Journal of the Optical Society of Korea, Vol. 8, No. 3, September 2004, pp. 108-114

Comparison of Irregular Quadtree Decomposition with Full-search Block Matching for Synthesizing Intermediate Images

Kyung-tae Kim*

School of Information & Communication, Hannam University, Daejeon, 306-791, KOREA

(Received May 20, 2004)

To allow multiple viewers to see the correct perspective and to provide a single viewer with motion parallax cues during head movement, more than two views are needed. Since it is prohibitive to acquire, process, and transmit a continuum of views, it would be preferable to acquire only a minimal set of views and to synthesize intermediate images. This paper presents how to synthesize the intermediate images using irregular quadtree decomposition and compares the proposed methods with full-search block matching. The image at the middle viewpoint between both viewpoints is synthesized and yields a 32.8 dB peak-signal-to-noise ratio (PSNR) which is 2.8 dB high and has a running time 30% of that for conventional full-search block matching.

OCIS codes : 100.6890, 110.4190

I. INTRODUCTION

Three dimensional (3D) display systems can be used in communication systems, education, medical science, and commercial advertisements. To be able to develop good 3D display systems, two conditions must be fulfilled: first, systems need to be realized by the current hardware capacity, and second, the systems need to become familiar to the human beings who are not used to seeing 3D images. Namely, 3D display systems must not make people tired, and they need to have wide viewing zones.

To allow multiple viewers to see the correct perspective and to provide a single viewer with motion parallax cues during head movement, more than two views are needed. Consequently, it requires larger and faster memory, networks, and computer systems. Since it is prohibitive to acquire, process, and transmit a continuum of views, it would be preferable to acquire only a minimal set of views and to use an estimated disparity map to synthesize intermediate images. Furthermore, the quality of the synthesized image depends on the accuracy of the estimated disparity map and how occlusions are handled.

The synthesizing of intermediate images can be defined as follows: by using two stereo images, intermediate images can be synthesized at any point between them. If a high quality of synthesized intermediate images can be obtained, then the systems can be simplified because there is a lesser amount of data to send and save.

Many researchers, including me, have worked on synthesizing intermediate images [1-4]. M. Fujii reconstructed intermediate view images from two images [5].

In this paper, the synthesizing of intermediate images using irregular quadtree decomposition for multiple views will be described and compared with traditional full-search block matching from the viewpoints of accuracy and computation time.

A set of cameras on a line, with equal distances between them, is used to acquire the multiple views. The images are preprocessed by means of a contrast stretch, histogram equalization, sharpening, and edge detection by the Canny algorithm. Irregular quadtree decomposition is used to align the block boundary which is the disparity discontinuity. Block boundaries are computed by locating the dominant horizontal and vertical edges.

To find the corresponding points between two images, a confidence measure of the accuracy of each disparity estimate is used. The single step disparity vector estimate associated with a region whose prediction similarity falls below a pre-defined threshold is assumed to be erroneous. The uniqueness constraint is also used, in other words, each pixel can have at most one match.

The next step is to detect and fill the occlusion regions that do not appear in either image. Regions of occlusion are decided by similarity comparisons among the matched block alternatives. I have been working on how to fill the occlusion regions of intermediate images and found some results [6].

Finally, a comparison of the proposed method with the full-search block matching method which is conventionally used is carried out. The image at the middle viewpoint between both viewpoints is synthesized and yields a 32.8 dB PSNR (peak-signal-to-noise ratio) which is 2.8 dB high and has a running time 30% of that for the conventional full-search block matching.

II. IRREGULAR QUADTREE DECOMPOSITION

Before irregular quadtree decomposition, the left and right images are preprocessed. In the first preprocessing step, the mean and variance of all pixels in one image are made equal to those of all pixels in the other image. This processing is for reducing the differences in luminance of the environments and image sensors of each camera, and each image is normalized to counteract the residual luminance and exposure differences in the images.

For detecting the boundaries of the image, the Canny operator [7] is applied to the images. The image is transformed by a thinning process into lines of pixels by the Hilditch procedure [8] and then the region surrounded by a certain area less than threshold is removed by a Freeman chain code [9], which is the most widely used method for lossless coding of line drawings.

The primary object of irregular quadtree decomposition is to align the block boundary with the boundary of the attribute that is used as the splitting criterion. This boundary is the disparity discontinuity. A disparity discontinuity which arises from an object boundary typically lies at an image intensity discontinuity (edge). In the absence of a disparity map, the edges in an image provide the most practical candidates as locations for partitioning. Because the image which is overlapped by the two edge images of the left and right original images is used, more detail can be detected and divided in the common block in both images. An algorithm to locate the dominant vertical and horizontal edge is needed. The effect of local details and noise are averaged out and the dominant edges along the horizontal and vertical directions become emphasized in the row and column averages. The dominant vertical and horizontal edge locating algorithm makes the shape of the block rectangular, not any other shape. The size of the rectangular block depends on the complexity of the image. Generally the edges representing the boundary of the object do not constrain the direction to vertical or horizontal only. Then, it may be better that the shape of the block is a polygon, representing the boundary of the object, and ultimately the edge of the object itself. If various polygons are used, the matching process becomes very complex and it may take a very long running time. It cannot be said that a polygon in one image must also be present in the other image. A rectangular block can also have almost the same function as the polygon, because a rectangular block will be divided into very detailed blocks when the edge in it is complex.

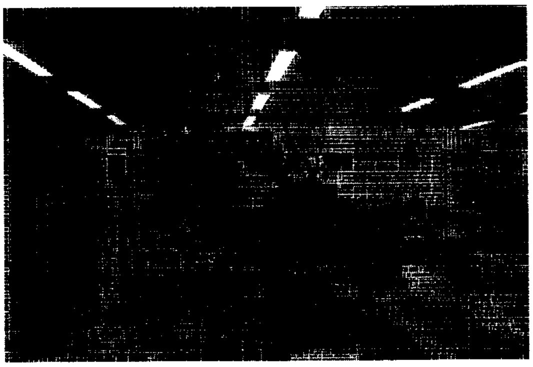


FIG. 1. Irregular quadtree decomposition image by block dividing.

A symmetric difference high pass filter is applied to the row and column averages. By finding the peak over the absolute values of the filter outputs, the horizontal and vertical boundaries are computed.

Typical high pass filters of order n=1 or n=2 (specially, if n=1, f=[-1, 0, 1]; and if n=2, f=[-1, -2, 0, 2, 1]) are used [4]. A large n provides a more reliable edge location by smoothing out local variations, but reduces the number of candidate boundaries due to edge effects. Figure 1 shows the irregular quadtree decomposition image by block division.

III. DISPARITY BY BLOCK MATCHING

The cameras used to capture stereo-pairs are separated by an accurately known horizontal distance, and they are arranged in a parallel configuration. A point in the object space synthesizes corresponding points in the left- and right-eye images. If the object point is unoccluded, i.e., if it is visible in both images, the disparity is defined as the distance in pixels between the image point in both images; otherwise it is undefined. In the parallel camera axis geometry, the disparity is entirely in the horizontal direction.

The disparity vector estimate associated with a region whose prediction similarity- the mean of the sum of the absolute difference within the block - falls below a user-defined threshold is assumed to be erroneous, and is discarded from any further consideration. The prediction similarity for a particular disparity estimate, where the statistics of the similarity are governed by whether the estimate is either correct or incorrect, is observed. The threshold then is based on some knowledge of these statistics and the relative costs of false alarms and misses. Once thresholded, the false estimates can be further eliminated by using the one-to-one relationship between the true disparities of corresponding points from the left image to the right image, and vice versa. If the estimated disparity field is inconsistent with both directions, it is assumed that an error has occurred, and the likely false estimate is eliminated.

Given a single-step disparity vector field in one direction, the usage of the pixels in the reference image is examined. If a reference pixel is used for the prediction of only one pixel in the desired image, it is assumed that this disparity estimate is accurate, and the reversed disparity vector is assigned too.

If a pixel is referenced by more than one disparity estimate, it is assumed that all of these estimates cannot be correct. Again for the decision as to which of these estimates has the lowest probability of error the prediction similarity is used. The disparity estimate for the reference pixel is assigned as the negative of the forward disparity vector with maximum prediction similarity, and all other vectors are invalidated. Finally, if a pixel in the reference image is not used in the prediction, it is assumed that this pixel is occluded and does not assign a reversed disparity estimate to it.

IV. INTERPOLATION

1. Corresponding points

Generating subjectively-pleasing intermediate views from two images should be an important concern. It is required that the given views have been captured by cameras whose offset is in only their horizontal and/or vertical position.

Improved depth perception is achieved by presenting the appropriate image (or pair of images for a binocular system) based on the viewer's location. Intermediate views also can be used to ensure that stereoscopic imagery is comfortable to view. Discomfort is often experienced when viewing stereoscopic images on a two-dimensional display device. The breakdown of the accommodation/convergence relationship has been widely reported as a cause of this discomfort [10]. An additional source of eye strain is related to the viewer's ability to fuse binocular information.

To achieve geometrically correct stereo-vision, in the sense that the viewer sees what would be seen by the naked-eye, the camera capturing the binocular image pair must be separated by the viewer's interocular separation. However, this quantity is viewer dependent and individuals often prefer varying degrees of depth perception based on individual stereoscopic viewing ability and the range of depth present in the scene. A greater sense of depth is provided by a relatively large inter-camera separation, but the larger the separation the more difficulty a marginal viewer has in fusing the image. If a continuum of views between two extreme viewpoints is available, a viewer can dynamically select the inter-camera separation for comfort and preferred sense of depth in a manner similar to the adjustments of brightness and contrast found on most display devices. Decreasing the camera separation also reduces the breakdown between the accommodation/convergence relationships.

Since infinitesimal camera spacing is required for the motion parallax and viewer-specified inter-camera separation applications, it is impractical to capture all of these views.

The actual interpolation from the complete and the interpolated disparity fields are performed. Pixels that are referenced by one or more valid vectors are assumed to be unoccluded, and are predicted from a weighted average of the corresponding pixels in the two images:

$$M(x,y,\beta) = (1-\beta)\cdot F(p_L) + \beta \cdot G(p_R) \tag{1}$$

$$p_L = \begin{bmatrix} x - \hat{d}_{x_{\beta\leftarrow L}}(x,y) \\ y - \hat{d}_{y_{\beta\leftarrow L}}(x,y) \end{bmatrix}, \quad \text{and} \quad p_R = \begin{bmatrix} x - \hat{d}_{x_{\beta\leftarrow R}}(x,y) \\ y - \hat{d}_{y_{\beta\leftarrow R}}(x,y) \end{bmatrix} \tag{2}$$

Here p_L and p_R respectively denote the pixel locations in view L and R. The intermediate image M is parameterized by its relative position between the given images L and R by β, where $0 \leq \beta \leq 1$. The intensities for fractional pixel locations are obtained through bilinear interpolation of the reference view. Averaging pixels from both views for unoccluded regions reduces the effect of high frequency noise components in the interpolation process, and blends the intensity variations between the two exterior views.

2. Occlusion points

It is assumed that a real-world object lies completely within a camera's field-of-view and is located between the camera and some scene background. An occlusion is caused by the foreground object if and only if the camera viewpoint changes or the object moves with respect to the background between images. For either case, the displacement of the foreground object differs from that of the scene background, i.e., a displacement discontinuity exits.

There are a variety of ways in which occlusion could be compensated. Figure 2 shows how to fill the intermediate image that does not have corresponding points between two images, i.e., occlusion regions. Four kinds of occlusions that can occur due to the placement of

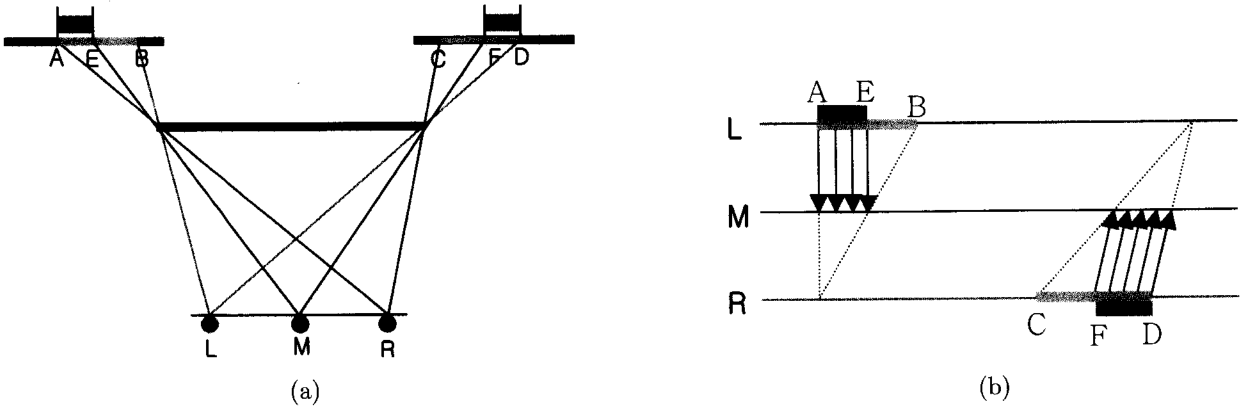


FIG. 2. How to fill the intermediate image for occlusion regions, the region AE and FD are ones in middle image to be synthesized. (a) configuration of real-world objects, and their regions appearing in the left, right and middle eyes, respectively, (b) one-dimensional images corresponding to (a).

objects are analyzed, and Fig. 2 (a) is one of them. Figure 2 (a) shows the configuration of real-world objects, and their regions that appear in the left, right and middle eyes, respectively, and Fig. 2 (b) shows a one-dimensional image corresponding to Fig. 2 (a). The regions of AB and CD appear in the left and right eyes (images) only. Region AE, which is on the left side of the occlusion region of the left image, appears in the middle image. In the same manner, region FD, which is on the right side of the occlusion region of the right image, appears in the middle image. As a result of this analysis, it is concluded that when the left image has occlusion regions, the corresponding region of the intermediate image is filled with the left part of the occlusion region of the left image; and when the right image has occlusion regions, the corresponding region of intermediate image is filled with the right part of the occlusion region of the right image.

V. FULL-SEARCH BLOCK MATCHING

The full-search block matching algorithm is conventionally used for finding the corresponding points with pixel units. The corresponding point between both images is found by the block information which includes the neighboring pixels of the observed ones. Each block within a given search window is compared to the current block and the best match is obtained with the comparison criterion, which is the mean absolute difference, MAD, for all positions. Generally, this algorithm is the best one in terms of the image quality and the simplicity of the algorithm, although it is very computationally intensive. While observed pixels are moved in some search range, it is assumed that all observed pixels must be in the whole image. The best matching region with respect to the MAD is always found. The computational load of this full-search algorithm is very demanding, and therefore many different fast algorithms have been developed. All of them try to reduce the computational load without, or with minor, loss of prediction error gain. An often-used approach is to apply a special search strategy, and therefore reduce the number of examined search positions.

VI. EXPERIMENTS AND RESULTS

An EOS-1 camera, manufactured by Cannon and with the Cannon lens EF 24 mm 1:1.4, is carefully aligned and is slid horizontally to avoid vertical disparity on the camera mount, which is manufactured very precisely. The experiments are performed with three images taken in an experimental room where a mannequin and some stuffed animal dolls are placed. Figure 3 shows the left, middle and right images for typical perspective input. Resolution of each image is 676 lines and 1016 pixels, and each pixel consists of 24 bits of color. The left and right cameras are 4 cm apart, and the objects are 80 cm to 1,000 cm from the camera. The maximum disparity is about 40 pixels for the nearest one of the objects (books). Lighting and exposure are carefully adjusted to match the luminance of the three images. Image data is first processed to an equal mean and variance between two images, and the histogram stretched to an intensity range of 0 to 255 for reducing the differences of brightness and contrast.

The intermediate image between both left and right images is synthesized by irregular quadtree decomposition and an exhaustive full-search block matching algorithms, and these two algorithms are compared with the items of running time and image quality (PSNR). The block size used for the exhaustive full-search block matching is 5 pixels by 5 pixels.

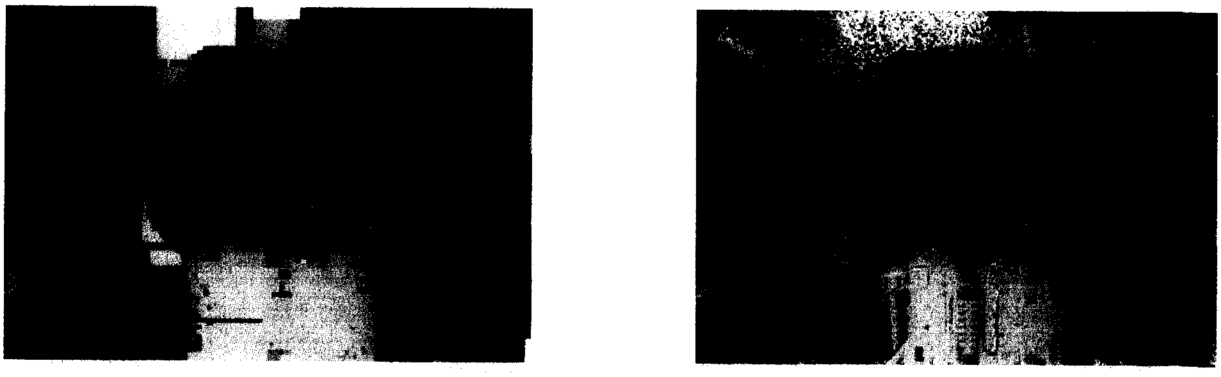
For dividing the irregular quadtree decomposition, edge images of the left and right images are obtained using the Canny algorithm. The edges obtained are thinned and small contours of edges with the threshold of 50 are removed by a Freeman chain code to reduce the unnecessary edges. Then the overlapping image of both edge images is used for dividing the block. The reason for using the overlap of both edge images is to synthesize the block so that block size of the input is smaller than the corresponding block in the reference image.

The high pass filter for finding the boundary whose coefficients are [-2, -1, 0, 1, 2] is applied to the row and column averages to detect the peak. The threshold of variance for block dividing is set to 100, and the maximum threshold of variance of the block is set to 13,000, therefore a block is divided into four or two horizontally or two vertically, according to the condition of the block when the variance has a value of between 100 and 13,000. The maximum threshold of variance is to avoid an unnecessary boundary of a too complicated image. The size of the block is set to 5, 100, 300 for minimum width and height, maximum width of x, and maximum width of y, respectively. The minimum value of 5 prevents the formation of extremely small blocks and also accounts for the inaccuracy at the edge of a block while using the symmetric high pass filter. The difference between the maximum width of the x and y directions is the reason why the variation of the image in the y direction may be less than that in the x direction.

Each block searches the corresponding block on the reference image in the range of -40 to 40 pixels in the x direction, but does not search in the y direction because of the assumption that the image has no vertical disparity owing to a well aligned camera mount. As a criterion for measuring the similarity between two images, the mean of the sum of the absolute difference (MAD) within the block is used. The processing of thresholding is carried out to meet the constraint of uniqueness (one-to-one matching) in which the threshold of MAD is 22. Actually, the value of MAD was not very sensitive, that is, the PSNR representing the quality of the image was almost the same while the MAD varies from 20 to 200.

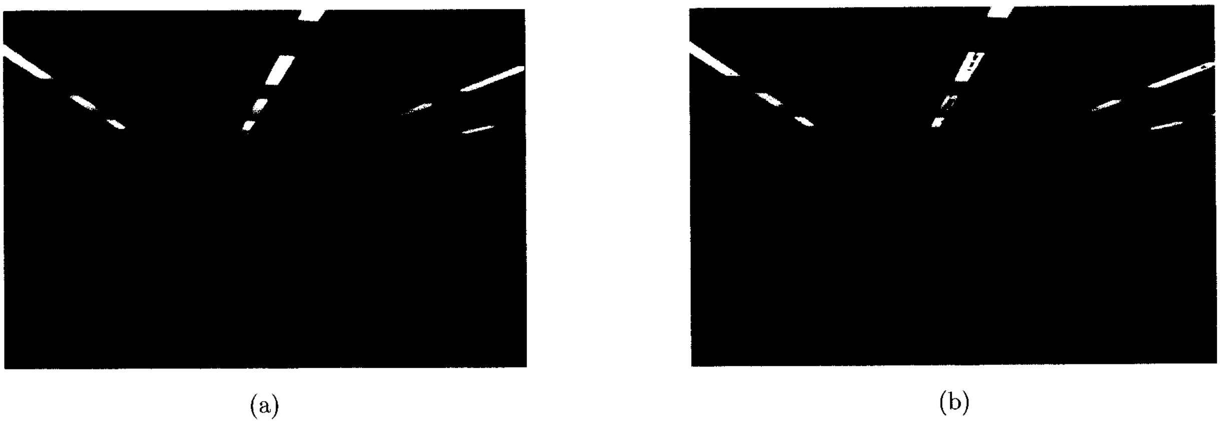
Figure 4 shows the disparity map obtained by using irregular quadtree decomposition and the exhaustive full-search block matching algorithm, with the values of disparities recalculated in the range of 0 (black: far from camera) to 255 (white: near to camera). The disparity map identifies regions for which apparently good disparity estimates have been obtained. From this disparity map, the regions around the objects closer to the camera (books) are almost white and the regions around the objects further from the camera (back

FIG. 3. Left, middle and right images for typical perspectives input; the middle image is for evaluation of the synthesized image.



(a) (b)

FIG. 4. Disparity map, (a) irregular quadtree decomposition, and (b) full-search block matching.



(a)

(b)

FIG. 5. Synthesized image at middle position between left and right image: (a) 32.8 dB PSNR by irregular quadtree decomposition, and (b) 30.0 dB PSNR by full-search block matching.

TABLE 1. Comparison of irregular quadtree decomposition and full-search block matching.

	Irregular quadtree decomposition	Full-search block matching
PSNR	32.8 dB	30.0 dB
Running time	13.7 seconds	46.8 seconds

ground) are almost black.

Figure 5 shows the synthesized intermediate image finally at the middle position between the left and right images. The synthesized image is evaluated by its peak-signal-to-noise ratio, PSNR, which is 32.8 dB and 30.0 dB for irregular decomposition and the full-search block matching algorithm, respectively. Also, computation times are 13.7 seconds and 46.8 seconds for both above algorithms, respectively. Table 1 summarizes the experimental results with respect to the items of PSNR and computation time. Even though the full-search block matching method is generally the best one for solving this problem, the reasons for the degradation of performance, compared to the proposed method, of the full-search block matching algorithm are due to :

(1) Finding the local minimum rather than the global one. Practically, a fixed block neighboring the observed pixel (a 5 x 5 block is used in this paper) is adopted for finding the corresponding point; and
(2) Dividing the region having the same disparities on the surface of the object into small blocks.

VII. CONCLUSIONS

To allow multiple viewers to see the correct perspective and provide a single viewer with motion parallax cues during head movement, more than two views are needed.

For this reason, technologies for synthesizing the intermediate image at an arbitrary location between both left and right images are proposed. The main technology of this paper is to divide the image into blocks by irregular quadtree decomposition whose criteria are variance in block and block size. All kinds of configurations of the objects in the real world are considered, then the general principles are derived, and they are applied to occlusion points. Finally, for evaluating this proposed algorithm, it is compared with exhaustive full-search block matching, and obtained a better quality of synthesized image and took 30% of the latter's computation time.

From now on, the appropriate matching algorithm for images with larger disparities between both images should be considered carefully, and multiple images more than two images are desirable.

ACKNOWLEDGMENT

The author thanks Mel Siegel, J.S. McVeigh and S. Sethuraman of CMU in USA for his basic comments and appreciates the help of the Communication Research Lab. Japan during his sabbatical year.

*Corresponding author : ktkim@hannam.ac.kr

REFERENCES

[1] V. Komar, Victor Komar, and Kyung-tae Kim, "Synthesizing of intermediate view images from two images," *Proceedings of Asia Display/International Display Workshop '01*, pp. 1381-1384, 2001.

[2] Kyung-tae Kim, Mel Siegel, and J. Y. Son, "Synthesis of a high-resolution 3D stereoscopic image pair from a high-resolution monoscopic image and a low resolution depth map," *Proceedings of SPIE Stereoscopic displays and virtual reality system V*, vol. 3295, pp. 76-85, 1998.

[3] J. S. McVeigh, *Efficient Compression of Arbitrary*

Multi-view Video Signals, Dissertation for the degree of Ph.D., Carnegie Mellon Univ., USA., 1996, pp. 65-68.

[4] Sriram Sethuraman, *Stereoscopic Image Sequence Compression Using Multiresolution and Quadtree Decomposition Based Disparity-And Motion-Adaptive Segmentation, Dissertation for the degree of Ph.D., Carnegie Mellon Univ., USA., 1996*, pp. 36-41.

[5] Fujii T., "A network model for binocular parallax extraction - an estimation of contour on isoparallax plans," *Trans. of IEICE D-2*, vol. Ja73-D-2, no. 8, 1990.

[6] Kyung-tae Kim, Y. Arakawa, and M. Siegel, "Intermediate image generation using multi-resolution and irregular quadtree decomposition", *Proceedings of SPIE*, vol. 5243, pp. 104-114, 2003.

[7] J. Canny, "A computational approach to edge detection," *Trans. Pattern Anal. Mach. Intell.*, PAMI-8 (6), pp. 679-698, 1986.

[8] C.J. Hilditch, Linear Skeletons From Square Cupboard, Machine Intelligence 4, *Edinburgh Univ. Press, UK., 1969*, pp. 403.

[9] H. Freeman, "Computer processing of line drawing image," *Computing Surveys*, vol. 6, pp. 57-59, 1974.

[10] J. S. McVeigh, M. W. Siegel, and A. G. Jordan, "Algorithm for automated eye-strain reduction in real stereoscopic images and sequences," *Proc. Human Vision and Electronic Imaging, SPIE*, vol. 2657, pp. 307-316, 1996.