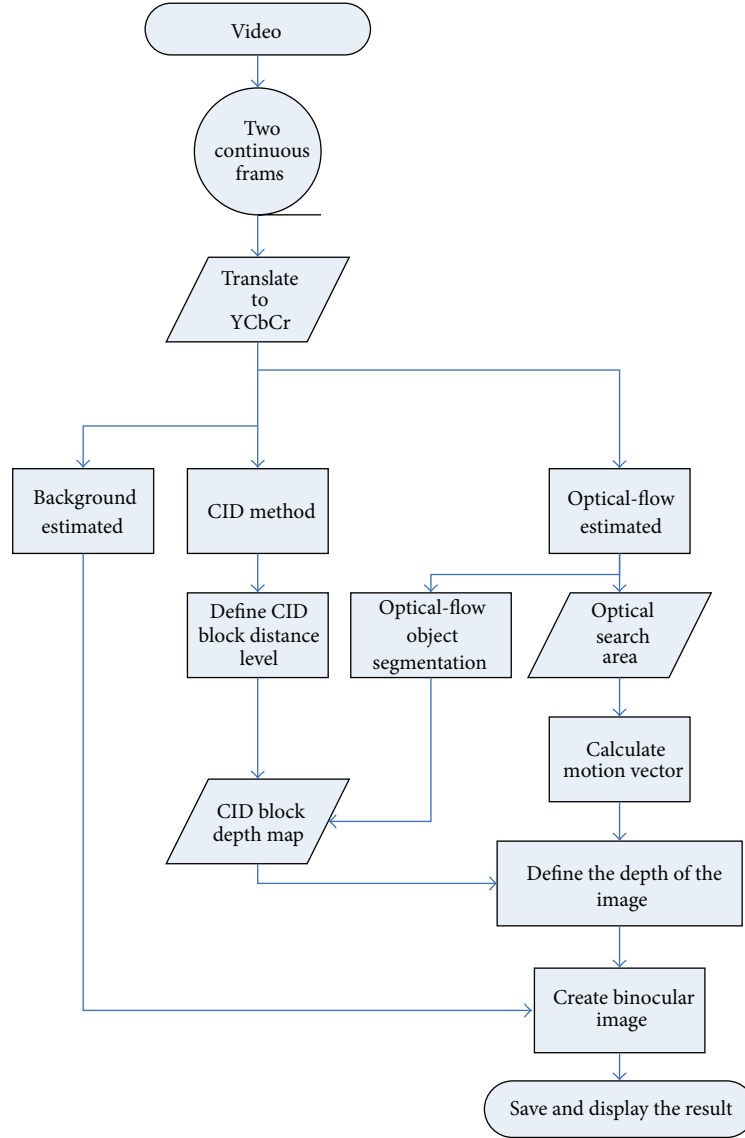


FIGURE 11: Flow chart of the proposed system.

TABLE 2: Relationship between distance level and pixel shift value.

CID distance level	Δs
Far	$(x_1 - x_2)/8$
Middle	$(x_1 - x_2)/4$
Near	$(x_1 - x_2)/2$

5. Experimental Results

The proposed overall system flow diagram is shown in Figure 11. Figure 12 is an example of two frames of an image sequence. The image is first converted to Y , C_r , and C_b spaces and then the luminance channel is used to compute the image depth and optical-flow vector. Figure 13 shows the experimental results by computing image depth according to clarity

and contrast of the image. Figure 14 depicts the estimated depth (map) computed from the optical-flow vector and modified with the CID method. Before creating the binocular image, the background area is set to the furthest from the foreground in order to create the best depth structure in the stereo image. To create the 3D stereo binocular image, the pixel shift is subtracted from the previous frame to create the left-eye image, and the pixel shift from the current frame is added to create the right-eye image based on the estimated depth map. Figure 15 shows the left- and right-eye images of the binocular image. By integrating the left- and right-eye images into an interlaced image and displaying it in a 3D display system, the 3D stereo effect can be perceived without wearing glasses.

In order to prove the feasibility of the proposed method, many famous 2D movies have been converted into 3D

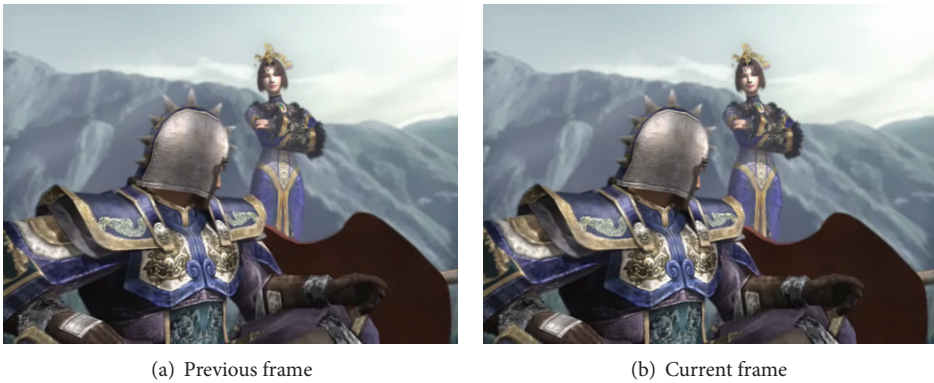


FIGURE 12: Two consecutive frames from image sequence.

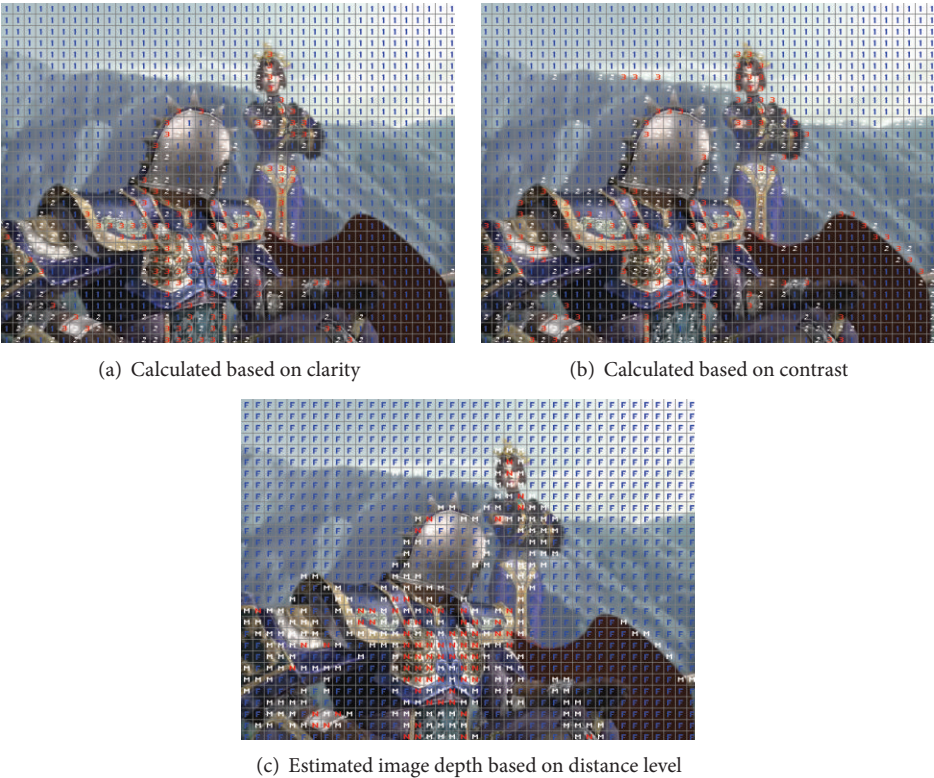


FIGURE 13: Experimental result using the computed image depth method.

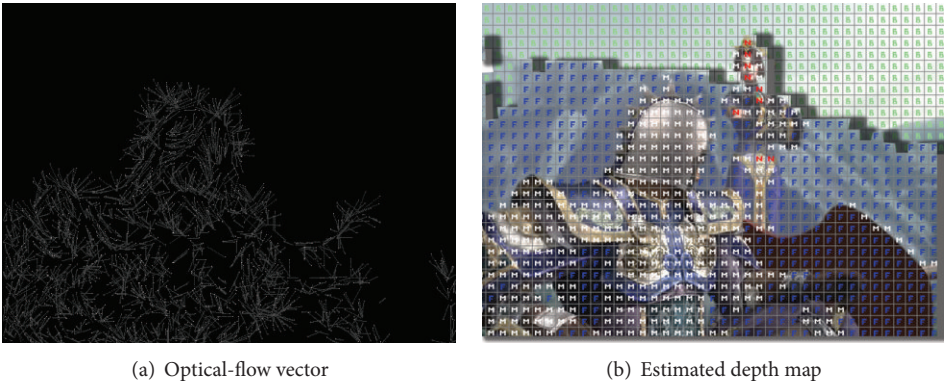


FIGURE 14: Estimated depth map computed from the optical-flow vector and modified with the CID method.



FIGURE 15: Binocular image creation.

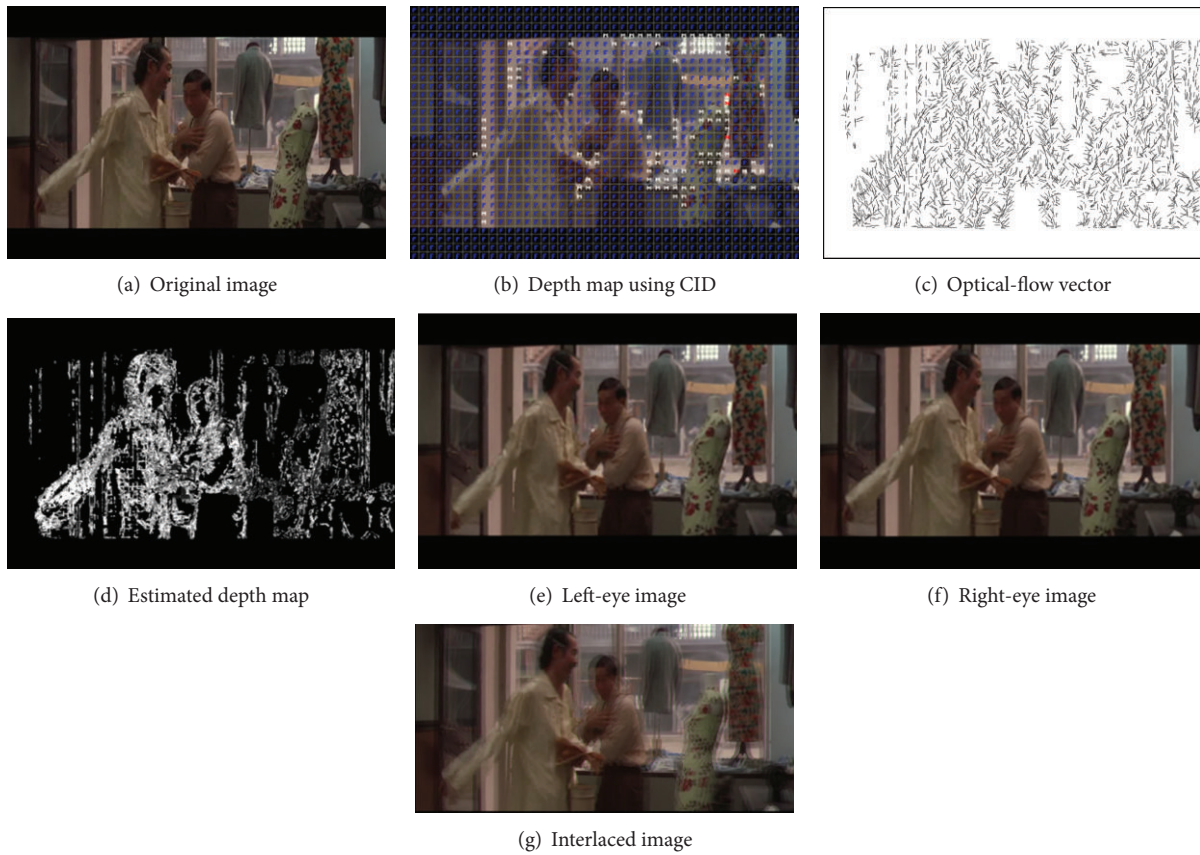


FIGURE 16: Example of 3D stereo binocular image creation from a 2D movie.

stereo movies. By viewing it on 3D auto-stereoscopic display equipment, 3D stereo can be experienced without wearing 3D stereo glasses. Figure 16 shows an example of 3D stereo binocular image creation from a 2D movie.

6. Conclusions and Future Works

This paper presents a method for creating a 3D stereo video by using an enhanced CID method combined with the optical-flow method. The static depth structure is analyzed; then the optical-flow method is used to determine the motion

distance between current and previous frames. The image depth can be estimated by merging static and dynamic depth cues and using the optical-flow recovery method to generate a binocular view. We do not need any prior information of camera parameter and depth map as with other methods. All we need is a monocular image sequence. After generating binocular image sequences, they can be viewed on 3D stereo display equipment and have a 3D experience.

In the experimental result, the effect was one of a background area being displayed behind the screen and the foreground area in front of screen. There is depth between

the foreground and background. With large moving areas, a multilayer depth is perceived, not just a single depth. The whole scene gives a satisfactory 3D experience. However, the 3D effect is not as good as the image sequences shot with a two-camera system. In order to improve the effect of the 3D stereo experience, further research need to be done. The following three directions are suggested for this.

- (1) Some image denoising algorithms [24] can be used to improve the accuracy of estimated depth information.
- (2) Because there is such a variety in the types of videos, it is difficult to obtain a common structure of the depth map. For true depth information, all research in scene analysis needs to be investigated. Creating several different scene modes is the most common approach. Some scene features such as affine moment [25] can be used to classify the different scene modes. By using different models to match the different scenes, the optimum depth information can be obtained.
- (3) Each kind of 3D display equipment has its own display theory and hardware specification with the result that there are different effects on different display equipment. The display effect is restricted by the software provided by the vendor. For better display quality and 3D stereo effect, the specifications of the display equipment need to be addressed, and the software must be matched for the proposed 3D stereo binocular image creation algorithm.

Acknowledgment

The authors would like to thank the National Science Council of Taiwan for the support of this research with Grant no. NSC101-2221-E-216-034.

References

- [1] C. Wheatstone, "Contributions to the physiology of vision. Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.
- [2] D. Brewster, *The Stereoscope*, Rudolf Kingslake, London, UK, 1971.
- [3] Anderton, U.S. Patent No. 542321, 1895.
- [4] L. Blondé, "3D stereo rendering challenges and techniques," in *Proceedings of the Conference on Information Sciences and Systems*, pp. 17–19, Princeton, NJ, USA, March 2010.
- [5] P. Merkle, K. Müller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, 2010.
- [6] Y. Yan, "A novel method for automatic 2D-to-3D video," in *Proceedings of the 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp. 7–9, June 2010.
- [7] L. Zhang, C. Vázquez, and S. Knorr, "3D-TV content creation: automatic 2D-to-3D video conversion," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 372–383, 2011.
- [8] Z. Zhang, "A compact stereoscopic video representation for 3D video generation and coding," in *Proceedings of the Data Compression Conference (DCC '12)*, pp. 189–198, Beijing, China, April 2012.
- [9] A. H. Wang, Q. H. Wang, X. F. Li, and D. H. Li, "Combined lenticular Lens for autostereoscopic three dimensional display," *Optik—International Journal for Light and Electron Optics*, vol. 123, no. 9, pp. 827–830, 2012.
- [10] J. Lee, "The 3D video processing method in the stereoscopic camera for mobile devices," in *Proceedings of the IEEE International Conference on Emerging Signal Processing Applications (ESPA)*, pp. 139–142, January 2012.
- [11] G. S. Lin, C. Y. Yeh, W. C. Chen, and W. N. Lie, "A 2D to 3D conversion scheme based on depth cues analysis for MPEG videos," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '10)*, pp. 1141–1145, Suntec City, Singapore, July 2010.
- [12] H. M. Wang, C. H. Huang, and J. F. Yang, "Depth maps interpolation from existing pairs of keyframes and depth maps for 3D video generation," in *Proceedings of the IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (ISCAS '10)*, pp. 3248–3251, June 2010.
- [13] P. J. Lee and E. Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 246–254, 2011.
- [14] J. Choi, "Sparsity based depth estimation and hole-filling algorithm for 2D to 3D video conversion," in *Proceedings of the International Conference on Signals and Electronic Systems (ICSES '12)*, pp. 1–4, Wroclaw, Poland, September 2012.
- [15] T. Hassner and R. Basri, "Example based 3D reconstruction from single 2D images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006.
- [16] K. C. Liu, Q. Wu, W. C. Chen, C. F. Wu, F. C. Jan, and T. Chen, "Example-based depth generation from single image for 3D content," in *Proceedings of the 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON '08)*, pp. 333–336, Istanbul, Turkey, May 2008.
- [17] Z. Wang, "Depth template based 2D-to-3D video conversion and coding system," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 308–313, Melbourne, VIC, USA, July 2012.
- [18] S. Battiato, S. Curti, M. la Cascia, M. Tortora, and E. Scordato, "Depth-map generation by image classification," in *Proceedings of the Three-Dimensional Image Capture and Applications VI*, vol. 5302, pp. 95–104, January 2004.
- [19] S. Okada, "32. 2: a real-time 2D to 3D conversion technique using computed image depth," *SID Digest of Technique Papers*, vol. 29, no. 1, pp. 919–922, 1998.
- [20] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.
- [21] M. B. Holte, T. B. Moeslund, and P. Fihl, "View-invariant gesture recognition using 3D optical flow and harmonic motion context," *Computer Vision and Image Understanding*, vol. 114, no. 12, pp. 1353–1361, 2010.
- [22] Y. Yang, Q. Liu, R. Ji, and Y. Gao, "Dynamic 3D scene depth reconstruction via optical flow field rectification," *Plos One*, vol. 7, no. 11, article e47041, 2012.
- [23] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [24] X. Zheng, Z. Liao, S. Hu, M. Li, and J. L. Zhou, "Improving spatial adaptivity of nonlocal means in low-dosed CT imaging

using pointwise fractal dimension,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 902148, 8 pages, 2013.

- [25] J. Yang, M. Li, Z. Chen, and Y. Chen, “Cutting affine moment invariants,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 928161, 12 pages, 2012.

