韓國컴퓨터情報學會 論文誌
第14卷 第10號, 2009. 10.
2009-14-10-2-1

A Revised Dynamic ROI Coding Method Based On The Automatic ROI Extraction For Low Depth-of-Field JPEG2000 Images

Jae Heung Park*, Hyun Joo Kim**, Jong Chae Shim***, Chang Yeul Yoo***,

Yeong Geon Seo****, Ki Jun Kang*****

낮은 피사계 심도 JPEG2000 이미지를 위한 자동 관심영역 추출기반의 개선된 동적 관심영역 코딩 방법

박재흥*, 김현주**, 심종채***, 유창열***, 서영건****, 강기준*****

Abstract

In this study, we propose a revised dynamic ROI (Region-of-Interest) coding method in which the focused ROI is automatically extracted without help from users during the recovery process of low DOF (Depth-of-Field) JPEG2000 image. The proposed method creates edge mask information using high frequency sub-band data on a specific level in DWT (Discrete Wavelet Transform), and then identifies the edge code block for a high-speed ROI extraction. The algorithm scans the edge mask data in four directions by the unit of code block and identifies the edge code block simply and fastly using a edge threshold. As the results of experimentation applying for Implicit method, the proposed method showed the superiority in the side of speed and quality comparing to the existing methods.

요 약

본 논문에서는 낮은 피사계 심도 JPEG2000 이미지의 복원 과정에서 관심영역을 자동으로 추출하여 우선적 처리하는 개선된 동적 관심영역 코딩 방법을 제안한다. 제안한 방법은 기존 방법과는 달리 사용자의 관심영역 지정 과정을 거치지 않고, DWT(Discrete Wavelet Transform)에서 특정 레벨의 고주파 서버 밴드를 사용하여 에지 마스크 정보를 생성한 후에 자동 에지 코드 블록 판별 알고리즘을 사용하여 관심영역을 빠르게 처리한다. 이 알고리즘은 에지 임계값과 4 방향(동, 서, 남, 북)으로 코드 블록 단위의 에지 마스크 정보를 이용하여 에지 코드 블록을 판별한다. 본 알고리즘을 기존의 Implicit 방법에 적용하여 실험한 결과, 제안한 방법이 기존의 방법들에 비해 속도와 품질 면에 있어서 우수함을 확인하였다.

• 제1저자 : 박재흥 교신저자 : 강기준
• 투고일 : 2009. 07. 24, 심사일 : 2009. 08. 13, 게재확정일 : 2009. 10. 26.
* 경상대학교 컴퓨터과학과, 컴·정보통신연구원 ** 진주산업대학교 컴퓨터공학과 *** 경남도립남해대학 인터넷정보기술과
**** 경상대학교 컴퓨터교육과, 컴·정보통신연구원 ***** 경상대학교 컴퓨터과학과
※ 본 연구는 2008년도 경상대학교 광대역연구개발망 활성화를 위한 연구과제 지원비를 받아 수행되었음.

▸ Keyword : 관심영역(Region-Of-Interest), Dynamic ROI coding, low DOF image, Automatic ROI extraction

I. Introduction

Due to the development of the Internet, image information is being used in various applicable fields as a highly effective medium in information transfer. However it is difficult to transfer large image data without a loss in real time, on low broadband networks, such as a mobile environment. Therefore rather than transferring and recovering all the image data, there have been studies on methods that process selected specific regions or a user's region of interest first. This study will mainly focus on the ROI coding method of JPEG2000. The ROI coding method not only enables viewing high-quality ROI at a low bit-rate, but also provides an excellent trade-off between quality and compression [1-3]. This coding method is divided into a static coding method, in which ROI definition and coding are carried out during the compression process of turning the original image into a JPEG2000 image, and a dynamic method in which ROI definition and coding are performed during the recovery process of a JPEG2000 image that was not ROI-coded. In most applied areas, the static coding method is fully applicable, but the dynamic method should be used when the ROI is unknown in an interactive environment during the image recovery process [3]. There also needs to be fast processing in order to provide a steady dynamic ROI coding service.

On the other hand, a low DOF (Depth-of-Field) image introduced in [4] is composed of a focused area (ROI) and a blurry area (background); the background can be expressed in a 2-dimensional Gaussian function. This image has strengths in bringing out a selected area and in tak and p only a small space, and th only a n tak and p onimage space, such as portraits, including wn ding or product photos.

This study explains the dynamic ROI coding method for a low DOF JPEG2000 image. In order to encode ROI, an ROI extraction process must be applied. The extraction automatically processes the image by using user-specified ROI shape data or certain data of the image. The user-specified ROI method takes too much time and cost, so various methods to automatically extract ROI without a user's help are being studied in [4-7]. However, the existing automatic ROI extraction methods use filtering processes such as the Roberts filter, Prewitt filter, and Laplacian filter, along with post-processing jobs which enable them to extract an exact ROI, but a fast extraction is difficult due to the amount of calculations requ ired. Therefore in this study, a dynamic ROI coding method is proposed where only partial edge mask data is scanned rather than additional filtering or processing, in order to quickly extract ROI automatically.

II. Related Works

1. Automatic ROI Extraction Method

Automatic extraction of ROI in images is a very difficult process, for not only do the ROI in images differ from person to person, there are situations where the image is composed of various colors, shapes and textures, and some contain various objects or images that cannot be specially displayed. In order to extract the ROI generally, an image partition process must be undertaken. Image partition methods include the region-based method, critical value-based method, edge-based method, and a mix of various algorithmd The region-based method [8-9] applies the image characteristic of the statistical values of colors and textures in the same region being similar; the critical value-based method [10-11] classifies the regions based on the value of some constant image feature; the edge-based method [12-13] uses the hypothesis that the edges between objects haves extreme changes in the definitions. After the image partition process, the ROI extraction process is undertaken. During this process, there are the Kim[4], Chen, O.T.-C.[5], and Fei Wang[6] methods in which feature extraction and post-morphologic operations automatically decide the ROI, and the Kong H-S[1] and Rene Rosenbaum[3], and H. Yang[14] methods where the user directly designates and extracts the ROI from the partitioned image.

The automatic ROI extraction method requires a large amount of calculation due to repeated filtering and post-processing, which makes it difficult to do quick processing. The manual ROI extraction method added a human element in order to compensate for the differing definitions of ROI depending on each person due to different cognitive elements, but it requires a lot of time and cost.

2. Dynamic ROI Coding Methods

The definition of ROI is usually done during the encoding process, but if the ROI is unknown at that point, the ROI coding is carried out by receiving ROI shape information in a mutual exchange during the decoding process. This defining of ROI during the decoding process for ROI coding is called dynamic ROI coding [1-3]. A general dynamic ROI coding process is as follows:

① Use the original image to automatically or semi-automatically extract the ROI by user.

② With the extracted ROI data, the preferred processing is carried out in the order of ROI importance.

③ The compressed ROI image is reconstructed.

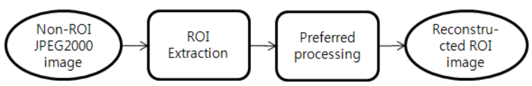


그림 1. 동적 관심영역 코딩 과정
Fig 1. Dynamic ROI coding process

The last stages of this method are totally different depending on how the ROI is defined. It requires many calculations, and could also create a non-compatible data stream. A typical dynamic ROI coding method includes the standard implicit method [15] and the modified implicit method [14] which complements the standard method, flexible and dynamic ROI method [3], prioritized ROI method [16], and fast ROI transcoding method [1]. These existing methods declare the ROI through the user looking at the rough image during the image recovery process, and thus are unable to provide real-time service.

III. Proposed Method

The processing stages of the proposed method are based on the edge mask data to eliminate noise and extract an automatic ROI, then uses the ROI information to do preferred processing and ultimately reconstruct the ROI image. For efficient processing, if the size of ROI takes up more than 90% of the overall image, the image is reconstructed without ROI coding; this is due to the overhead on ROI coding.

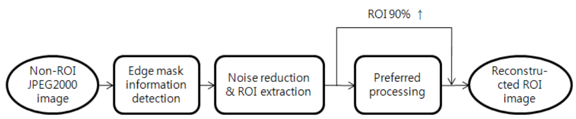


그림 2. 제안된 방법의 처리 과정
Fig 2. The proposed method process

1. Mask Information Detection

Discrete Wavelet Transform (DWT) has been a very popular and commonly used tool for image analysis and has become part of the JPEG2000 standard. JPEG2000 uses a Cohen-Daubechies-Feauveau (CDF) 9/7 wavelet as a default wavelet for lossy compression.

A DWT is, in practice, created by passing the image through a series of filter bank stages. An image is first filtered in the horizontal direction. The high-pass filter and the low-pass filter are finite impulse response filters. The filtered outputs are then down sampled by a factor of 2 in the horizontal direction. This simply means that every second row of the resulting image is kept. Each of these signals (sub images) are then filtered by an identical filter pair in the vertical direction and down sampled by a factor of 2 again. The transform results in four new sub-bands at each level of decomposition; namely, an approximation sub-band at low resolution, LL, and three directionally sensitive detail sub-bands:HL-vertical features (horizontally high pass), LH-horizontal image features(vertically high pass), and HH - diagonal features (horizontally and vertically high pass). Two-level decomposition, seen in Figure 3, is processed to the second decomposition, only using the LL sub- as input.

In JPEG2000 compression capacity, it is known that the best capacity is shown when the decomposition level is 4, but the number of levels used in this study for edge extraction can be decided upon depending on the size of the image and the code block.

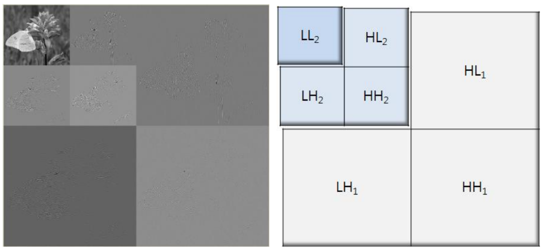


그림 3. 2 레벨 분해의 예제
Fig 3. Example of two-level decomposition

Using the HL, LH, LL sub-bands of a specific level, the edge mask data of vertical, horizontal, and diagonal directions is extracted. The following is the functional equation ($EM_{HL}(u,v)$) for edge mask extraction of the HL sub-band.

$$EM_{HL}(u,v) = \begin{cases} 1, \text{ if } WC_{HL}(u,v) \geq Edge\ Threshold \\ 0, \ otherwise \end{cases} \qquad \cdots (1)$$

Here, $WC_{HL}(u,v)$ means the wavelet coefficient in the HL sub-band; Edge Threshold means the critical value of the edge. The functional equation for edge threshold is as follows.

$$Edge\ Threshold = MinWC_{HL}(u,v) + (MaxWC_{HL}(u,v) - MinWC_{HL}(u,v)) \times 0.488 \qquad \cdots (2)$$

$MaxWC_{HL}(u,v)$ means the largest wavelet value in the HL sub-band, and $MinWC_{HL}(u,v)$ means the smallest wavelet value in the HL sub-band, while, 0.488 is the edge weight value; the edge was best displayed when the value was 0.488. The edge threshold is shown differently depending on the sub-band. The edge information from the three directions are put together to form the final edge mask information. The following is the functional equation for final edge mask information.

$$EM(u,v) = EM_{HL}(u,v) \text{ or } EM_{LH}(u,v) \text{ or } EM_{HH}(u,v) \qquad \cdots (3)$$

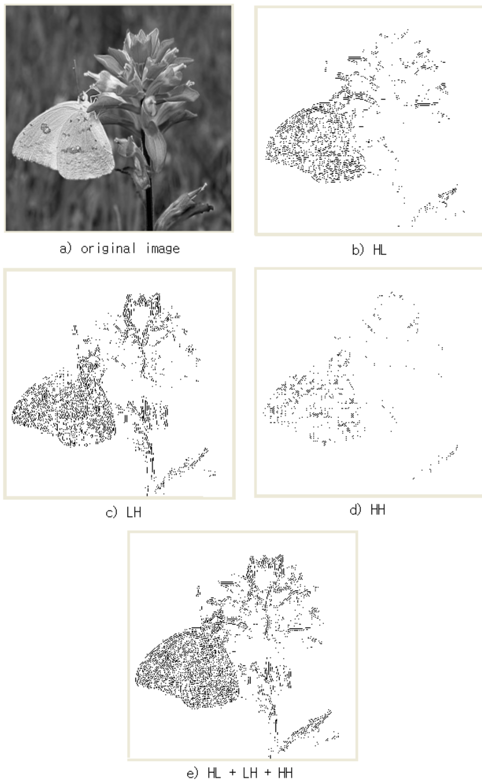


그림 4. 2 레벨 서브 밴드에 의해 생성된 최종 에지 마스크
Fig 4. Final edge mask made by two-level sub-bands

Figure 4 displays the final edge mask results using level 2. a) represents the edge mask information of HL sub-band which is the vertical direction edge, b) represents the LH sub-band in the horizontal direction edge, c) represents the HH sub-band in the diagonal direction edge, and d) represents the final edge mask information which combines a), b), and c).

2. ROI Extraction

Generally, dynamic ROI coding is processed in code block units for faster processing. The code block size is chosen from 16×16, 32×32, or 64×64. The proposedchothod uses noise reduction and the edge code block dotection process in a code block unit for fast ROI extraction. The following is the proposed process for automatic ROI extraction of this study.

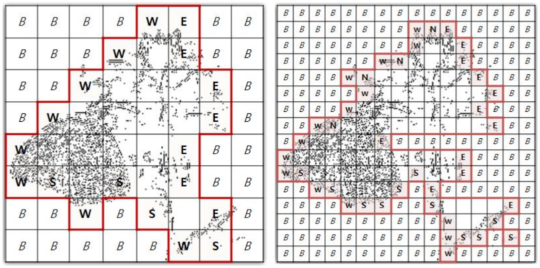
(1) The final edge information is scanned to 4 directions (north, south, east, west) in code block units, and the code block detection processing is done depending on the amount of edge information. If the edge information in the code block unit is smaller than the code block critical value, it is distinguished as background code block for the noise reduction process. Next is the functional equation ($EMCB(x,y)$) which finds out whether the edge of the (x,y) code block unit is distinguished. If the value is 1, it means the edge code block, and 0 means the code block that is not edge (background code block).

$$EMCB(x,y) = \begin{cases} 0, \text{ if } \sum_{u,v \in B_j} EM(u,v) < Edge\ Codeblock\ Threshold \\ 1, otherwise \end{cases} \quad (4)$$

B_j is the j-th code block, $EM(u,v)$ is the edge mask information in the code block, Edge code block threshold is the critical value which decides edge code block distinction, set at 10%. The bigger this number, the less the chance of edge code block distinction, and the smaller the number, the greater the chance. If the value was extreme with 0%, all will be distinguished as edge code block, and the whole image will be ROI-extracted.

(2) If the edge code block is found, the scan moves on to the first code block of the next line.

(3) After the four-way scanning is complete, the distinguished mask information is categorized into background, edge, and unscanned code block. Here, the edge and the unscanned code block are distinguished by ROI.



a) 64 × 64 b) 32 × 32

그림 4. 코드 블록 크기에 따른 자동 관심영역 추출

Fig 5. Automatic ROI extraction according to code block size

Figure 5 shows the results of automatic ROI extraction from a 512 ×512 size image in code block units of 64×64 and 32×32. B is the background, and W, S, E, N each stand for the directions of the distinguished edge code blocks. This algorithm is able to extract without additional calculations in the case of multi-ROIs.

3. The preferred processing of the ROI

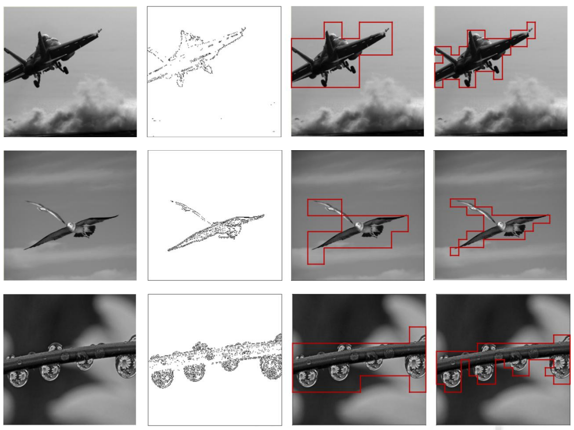
The preferred processing of the proposed coding method is the same as the Implicit ROI coding[15] expected for the ROI extraction method. In EBCOT, each quality layer comprises an arbitrary contribution from the embedded bit stream of each code block (packet or precinct) of each sub-band. Therefore, the preferred processing of the ROI is possible by applying relatively larger contributions from code blocks involved in the ROI reconstruction to quality layers. Since EBCOT allocates the code block contributions in terms of the overall distortion minimization, the ROI encoding allocates code block contributions in accordance with distortion reduction and the ROI. As a result, the distortion measure becomes

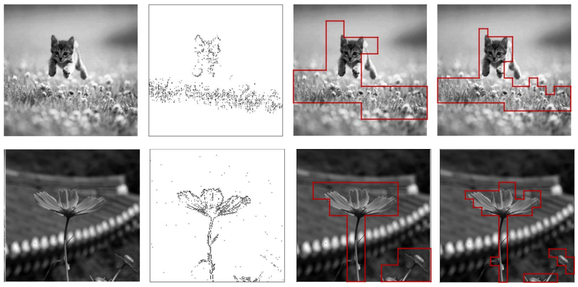
$$D_j^{n_i} = \begin{cases} W_{ROI}\, W_b \sum_{u,v \in B_j} \overline{(a^{n_i}(u,v) - a(u,v))^2}, & ROI\ code\ block \\ W_b \sum_{u,v \in B_j} \overline{(a^{n_i}(u,v) - a(u,v))^2}, & others \end{cases} \quad (5)$$

In Equation 5, W_{ROI} is the weight of the ROI, $D_j^{n_i}$ is the weighted MSE loss in n_i, W_b is the weight of sub-band including B_j, B_j is the j-th code block, $a(u,v)$ is the coefficients in the code block, $\overline{a^{n_i}(u,v)}$ is the quantized coefficients by n_i, and n_i is the truncation point. The decoding of the proposed coding method is the same as the Implicit ROI coding or the JPEG2000 part 1 baseline decoder. Note that any JPEG2000 part 1 compliant decoder can correctly decode the bit stream compressed by the proposed method.

Ⅳ. Experiments and Evaluations

In our experiments we tested a variety of methods including: the EBCOT method where any ROI are not coded, the Maxshift method, and the proposed method. The first two methods were evaluated by using Kakadu software V3.0.7 [17], and the proposed method was evaluated by modifying the Kakadu software. For fair experimental conditions, we set W_{ROI} to 4096. In the proposed method, the edge weight is 0.488, and the edge code block threshold is 10%. The experimental images are shown in Figure 6 a) as gray JPEG2000 images (8bpp).



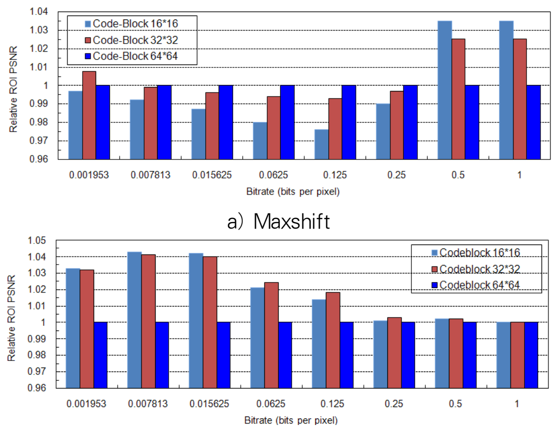


a) Original image b) Edge mask information c) 64×64 d) 32×32
그림 6. 자동 관심영역 추출 결과
Fig 6. Automatic ROI extraction results

Figure 6 shows the edge mask information and ROI extraction results according to code block size. Results from 32×32 unit ROI extraction show a more precise result than 64×64. The last images show that the multi-ROI is also well extracted.

a) Original image b) EBCOT c) Maxshift d)Proposed
그림 7. 다양한 코딩 방법을 사용한 재구성된 이미지
Fig 7. Reconstructed images using different coding methods

Figure 7 shows reconstructed images using different coding methods in 0.25bpp. The original images are shown in a). Images b) show us blurred versions of the original images, indicating that no ROI coding was executed. In images c), we can only see the ROI part in high quality since the Maxshift method codes the backgrounds after processing all of the ROI. In d), we can see the ROI in high quality, whereas the background shows poor quality. Finally, images d) look superior to images a) and b) in subjective evaluation of the whole image at low bit rates.



a) Maxshift

b) Proposed

그림 8. 64×64 ROI PSNR 기준으로 다양한 크기의 ROI PSNR 차이
Fig 8. Difference of ROI PSNR of different sizes(32×32, 16×16) on the basis of 64×64

Figure 8 a) illustrates the effect of code block size on the ROI performance for the Maxshift method and the proposed method. The ROI performance using a code block size of 64×64 is generally superior, especially at low bit rates, to the performances of those using reduced code block sizes. At higher bit rates (i.e. 0.5 and 1 bpp), smaller code block size begins to perform better, since its better spatial localization of the ROI compensates for the overhead in coding an increased number of code blocks. However, at these higher bit rates, the ROI becomes visually (or near visually) lossless regardless of the code block size.

Figure 8 b) which is Implicit coding applying for the proposed ROI extraction algorithm can be seen that an increased ROI performance can be achieved with a reduced code block size. The use of smaller code block size allows for increased spatial refinement. The final code stream bit rates, however, are detrimentally affected due to the overhead from an increased number of code blocks that needs to be represented in the code stream. It can be shown that Proposed using 32 ×32 code blocks generate code streams with larger final bit rates than that obtained by the max-shift using 64×64 code blocks. Smaller code block sizes should be used when the ROI is of primary visual importance at lower bit rates.

〔18〕 repeatedly compares the edge information by the unit of neighbor pixel, deletes the noise, classifies the ROI and codes them. So it is possible to handle the precise ROI, real time execution can not be possible caused by the increasing time complexity. On the contrary, the proposed method is possible to process fastly, since it identifies and codes independently ROI code blocks by the edge threshold using the edge information by the unit of code block.

Ⅴ. Conclusion

In this study, the DWT characteristics of JPEG2000 were used to propose a revised dynamic ROI coding method based on automatic ROI extraction. The proposed method uses edge critical value and edge code block critical value for noise reduction and the edge distinction process; and for a faster ROI extraction, a four-way scan was done to search the edge code block that is first contacted for processing. Various experiments showed that the proposed method functioned the best under a low bit-rate with code block size 32×32. The proposed method is available for the interactive environments. Later studies will extend the applicability of the proposed algorithm to MJPEG2000.

REFERENCES

[1] Kong H-S, A. Vetro, T. Hata, N. Kuwahara, "Fast Region-of-Interest transcoding for JPEG2000 images," IEEE International Symposium on Circuits and Systems, vol.2, pp. 952-955, 2005.

[2] Zhang Li-bao, Wang Ke, "New regions of interest image coding and its applications for sensing image," International Conference on Microwave and Millimeter Wave Technology, pp.952-956, 2004.

[3] Rene Rosenbaum, Heidrun Schumanm, "Flexible, dynamic and compliant Region of Interest coding in JPEG2000," IEEE The International Conference on Image Processing, vol.3, pp.101-104, 2002.

[4] Changick Kim, "Segmenting a low depth-of-field image using morphological filters and region merging," IEEE Transactions on Image Processing, vol. 14, pp.1503-1511, 2005.

[5] Chen, O.T.-C., Chih-Chang Chen, "Automatically-Determined Region of Interest in JPEG 2000," IEEE Transactions on Multimedia, vol.9, pp. 1333-1345, 2007.

[6] Fei Wang, Jishang Wei, "Nanning Zheng, Shaoyi Du, Bin Gao, Automatic ROI Selection for JPEG2000 Compression of Remote Sensing Images," International Conference on Semantic Computing, pp. 615-621, 2007.

[7] C. S. Won, K. Pyun, R. M. Gray, "Automatic Object Segmentation in Images with Low Depth of Field," In Proceedings of IEEE International Conference on Image Processing, vol. 3, pp. 806-808, 2002.

[8] J. Z. Wang, J. Li, R. M. Gray, G. "Wiederhold, Unsupervised Multiresolution Segmentation for Images with Low Depth of Field," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 85-90, 2001.

[9] R. M. Harralick, L. G. Shapiro. Survey, "Image Segmentation Techniques," Computer Vision, Graphics and Image Processing, vol. 29, pp. 100-132, 1985.

[10] P. K. Sahoo, S. Soltani, A. K. C. Wong, " *A Survey of Thresholding Techniques*" Computer Vision, Graphics and Image Processing, vol.41, pp. 233-260, 1988.

[11] S. D. Kang, H. W. Yoo, D. S. Jang, "Color Image Segmentation Based on the Normal Distribution and the Dynamic Thresholding," Lecture Notes in Computer Science, vol. 4705, pp. 374-384, 2007.

[12] Q. Gao, "Extracting Object Silhouettes by Perceptual Edge Tracking," In Proceedings of IEEE International Conference on Systems, vol. 3, pp. 2450-2454, 1997.

[13] S. Mahamud, L. R. Williams, K. K. Thornber, K. Xu, "Segmentation of Multiple Salient Closed Contours from Real Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 433-444, 2003.

[14] H. Yang, M. Long, H. M. Tai, "Region-of-Interest image coding based on EBCOT," IEE Proceedings-Vision, Image, and Signal Processing, vol. 152, pp. 590-596, 2005.

[15] Martin Boliek, Charilas Christopoulos, "JPEG 2000 part I final committee draft version 1.0," ISO/IEC JTC 1/SC 29/WG 1 N1646R, 2000.

[16] Victor Sanchez, Anup Basu, Mrinal K. Mandal, "Prioritized region of interest coding in JPEG2000," IEEE Transaction on Circuits and Systems for Video Technology, vol.14, no.9, 2004.

[17] D. Taubman:Kakadu software, "A comprehensive framework for JPEG2000," *http://www.kakadusoftware.com/.*

[18] 박순화, 서영건, 이부권, 강기준, 김호용, 김형준, 김상복, "피사계 심도가 낮은 이미지에서 웨이블릿 기반의 자동 ROI 추출 및 마스크 생성," 한국컴퓨터정보학회 논문지, 제14권, 제 3호, 93-101쪽, 2009년 3월.

저 자 소 개

박 재 흥

1978년 : 충북대학교 수학교육과 학사
1980년 : 중앙대학교 전산학과 석사
1989년 : 중앙대학교 전산학과 박사
1983년~현재 :
경상대학교 컴퓨터과학과 교수
관심분야 : Software Testing,
JPEG2000, Software Process

김 현 주

2000년 8월 :
경상대학교 컴퓨터과학과 공학박사
2002년~현재 :
진주산업대학교 컴퓨터공학부 부교수
관심분야 : 정보검색, XML, 컴퓨터교육

심 종 채

1987년 : 경상대학교 학사
1987년 : (주)정원시스템 SE
1989~1998년 : 호남석유화학(주) 전산실
1995년 : 중앙대학교 석사
2002년 : 경상대학교 컴퓨터과학과 박사
1998~현재 : 경남도립남해대학
인터넷정보기술과 교수
관심분야 : 데이터베이스, 멀티미디어,
가상대학

유 창 열

1987년 : 경상대학교전산통계학과 학사
1993년 : 경상대학교 전자계산학과 대학원 석사
1999년 : 경상대학교 전자계산학과 대학원 박사
1996년 3월~현재 : 남해전문대학 인터넷 정보기술과 교수
관심분야 : 시스템분석, ERP 성능평가,
컨설팅 등

서 영 건

1987년 : 경상대학교전산통계학과학사
1989년 : 숭실대학교 대학원 전자계산학과 석사
1997년 : 숭실대학교대학원 전자계산학과 박사
1989년~1992년 : 삼보컴퓨터 근무
1997년~현재 : 경상대학교 컴퓨터교육과 부교수
2001년~현재 : 경상대학교 컴퓨터정보통신 연구소원
관심분야 : 멀티미디어통신, 영상인식, 원격교육

강 기 준

2000년 : 경상대학교 컴퓨터과학과학사
2002년 : 경상대학교 교육대학원 컴퓨터교육과 석사
2007년 : 경상대학교 대학원 컴퓨터과 학과 박사
관심분야 : JPEG2000, MPEG, 원격교육