Hardware realization of panoramic camera with speaker-oriented face extraction for teleconferencing

Yukinori NAGASE, Takahiko YAMAMOTO, Takao KAWAMURA and Kazunori SUGAHARA
Faculty of Engineering, Tottori University
Koyama-Minami, Tottori 680-8552, JAPAN
Email: {kawamura,sugahara}@ike.tottori-u.ac.jp

***Abstract*— In this paper, a panoramic camera with speaker oriented face extraction function is proposed. For the face extraction, the Genetic Algorithm(GA) is implemented in a Field Prgramble Gate Array(FPGA) chip. Panoramic video signal generator circuit is also implemented in the FPGA for teleconferencing. Experimental results are included.**

I. Introduction

The advancement of communication speed in digital networks makes it possible to exchange a large amount of information, such as movies, images or human voices, between distant places. Teleconferencing is one typical example of the practical applications in such fields.

In most of the current teleconferencing systems, a full view image of a conference room and voice of speakers are bidirectionally exchanged between distant places. However, in teleconferencing situations, face images of speakers are also requested to perceive facial expression. Fig.1 shows a typical example. To take both of full view images of conference rooms and face images of speakers, it is required to develop a multi camera system with human operations.

In this paper, a panoramic camera with speaker-oriented face extraction function is proposed. The proposed camera is constructed with three NTSC video camera elements, video decoder and encoder LSIs and the proposed video image processor. The video image processor has a panoramic image generation function from output video signals of three video camera elements and a face extraction function. The video image processor is implemented on one FPGA chip, and is applicable to develop stand alone systems.

II. Hardware configuration

The hardware configuration of the proposed camera is shown in Fig.2. As shown in the figure, the camera is constructed with the following 5 blocks.

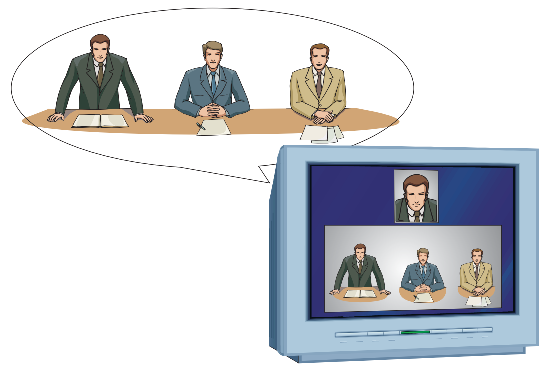

Fig. 1. Example of a preferred teleconferencing.

1) Image input block
   Image input block is constructed with three NTSC video camera elements as shown in Fig.3. Each camera element in Fig.3 outputs NTSC video signal.

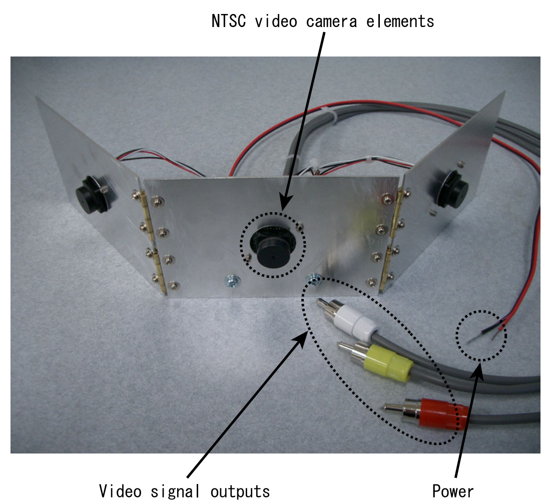

Fig. 3. Image input block.

2) Video signal decoder block
   Video signal decoder LSI, MSM7664B[1] pro-

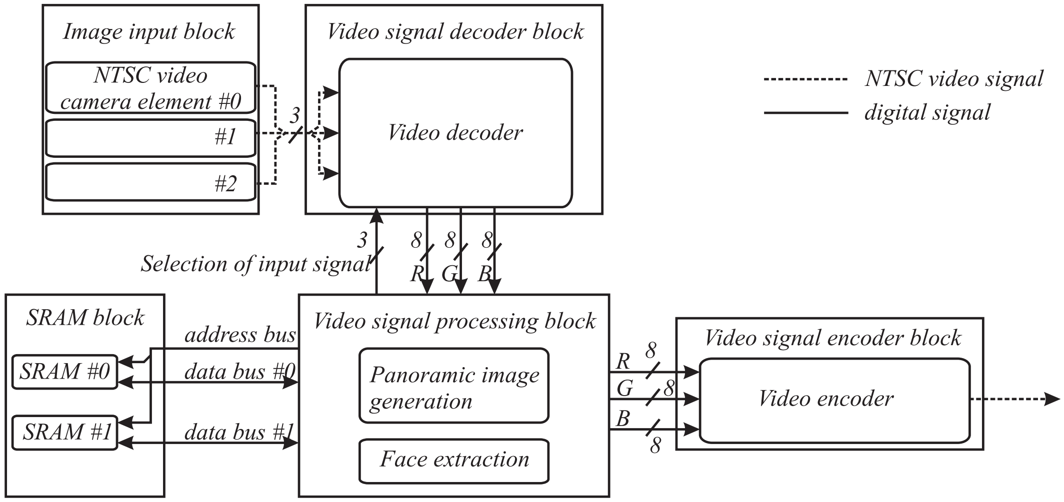

Fig. 2. Hardware configuration.

duced by Oki Electric Industry Co., Ltd., converts NTSC video signals of three video camera elements to 640 × 480 [pixels] size RGB digital image signals. The decoder chip has 4 input ports and 3 of them are utilized in this camera. The switching of these input ports is controlled by the address signals from the video signal processing block.

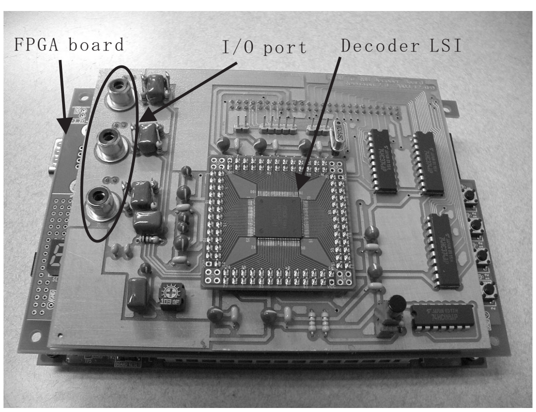

Fig. 4. FPGA, video decode and encode board.

3) Video signal processing block
In the video signal processing block, functions of panoramic image generation from three RGB digital images and of speaker face extraction are implemented in one FPGA. Input signal selection signals for the video signal decoder is also generated in this block. The details of the panoramic image generation function and of the face extraction function are described in the following chapters.
4) SRAM block
The input images and the generated panoramic image is stored in the SRAM block. The SRAM block is constructed with two SS-RAM(Synchronous Statical Random Access Memory) chips and they are able to be accessed independently.
5) Video signal encoder block
Video signal encoder LSI, MSM7654[2] produced by Oki Electric Industry Co., Ltd., generates a NTSC output signal according to the output digital image which the video signal processing block generated.

III. Generation of panoramic images

The 3 input images from the image input block are compliant with the NTSC specifications, they are 640 × 480 [pixels] size. The output image also should be compliant with the specifications, it should have the same size. Considering these points, the output image is designed as shown in Fig.5. The lower half area of the output image is arranged for the panoramic image and the extracted face image is enlarged and is represented on the upper half of it.

The panoramic image part in the output image is 640 × 180 [pixels] size and it is generated with $3/8$ reduced 3 input images with two overlapped area in 40 pixel width.

IV. Face extraction

The processing for the face extraction of the speaker is accomplished by the technique based on the GA (Genetic Algorithm)[3], [4]. The block diagram of the face extraction processing is shown in Fig.6. As shown

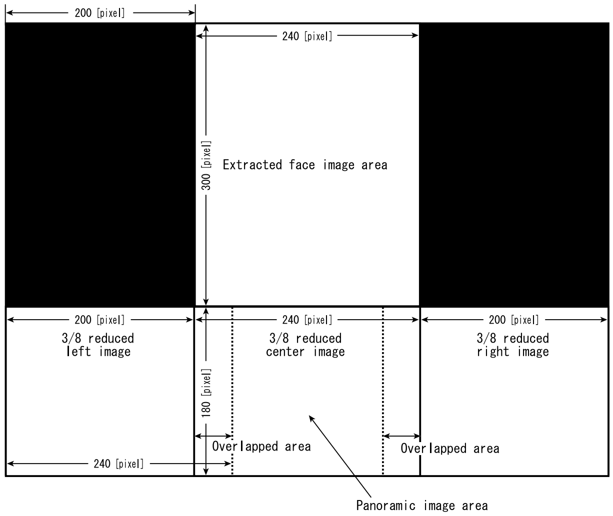

Fig. 5. Generation of panorama image.

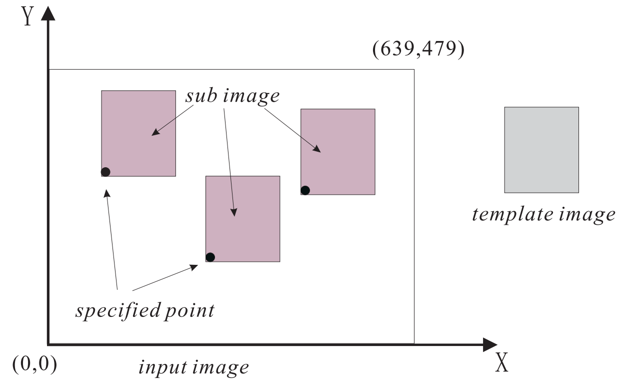

Fig. 8. Comparison of input and template image.

in the figure, the GA process is constructed with the following functions,

1) The random number generator.
2) The evaluation of obtained intermediate results.
3) The selection of preferred results as parents genes.
4) The crossover and mutation of parent genes and generate child genes.

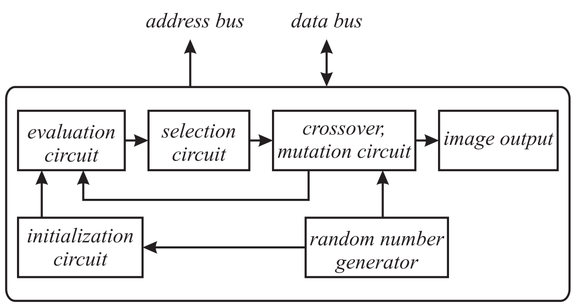

Fig. 6. Face extaction.

In this paper, the number of genes is set as 32 and the genetic code is defined as shown in Fig.7. The values of X-axis and Y-axis in this code represent the coordinates of the pixel in the bottom left corner of a sub-image, as shown in Fig.8.

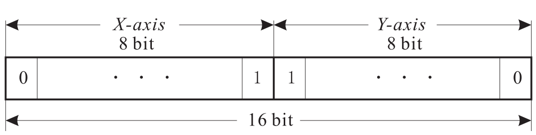

Fig. 7. Genetic code.

The evaluation of intermediate results are accomplished by the comparison of RGB values of each pixel in sub image which addressed by the each genetic code with those in the given template image. The genes are classified into parents group and children group according to the evaluation results. The upper half of genes is classified into the parents group and the lower half is classified into the children group. And two genes selected randomly in the parents group generate two children by the crossover process. The crossover process is repeated 8 times and the generated 16 genes updates the children group. The evaluation and crossover processes are shown in Fig.8 and Fig.9, respectively.

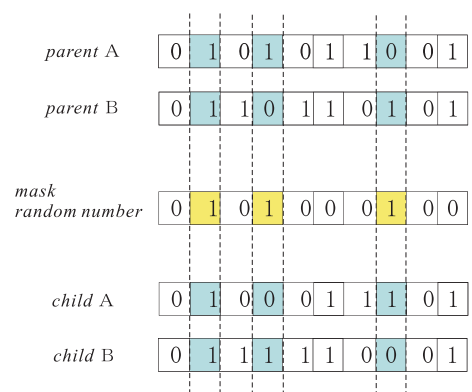

Fig. 9. Crossover.

After the crossover processes, each gene is updated by the mutation process as shown in Fig.10. In the mutation process, bit number of randomly selected genes are inverted according to the given mask data. Here, the mask data is also decided according to the random number.

V. Experimental Results

A. Face extraction

To confirm the operation of the face extraction function in the produced FPGA, several images are examined.

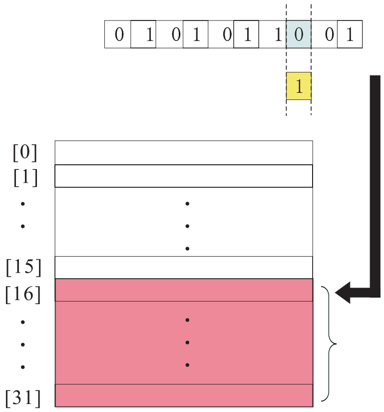

Fig. 10. Mutation.

Fig.11 shows one of the examined images and the locations of sub images corresponding to the initial genes. The template face image used here is shown in Fig.12.

The operation of the face extraction process is confirmed by the obtained final result shown in Fig.13.

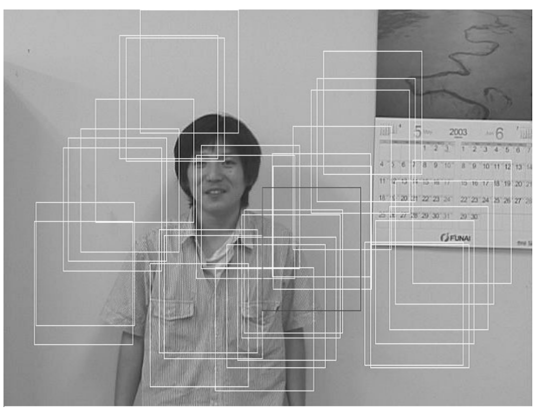

Fig. 11. Initial state.

Fig. 12. Template image.

B. Generation of panoramic images

The example of the generated panoramic images is shown in Fig.14. In Fig.14, the final result of face extraction of speaker is also included.

Fig. 13. Final result.

Fig. 14. Composite image.

VI. Conclusion

In this paper, the panoramic camera with speaker-oriented face extraction for teleconferencing is proposed. By using the proposed camera, not only full view of conference room but also facial expressions of speakers are able to be exchanged between distant places.

The functions of the face extraction and those of the panoramic image generation are implemented in one FPGA chip. The inputs and output images of the camera are compliant with the NTSC specifications, it is easily applicalbe to stand alone systems.

References

[1] Oki Electric Industry: MSM7664B data sheets(in Japanese), (2001).
[2] Oki Electric Industry: MSM7654B data sheets(in Japanese), (1999).
[3] D.E.Goldberg:"Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley (1989).
[4] Vose, Michael D.:"The Simple Genetic Algorithm", (1998).