트리 기반 정적/동적 영상 모자이크

강오형* · 이양원*

Tree-Based Static/Dynamic Image Mosaicing

Oh-hyung Kang* · Yang-won Rhee*

요 약

본 논문에서는 효율적인 비디오 데이터베이스를 구축하기 위하여 카메라와 객체 파라미터를 이용한 트리-기반 계층형 영상 모자이크 시스템을 제시한다. 장면 전환 검출을 위하여 그레이-레벨 히스토그램 차이와 평균 명암도 차이를 이용한 방법을 제시하였다. 카메라 파라미터는 최소 사각형 오류 기법과 어파인 모델을 이용하여 측정하고, 두 입력 영상의 유사성을 측정하기 위하여 차영상을 이용한다. 또한 동적 객체는 매크로 블록 설정에 의하여 검색되고 영역 분할과 4-분할 탐색에 의하여 추출한다. 동적 객체의 표현은 동적 궤도 평가 함수에 의하여 수행되고 블러링을 통하여 부드럽고 완만한 모자이크 영상을 구축한다.

ABSTRACT

This paper proposes a tree-based hierarchical image mosaicing system using camera and object parameters for efficient video database construction. Gray level histogram difference and average intensity difference are proposed for scene change detection of input video. Camera parameter measured by utilizing least sum of square difference and affine model, and difference image is used for similarity measure of two input images. Also, dynamic objects are searched by through macro block setting and extracted by using region splitting and 4-split detection methods. Dynamic trajectory evaluation function is used for expression of dynamic objects, and blurring is performed for construction of soft and slow mosaic image.

키워드

image mosaic, scene change detection, least sum of squared difference, affine model, difference image

Ⅰ. Introduction

By enhancement of computer information technology and increment of user's requirements, research concerned with multimedia data processing is performed widely. However, because of large amount of multimedia data and much process time, it remains to hard problem to process, transmits and understands information. Especially, video has problem that users are not easy understand content of video fast because video has much wide information as integrated form with text, still image, sound and voice etc. as well as moving picture itself. Also, implementation of method that can express easily their information because scene

*군산대학교 컴퓨터공학과
접수일자 : 2003. 7. 7

and contents of scene differ in one video is not easy, and in case of still image that is retaining resemble contents has a problem that in occupy many storage spaces.

Mosaic image construction refers to the creation of single new image that compose several videos and still images related together[1]. Mosaic is classed by static mosaic that focus on background, dynamic mosaic for dynamic object description, and synopsis mosaic that appear representatively integrating static and dynamic description[2].

Technologies of mosaic construction process consist of arrangement, integration and overlapping analysis of continuous images. The most traditional field of application for panoramic mosaic system is construction of air artificial satellite picture, recently scene fixing and scene sensing, video compression and indexing, and research of camera resolution as well as, even study in very various field to simple picture edit[2-4]. Frame image that represent one scene in video is known as representative frame image, it have many difficulty in understanding scene of moving picture only by representative frame. Mosaic image solves ambiguity problem of information that can drop in representative frame of video because it make resemble much images to single image. Mosaic image has great advantage that is use of minimum storage space, fast transmission of data, and understanding of whole scene because it is including all adjacent images.

Many papers suggested difference image, coordinate conversion technique of image, optical flow and problem and solution of motion estimation. Shum et al.[3] proposes a description method of mosaic image that each input image is concerned with transformation matrix that is not method to reflect on general plane. However, this method has a disadvantage that is very sensitive at noise. Xiong et al.[4] proposed way to construct the virtual world by getting 4 images rotating camera 90 degrees. IBM's system called ImageMiner is system that integrate still image analysis in video analysis as IBM's trademark, described method to recognize object using color, texture and shape information as step that analyze mosaic image[5]. However, ImageMiner system has problem that it does not recognize dynamic object because mosaic image is recognized object only with color, texture, and shape information. Irani et al.[2] proposed about dynamic mosaic construction for dynamic object, but specific method that can identify camera and dynamic object that move did not describe.

The easiest method in that create mosaic image is that consisted transfer only between images. Images moved so only can be implemented easily on minimum restriction, and mosaic of high resolution that quality is high of mosaic image can be created, and advantage is as well as computing time is fast[6-9].

Camera motion must be measured necessarily to construct mosaic. One of the camera motion estimation is a method that perform through parameter calculation of camera that use optical flow[3, 10]. Also, motion models that used to extract camera parameter have used as two-dimensions parameter motion model and complicated three-dimension motion model[3]. Affine model that can measure rotation at the same time including movement and scaling among 2 dimensions motion model is utilized most in camera parameter measuring[4].

This paper proposes, after preprocessing of input as still image and video frame, method that measure parameters of camera and dynamic objects, and method that construct mosaic image by object extraction. Whole system structure is shown as Fig. 1.

Chapter 2 present about scene change detection of input video. Chapter 3 explains about hierarchical mosaic image construction method and Chapter 4 explains about static and dynamic mosaic image construction. Chapter 5 experiment static and a dynamic mosaic image construction and Chapter 6 conclude this paper and find about future work.

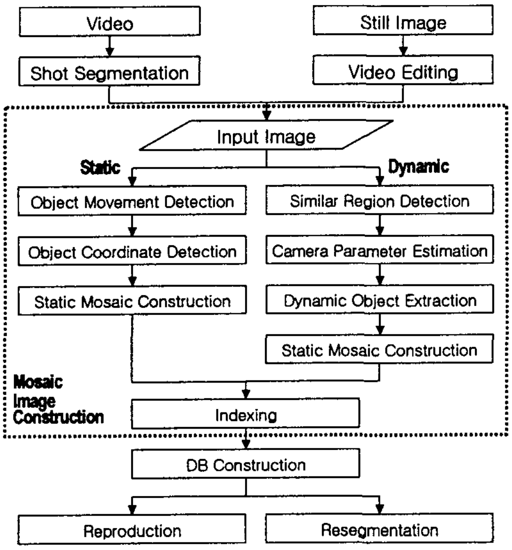

Fig. 1 Whole system structure

II. Scene Change Detection of Video

In this paper, scene change detection is based on the framework of the twin-threshold method proposed by Zhang[11] that is able to detect both abrupt and gradual transition. To segment the video, suitable metrics is defined, so that a shot boundary is declared whenever that metric exceeds a given threshold. Gray level histogram difference is used as the first metric in our algorithm because histogram is less sensitive to object motion than other metrics, shown in Equation (1). The second metric is average intensity difference, shown in Equation (2).

$$D_i = \sum_{j=1}^{Bins} \left| H_i(j) - H_{i-1}(j) \right| \tag{1}$$

$$AI_i = \frac{\sum_{j=1}^{Bins} j \times H_i(j)}{\sum_{j=1}^{Bins} H_i(j)} \qquad AI_{i-1} = \frac{\sum_{j=1}^{Bins} j \times H_{i-1}(j)}{\sum_{j=1}^{Bins} H_{i-1}(j)}$$

$$AID_i = AI_i - AI_{i-1} \tag{2}$$

Hi(j) indicates j-th bin of the gray level histogram belonging to frame i. 256 bins are choose for gray level histogram. Di denotes the histogram difference between frame i and its preceding frame (i-1). AIi is the average intensity value of the frame i, and AIDi is the average intensity difference between frame i and (i-1).

III. Hierarchical Image Mosaicing

Requirement that is gone ahead first in mosaic image creation is that do all image to arrange resemble images consecutively. This needs to match overlapping area between two images to compare to one part.

3.1 Mosaicing Flow

In mosaic system as Fig. 2, image in scene that acquired by video camera and arranged image that acquired by general camera are used for input. Image that is acquired by general camera can be greatly difference of each image. In this case, it is effective that construct mosaic through similarity measure.

Fig. 2 shows flow for mosaic image

construction. After videos of real life that is acquired from camera are input by continuous frame as segmented scene and still images are

input automatically arranged image, then two frames or images are compared. For measure camera movement between two images, least sum of square difference and affine model is used. In this paper, macro block setting, region splitting, and 4-split detection method are proposed to extract and recognize dynamic object. After extract and recognize dynamic object, construct smooth mosaic image generally applying trajectory description of dynamic object and blurring method.

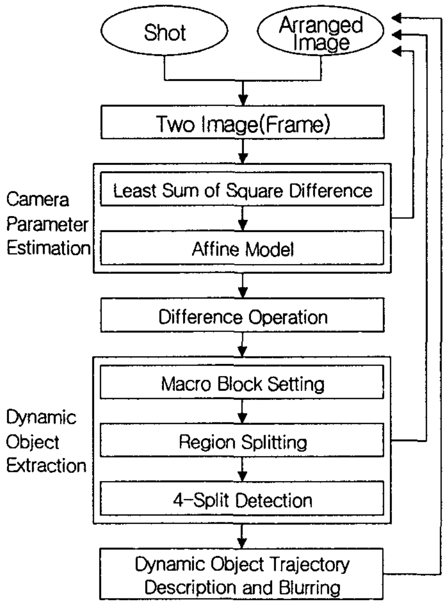

Fig. 2 Mosaicing Flow

3.2 Tree-based Mosaicing

This paper proposes creation method of tree-based mosaic image to get fast computing time of mosaic image creation. Tree-based mosaic image creation is not that adjacent images are created to consecutively mosaic image. Image mosaicing creates one partial mosaic image by comparison of only two of adjacent images. Therefore all source images are created to partial mosaic image. One general mosaic image is created by creation of new partial mosaic image repeatedly by such created partial mosaic image. If source image is 16, 4 levels of partial mosaic exist.

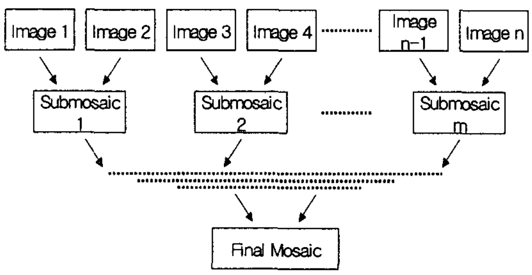

Fig. 3 Tree-based Mosaic Image

Ⅳ. Static and Dynamic Image Mosaic

4.1 Static Image Mosaicing

Method to extract dynamic object in static background uses difference image technique, is that calculate by using difference operation between pixels of two images as Equation (3).

$$D(x,y)=|I_a(x,y)-I_b(x,y)| \tag{3}$$

Here Ia is a base image and Ib is a reference image. Also, need local detection to detect object region. This paper does to detect object using 16×16 macro block, detect by performing local difference operation as Equation (4).

$$LD(x,y)=\sum_{y=1}^{N}\sum_{x=1}^{N}|I_a(x,y)-I_b(x,y)| \tag{4}$$

Local difference operation is that subtract reference image Ib from base image Ia and calculates this as size of 16×16 pixel. In this way, after detect object region by performing local difference operation using macro block, similarity of two images is measured by using least sum of square difference, and measure affine parameter using coordinate value three point or more obtained by calculating camera movement, then two images are matched for constructing static mosaic image. Least sum of square difference and affine parameter measure is described in detail in mosaic construction using dynamic background and object in next section.

4.2 Dynamic Image Mosaicing

There is difficult to detect movement of camera and object between two images that have dynamic object. Even if there is camera movement, if object is filling image as a whole, it do not recognize camera motion. Also, if big object moves when camera does not move, it can not recognize this as camera motion.

This paper proposes following method to solve this ambiguity. Simple method that can think most usually is that object is occupying middle of screen mainly when camera takes an important object. So, similarity is measured by based on macro block of outer region except middle of base image in this method.

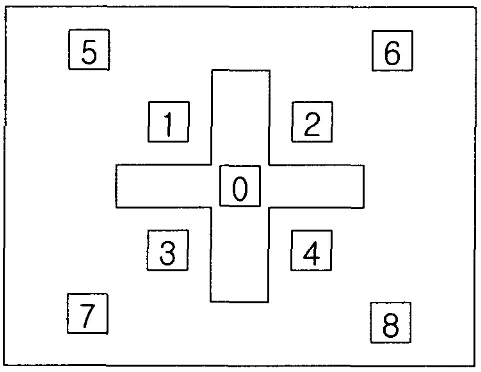

Fig. 4 Macro Block of Base Image

Fig. 4 shows macro block of base image for comparison. Each macro block selects macro block by non-linear as central outer region of cross shape. Preferential assumption is that the size of object does not occupy half of image. Algorithm that detect dynamic object by setting macro block is as Fig. 5.

```
(1) input two image(frame);
(2) similarity measure of macro block 1, 2, 3 and 4;
    if(value of least sum of square difference
                                         < threshold)
      then {measure of motion vector;
          goto (3);}
      else goto (1);
(3) if(motion vector is similar to all four macro block)
      then {camera parameter = motion vector;
          goto (4);}
      else {measure of motion vector by extending
                            to macro block 5, 6, 7 and 8;
          goto (4);}
(4) measure of camera parameter through affine
                                       transformation;
(5) local difference operation between two images
```

Fig. 5 Dynamic Object Detection Algorithm

If result that get performing difference operation using above algorithm is more than threshold, it is considered that dynamic object exists. Next to, look around each methods that is used in above algorithm.

4.2.1 Least Sum of Square Difference

First, must extract correct camera parameter to look for similarity between two image, this paper proposes least sum of square difference about fixed window block as Equation (5).

$$E(C)=\sum_{b\in W}[I_i(X+b)-I_j(X+b+d_k)]^2 \tag{5}$$

In equation (5), X refers to pixel position of x and y, and b refers windows of a regular square of an image. Value of lease sum of square difference is calculated through block of all dks of reference image Ij in place that is calculated by square of difference value between reference image Ij and base image Ii. At this time, minimum value among square difference that is calculated is selected.

4.2.2 Affine Model

Camera parameter including rotation, scaling and movement of image is measured by using affine model based on the most similar pixel value that is detected by using Equation (6).

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \tag{6}$$

Three nonlinear points $(x_1', y_1', x_2', y_2', x_3', y_3')$ $(x_1, y_1, x_2, y_2, x_3, y_3)$ of base image and reference image are need to calculate affine model. That is, there must be three linear equations as Equation (7).

$$x' = ax+by+e \qquad y' = cx+dy+f \tag{7}$$

Through making of three linear equations using these three points, six parameters(a, b, c, d, e, f) of affine model is calculated. After calculate six parameter by calculating three equations through three points of base image and three points of reference image, arrange pixels that correspond to all pixels of reference image. In equation (7), a is a scaling coefficient of x axis, b is a rotation coefficient of x axis, c is a movement coefficient of x axis, d is a scaling coefficient of y axis, e is a rotation coefficient of y axis, f is a movement coefficient of y axis.

4.2.3 Local Difference Operation

By next time, must detect movement of object, local detection must performed for this. Method to present in this paper performs by comparison through difference operation using motion vector and macro block between two images with camera motion that is calculated, as Equation (8).

$$E(O)=\sum_{b\in W}[I_i(X+b)-I_j(X-u(k)+b+d_k)]^2 \quad (8)$$

Here u(k) refers to distance of camera movement as motion vectors of x axis and y axis, and error value between two images is calculated by subtraction of it. At this time, large movement is detected if give do larger threshold that detect movement of object, and small motion can be extracted if give as is small.

4.2.4 Region Splitting

Also, this paper proposes a region splitting method through basic assumption for dynamic object detection. In Fig. 6, if suppose that b and c region of two images show similar region between two images, following assumption is followed for detecting dynamic object.

First, if some part of region a and some part of region c or d is corresponded, this region becomes object.

Second, it some part of region b and d is corresponded, this region becomes dynamic object.

Finally, conflicting region between region b and c that corresponding between two images becomes object.

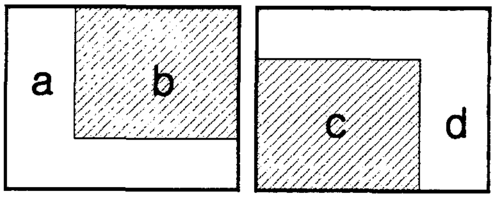

Fig. 6 Region Splitting for Dynamic Object Detection

4.2.5 4-Split Detection

Dynamic object extraction is that extract only dynamic object by detection of similar region that existing dynamic object in two image. If value of difference image between two similar region is large, there is assumption that dynamic object is existed within two images. As shown in Fig. 7, region of 1 quadrant is calculated and compared. If value of difference image is small, computation proceeds to 2 quadrants and so on continuously. Size of dynamic object ignores object fewer than smallest 7×7 pixel size, and when multiple dynamic object is detected, the largest dynamic object between them is extracted. Detection process is performed up to last 8×8 block and dynamic object region is created by sum of blocks.

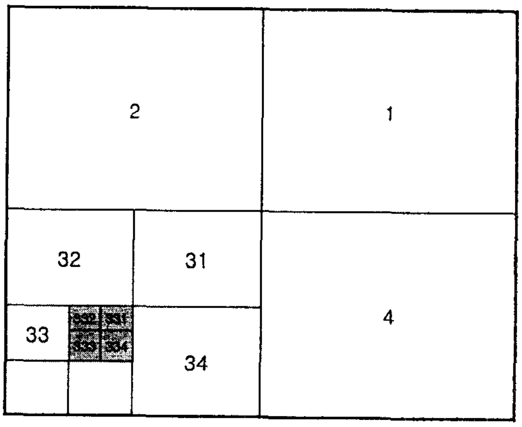

Fig. 7 4-Split Detection Method

4.2.6 Dynamic Trajectory Description and Blurring

Background image composition compose only remainder background image after dynamic object extraction. At this time, background part of other remainder image is inserted on part of extracted dynamic object. After create background mosaic image, the description of dynamic object express object that distance is more than 1.5 times of maximum width and height size of the extracted object. Evaluation function that present in this paper is as shown in Equation (9).

$$\begin{aligned} &\text{if } (A > 1.5B) \text{ then describe dynamic object} \\ &\text{where } A = Ii(x2,\ y2) - Ii{-}1(x1,\ y1) \qquad (9) \\ &\quad B = \text{Length}(O_{MAX}[(x_1,\ x_2),\ (y_1,\ y_2)]) \end{aligned}$$

In Equation (9), after calculating maximum size(OMAX) of dynamic object, when dynamic object moved to right, it is described if distance difference((Ii(x2, y2) - Ii-1(x1, y1))) between left region of present image and right region of preceding image is more than 1.5 times of maximum object size.

Blurring creates visually smoothing mosaic image using the most general method that sum of whole mask set to 1 using 3 x 3 mask on the border part where each image may integrated.

Such created mosaic image is stored to video database in form of tree-based mosaic that explained in section 3.2. That is, image that is actually stored to video database is a complete constructed mosaic image. All actuality each object and backgrounds are stored being implied in this mosaic image and partial description and omission according to method that express this are possible. Also, it can use storage space more efficiently because do not store each all images and can search fast because partial searches for single mosaic image are possible.

V. Experimental Result

To implement mosaic system that proposed in this paper, image is acquired from HITACH Digital Zoom camera. Input image is used after normalizing of 320 * 240 sizes and implementation is performed by using Visual C++6.0 in Pentium-4 1.3GHz.

Fig. 8 display some of whole frame in 30 second video that take interior of laboratory.

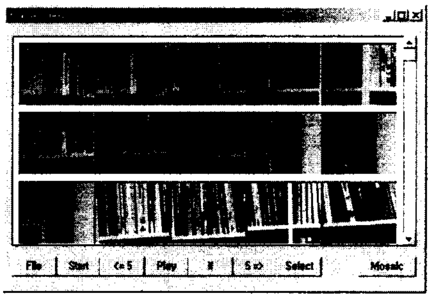

Fig. 8 Frame Images

Fig. 9 display 6 image that is selected arbitrarily among whole frame of Fig. 7 These are image that is used in input to create mosaic image.

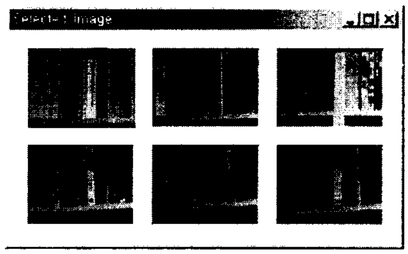

Fig. 9 Selected Input Image

Fig. 10 is static mosaic image that is created from 6 input images which appeared in Fig. 9. In Fig. 10, we can know that image length from left to right is prolonged fairy as it is panoramic mosaic image that dynamic object does not exist. Through this mosaic image, we can understand easily contents of whole video.

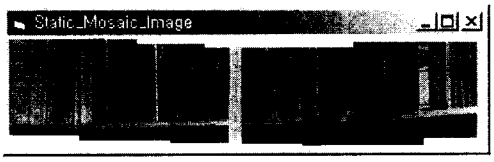

Fig. 10 Static Mosaic Image

Experiment that construct dynamic mosaic image in video that dynamic object exists is as following. First, Fig. 11 is window that extract key frame by proposed scene change detection method for a video 30 second long that dynamic object exists in it.

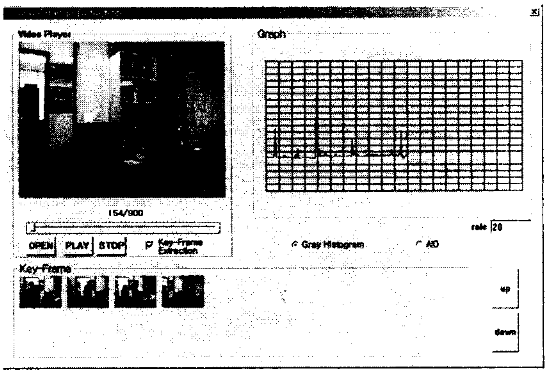

Fig. 11 Scene Change Detection

Fig. 12 is dynamic mosaic image that describing dynamic object. When dynamic object is moved more than 80 pixels by an experiment, dynamic object is appeared in mosaic image.

Fig. 12 Dynamic Mosaic Image

Result that construct mosaic image for 5 videos 30 seconds long that captured in campus is as Table 1.

Table 1. Mosaic Image Construction Result

Input Video		Mosaic Image		Reduction Ratio
Average Frame Number	Average Volume	Average Frame Number	Average Volume	
900	1.16G	97	20.81M	98.25%

Constructed mosaic image reduces storage space of 98.25% than input video's volume. That is, it does so that can understand whole video's contents by one image of very small volume.

In this paper, unique and notable characteristics of research and experiment that compare with existing mosaic system can be summarized as Table 2.

Table 2. System Compare and Evaluation

item	existing system	proposed system
construction base	consecutive frame-based	tree-based
utilizing type	static or dynamic	both static and dynamic
computation area	whole image	partial image
parameter	mainly camera parameter	both camera and object parameter
trajectory description	ambiguous	clear and flexible
mosaic image border	no-soft	soft

Ⅵ. Conclusion

This paper proposed way to construct tree-based hierarchical image mosaicing using camera and object parameter for efficient video database construction.

Camera parameter measured by utilizing least sum of square difference and affine model, and difference image is used for similarity measure of two input images. Also, dynamic objects are detected by through macro block setting and extracted by using region splitting and 4-split detection methods. And soft mosaic image is constructed through blurring after used dynamic trajectory evaluation function for expression of dynamic object. This tree-based mosaic system presented foundation that can reduce time and storage space to construct video database.

In future, consecutive research about more efficient scene change detection, object detection and image reappearance is needed for creating enhanced mosaic image.

참고문헌

[1] Wonbo Y., Jaewon K., Keunhwan J., and Yangweon R., "Tree-Based Mosaic Image and Redistribution," Proc. KIPS, Oct. 8, 1999.

[2] M. Irani, P. Anandan, and S. Hsu, "Mosaic Based Representation of Video Sequence and Their Applications," IEEE Proc. ICCV'95, pp. 605-611, 1995.

[3] Heung-Yeung Shum and Richard Szeliski, " Panoramic Image Mosaic," MSR-TR-97-23, 1997.

[4] Y. Xiong and K. Turkowski, "Creating Image-Based VR Using A Selfcalibrating Fisheye Lens," Proc. CVPR' 97, pp. 237-243, 1997.

[5] J. Krey ß, M. Roper, P. Alshuth, Th. Hermes, and O. Herzog, "Video Retrieval by Still-Image Analysis with ImageMiner," Storage and Retrieval for Image and Video Databases (SPIE), pp. 36-44, 1997.

[6] A. Tekalp, M. Ozkan, and M. Sezan, " High-Resolution Image Reconstruction from Lower-Resolution Image Sequences and Space-Varying Image Restoration," Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. III-169-172, March 23-26, 1992.

[7] L. Teodosio and W.Bender, "Salient Video Stills : Content and Context Preserved," Proc. ACM Multimedia Conf., 1993.

[8] Gulrukh Ahanger and Thomas D.C.Little, "A Survey of Technologies for Parsing and Indexing Digital Video", Journal of Visual Communication and Image Representation, Vol. 7, No.1, March, pp.28-43, 1996.

[9] D. J. Heeger, "Optical Flow Using spatiotemporal Filters", International Journal of Computer Vision, pp. 279-302, 1998.

[10] Richard Szelisk, and H. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps," In Proc. of SIGGRAPH, pp. 251-258, 1997.

[11] H. J. Zhang, A. Kankanhalli, S. W. Smoliar, "Automatic Partitioning of Full-Motion Video", ACM Multimedia System, Vol. 1, No. 1, pp. 10-28, 1993.

저자소개

강오형(Oh-hyung Kang)

1993년 군산대학교 전자계산학과 졸업(학사)
1997년 군산대학교 대학원 정보통신공학과 졸업(석사)
2001년 군산대학교 대학원 컴퓨터과학과 박사과정
1993년~1997년 (주)동양기술개발공사 전산실
1998년~현재 군산대학교 전자계산소 재직
※관심분야 : 멀티미디어, 컴퓨터비전, 네트워크, 원격교육

이양원(Yang-won Rhee)

1978년 숭실대학교 전자계산학과 졸업(학사)
1983년 연세대학교 산업대학원 전자계산학과 졸업(석사)
1994년 숭실대학교 대학원 전자계산학과 졸업(박사)
1979년~1986년 한국국방연구원 정보관리위원회 연구원
1986년~현재 군산대학교 컴퓨터정보과학과 교수
※관심분야 : 멀티미디어, 컴퓨터비전, 인공지능, 가상현실