

논문 2008-6-8

모바일 GIS를 위한 셀 기반의 공간 데이터 압축 기법

A Cell-based Compression Technique of the Spatial Data for the Mobile GIS

이기영*, 임 근**, 최규석***

Ki-Young Lee*, Keun Lim**, Gyoos Seok Choi***

요 약 최근에 무선 통신 기술 및 지리정보시스템(GIS)의 발달과 더불어 모바일 GIS에 대한 관심이 높아지고 있다. 이러한 모바일 GIS는 기존의 서버 컴퓨팅 환경과는 다르게 자원이 매우 제한적인 모바일 환경에 최적화되어야 한다. 특히, 모바일 GIS에서는 대용량 공간 데이터를 모바일 장치에 적합하게 줄이는 것이 매우 중요하다. 그러나 기존의 공간 데이터 압축 기법은 데이터를 가변적으로 처리하는 기법으로써 데이터의 크기를 예측하기가 매우 힘들며, 순차적으로 처리해야만 하는 단점이 존재한다. 따라서, 본 논문에서는 모바일 GIS를 위한 셀기반 공간 데이터 압축 기법을 제시한다. 본 논문에서 제시하는 셀기반 공간 데이터 압축 기법은 모바일 환경에서 클라이언트로 공간 데이터를 제공할 때 셀 단위로 전송하는 것을 기반으로 하여 압축하는 방법이다. 마지막으로, 본 논문에서는 실제 데이터를 통한 성능 평가를 수행함으로써 셀기반 공간 데이터 압축 기법의 우수성을 입증하였다.

Abstract Recently, with the development of wireless communications and GIS, interest about mobile computing is rising. In this case, the GIS needs to be placed in a constricted environment than the environment of server computing. For this reason large amounts of spatial data must be compressed into the mobile device. The compression technique of the spatial data is difficult and must be processed in the correct order because the size of the data is unpredictable. Therefore, in this paper, the cell-based compression technique of the spatial data for mobile GIS is presented. This paper includes the process of transformation of spatial data from a certain server to a mobile device by cell-based compression technique. In this paper, the cell-based compression technique of the spatial data has been proven practically efficient.

Key Words : the cell-based compression technique, mobile GIS, spatial data

1. 서 론

최근 지리 정보 시스템 기술이 발전함에 따라 위치 기반 서비스(LBS:Location-Based Services), 텔레매틱스 등과 같이 공간 데이터를 신속하고 효율적으로 사용자에게 제공하는 다양한 지리정보시스템(GIS:Geographic Information System) 서비스가 늘어나고 있다^{[1][2]}. 특히,

정보통신 기술이 발전함에 따라 편리성과 신속성을 추구하기 위해 모바일 장치를 통한 업무처리의 활용이 다양한 분야에서 많이 요구되고 있다. 이러한 추세에 발맞추어 PDA, 인터넷 접속이 가능한 스마트 폰 등의 기술이 발전하자 이러한 모바일 장치에서 정보를 효율적으로 제공하기 위한 많은 연구가 수행되고 있다^{[3][4][5]}.

모바일 데이터베이스는 이러한 모바일 장치에서 운영되는 데이터베이스를 말한다. 이는 단순히 서버와 연결을 통해 질의 수행하는 클라이언트를 말하는 것이 아니라, 서버로부터 제공받은 데이터를 모바일 장치에서 효율적으로 처리할 수 있어야 함을 의미한다^[6]. 특히, 모바일

*정회원, 을지대학교 의료산업학부

**정회원, 을지대학교 의료산업학부(교신저자)

***중신회원, 청운대학교 컴퓨터학과

접수일자 2008.9.25, 수정완료 2008.12.7

일 GIS에서는 모바일 장치에서 GIS에 필요한 공간 데이터를 저장 및 처리하는 것이 필요하다^{[5][7]}.

현재 모바일 장치는 기존 PC에 비해 낮은 성능의 프로세서와 작은 용량의 저장 공간을 사용하고 있으며, 자체적인 화일 시스템에서 데이터를 처리 및 저장하고 있다^[8]. 이는 작은 용량의 데이터를 처리하기는 편리할지 모르나 나날이 증가하고 있는 대용량의 공간 데이터를 저장하고 매우 복잡하고 다양한 공간 질의를 처리하기에는 효율적이지 못하다^{[9][10]}. 그러므로 모바일 GIS에서 활용되는 대용량의 공간 데이터를 모바일 장치에서 효율적으로 저장 및 처리하기 위해서는 서버의 대용량 공간 데이터를 모바일 장치에 적합하도록 공간 데이터를 변환 및 압축하는 기술이 반드시 필요하다.

따라서 본 논문에서는 모바일 GIS를 위한 셀기반 공간 데이터 압축 기법을 제시한다. 본 논문에서 제시하는 셀기반 공간 데이터 압축 기법은 일반적으로 서버에서 모바일 환경으로 공간 데이터를 제공할 때 질의 영역을 포함하는 셀 단위로 데이터를 전송하는 셀기반 저장 방식을 기반으로 공간 데이터를 압축하는 방법이다. 따라서 공간 데이터를 데이터 분포에 적합한 셀로 구성하며 각 셀의 기준점을 중심으로 상대 좌표를 이용하여 공간 데이터를 표현한다. 마지막으로, 본 논문에서는 실제 데이터를 통한 성능 평가를 수행하여 본 논문에서 제시한 셀기반 공간 데이터 압축 기법의 우수성을 입증하였다.

본 논문은 다음과 같이 구성되어 있다. 제 II장에서는 모바일 GIS와 기존의 공간 데이터 압축 기법을 설명한다. 제 III장에서는 본 논문에서 제시한 모바일 GIS를 위한 셀기반 공간 데이터 압축 기법에 대해서 구체적으로 설명한다. 제 IV장에서는 성능 평가를 통해 본 논문에서 제안하는 셀기반 공간 데이터 압축 기법의 우수성을 입증하고, 마지막으로 제 V장에서는 결론에 대하여 언급한다.

II. 관련 연구

본 장에서는 모바일 GIS와 기존의 공간 데이터 압축 기법인 산술 연산 코딩 기법에 대하여 설명한다.

1. 모바일 GIS

모바일 컴퓨팅이란 사용자가 이동하는 환경에서 컴퓨

터 자원을 사용하는 것을 말한다. 따라서 일반 PDA나 스마트 폰은 물론 노트북이나 차량 내부에 설치되어 있는 컴퓨터를 활용하는 것도 모바일 컴퓨팅이라고 할 수 있다^{[11][12][13]}. 모바일 GIS란 모바일의 개념과 GIS의 개념이 합쳐진 것으로 PC 환경이 아닌 모바일 컴퓨팅 환경에서 운영되는 GIS 솔루션을 의미한다.

모바일 GIS의 초기에는 성능이 미비한 모바일 컴퓨터에 국한되어 단순한 사용자의 위치 정보만을 제공해 왔지만, 현재는 휴대용 단말기, 무선 인터넷, GPS 장치 등의 발전으로 PDA나 스마트 폰에 이르기까지 점차 그 활용 범위 및 제공되는 정보가 확대되어 가고 있다.

또한, 모바일 GIS 서비스는 무선 데이터 통신 기술과 GPS의 위치 추적 기능을 이용하여 화물 위치 추적 서비스, 교통 정보, 지역 정보 등이 실시간으로 요구되는 다양한 분야에서 활용될 수 있다. 또한, 향후 유비쿼터스 컴퓨팅 환경에서는 다른 다양한 서비스와 결합되어 활용될 것으로 예상되고 있다. 이러한 모바일 GIS에서는 한정된 자원에서 데이터를 처리하는 것이 매우 중요하며, 이를 해결하기 위해 많은 연구가 수행되었다.

2. 산술 연산 코딩 기법

본 논문에서 제시한 공간 데이터 압축 방법과 비교하기 위해 기존에 많이 사용되던 산술 연산 코딩 기법을 적용하였다^[14]. 일반적으로 GIS에서 가장 많은 크기를 차지하는 것은 점(Point)이 아닌 선(MultiLine)이나 면(Polygon)이다. 특히, 이러한 선이나 면은 제한된 영역에 클러스터링된 형태를 띠게 된다. 그러므로 이러한 선이나 면을 특정 기준점을 기준으로 하여 거리의 차이를 구하게 되면 일반적으로 8바이트를 차지하는 실제 데이터보다 그 값이 매우 작아지게 된다.

예를 들어 (123456, 123456), (123457, 123456), (123457, 123457), (123458, 123458) 등과 같은 5개의 점으로 이루어진 면이 있을 경우 맨 처음 값인 123456과 이어지는 다음 X 좌표 값들의 차는 1, 0, 1과 같은 매우 작은 값을 가지게 된다. 이러한 차이값을 저장함으로써 실제 데이터에 비해 값을 저장하기 위해 필요한 공간을 충분히 줄일 수 있게 되며, 작아진 값을 저장할 때 길이 플래그를 이용함으로써 저장할 값에 따라 매우 작은 크기를 이용해 데이터를 저장할 수 있다. 그림1은 산술 연산 코딩 기법에서 활용되는 좌표의 저장 구조를 보여준다.

부호 플래그 길이 플래그 차이 값

그림 1. 좌표 저장 구조
Fig. 1. Coordinate Storage Structure

그림 1에서 보듯이, 맨 앞의 플래그는 부호 플래그로써 양수인지 음수인지를 표시하고, 길이 플래그는 실제 값이 차지하는 비트의 크기를 표시한다. 즉, 길이 플래그 값이 1이면 실제 값이 4비트임을, 2이면 8비트, 3이면 12비트, 4이면 16비트임을 나타낸다. 각 비트 단위마다 값은 15, 255, 4095, 65535의 크기까지 표현할 수 있다. 따라서 길이 플래그의 기본 크기는 차이값의 최대 크기에 따라 설정되어야 한다.

산술 연산 코딩 기법은 지정된 기준 점이 변경될 때마다 전체 데이터를 다시 계산 해주어야 하는 단점이 존재하며, 데이터의 크기가 가변적이므로 직접 접근에 매우 어렵다는 단점이 발생한다.

III. 셀 기반 공간 데이터 압축 기법

본 장에서는 모바일 GIS를 위한 셀 기반 압축 기법에 대해서 설명한다. 셀 기반 공간 데이터 압축 기법은 일반적으로 서버에서 모바일 환경으로 공간 데이터를 제공할 때 요구하는 영역을 포함한 셀 단위로 전송하는 것을 기반으로 하여 압축하는 방법이다.

따라서, 셀 기반 공간 데이터 압축 기법은 셀 단위로 압축되며, 크게 압축된 셀의 기본 정보를 저장하기 위한 헤더와 실제 셀에 포함된 공간 데이터를 압축한 바디로 구성된다. 그림 2는 셀 정보를 저장하기 위한 헤더 정보와 그 예를 보여준다.

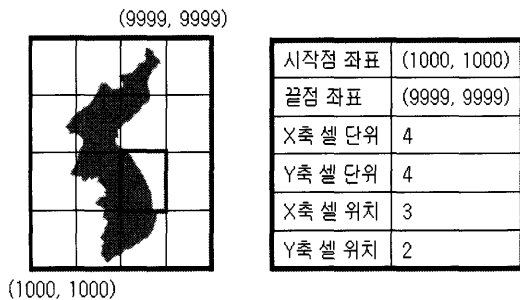


그림 2. 셀과 헤더 구조의 예
Fig. 2. Example of Structure of Header and Cell

그림 2에서 왼쪽 그림은 공간 데이터에 대해서 두 기준점을 기준으로 4X4로 구성된 셀 분할의 예를 보여준다. 그림 2에서 오른쪽 그림은 왼쪽 그림에서 진한색 사각형을 이용해 구성되는 셀의 헤더 정보를 보여준다.

그림 2의 오른쪽 그림에서 보듯이 셀의 압축 헤더는 시작점 좌표, 끝점 좌표, X축 셀 단위, Y축 셀 단위, X축 셀 위치, Y축 셀 위치를 저장한다. 시작점과 끝점은 전체 공간 데이터의 크기를 나타내며, X축과 Y축 셀 단위는 전체 공간 데이터에서 각 셀 단위의 분할 횟수를 의미한다. 이 때, 시작점, 끝점, X축 셀 단위, Y축 셀 단위가 고정적인 경우, 즉 크기가 변화되지 않는 경우에는 한번만 전송하고 다음부터 헤더에 정보를 포함하지 않음으로써 압축된 셀의 헤더의 크기를 더욱 줄일 수 있다. 마지막으로 X축과 Y축 셀 위치는 각각의 축을 분할 했을 때 몇 번째 셀인지를 나타낸다.

각 셀을 압축하기 위해서는 각 공간 데이터가 포함된 셀의 시작점(c.start)과 끝점(c.end) 값을 얻어내야 되며, 이를 구하는 공식은 아래와 같다.

$$c.start.x = start.x + \left(\frac{(x-1)(end.x - start.x)}{cx} \right)$$

$$c.start.y = start.y + \left(\frac{(y-1)(end.y - start.y)}{cy} \right)$$

$$c.end.x = start.x + \left(\frac{(x)(end.x - start.x)}{cx} \right)$$

$$c.end.y = start.y + \left(\frac{(y)(end.y - start.y)}{cy} \right)$$

위의 식은 헤더에서 시작점과 끝점의 X축 좌표와 Y축 좌표를 각각 start.x, start.y, end.x, end.y라고 하고 X축과 Y축의 셀 단위를 cx, cy라고 할 경우 X축과 Y축 셀의 위치 값이 x, y일 때의 시작점의 X축 좌표(c.start.x), Y축 좌표(c.start.y), 끝점의 X축 좌표(c.end.x), Y축 좌표(c.end.y)를 구한다.

셀 단위로 압축할 때 실제 셀에 포함된 공간 데이터를 압축한 바디는 셀 별로 시작점과 끝점을 기준으로 하여 셀 내부에서 셀을 분할 한 후 해당 좌표에 가장 가까운 셀의 상대 위치로 표현된다. 이때, 압축된 공간 데이터의 정밀도는 셀 내부를 분할하는 횟수로 인해 정해지며 이는 추후 공간 데이터를 저장하는 기본 크기가 된다. 즉, 분할 횟수가 256일 경우 하나의 좌표를 저장하는데 2Byte가 필요하게 된다.

그림 3은 공간 데이터를 압축하는 모습을 보여준다.

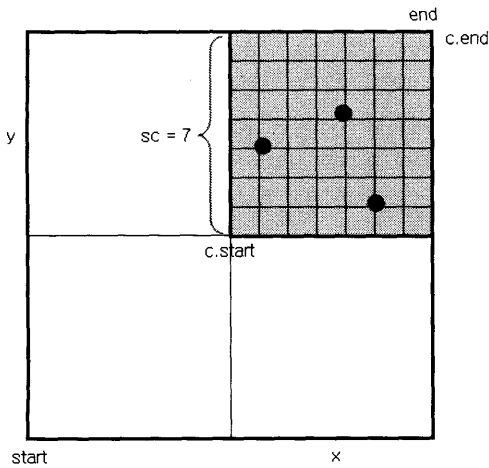


그림 3. 공간 데이터 압축의 예
Fig. 3. Example of Compression of Spatial Data

그림 3에서 보듯이 각 셀을 정해진 크기로 분할하며 공간 데이터를 분할된 내부 셀의 가장 가까운 교차점값으로 대응 시킨다. 압축된 공간 데이터(c) 값을 얻어내야 되며, 이를 구하는 공식은 아래와 같다.

$$c.x = \frac{a.x - c.start.x}{\left(\frac{c.end.x - c.start.x}{sc}\right)}$$

$$c.y = \frac{a.y - c.start.y}{\left(\frac{c.end.y - c.start.y}{sc}\right)}$$

위의 식은 앞에서 구한 공간 데이터를 포함하는 셀의 시작점과 끝점을 이용해 공간 데이터의 좌표를 a.x, a.y 라고 하고 내부 셀의 분할 횟수가 sc라고 할 때 압축한 새로운 공간 데이터의 압축 좌표(c.x, c.y)를 구한다. 압축된 공간 데이터는 셀의 바디 부분에 공간 데이터의 식별자와 같이 저장된다.

IV. 성능 평가

본 논문에서 성능 평가는 CPU 펜티엄 콘로 1.6GHz, 메모리 1GB, 그리고 Redhat 9.0 환경에서 수행되었다. 성능 평가에 사용된 공간 데이터는 약 250,000개의 2차원 공간 데이터인 서울시의 건물 데이터를 사용하였으며, 전체 크기는 약 40MB이다. 본 논문에서는 산술 연산 코

딩 기법과 셀 기반 압축 기법을 이용해 인덱스 크기와 접근 속도를 비교하였다.

그림 4는 공간 데이터를 압축하였을 경우 실제 데이터, 산술 연산 코딩 기법, 셀 기반 압축 기법의 크기를 비교한 것이다.

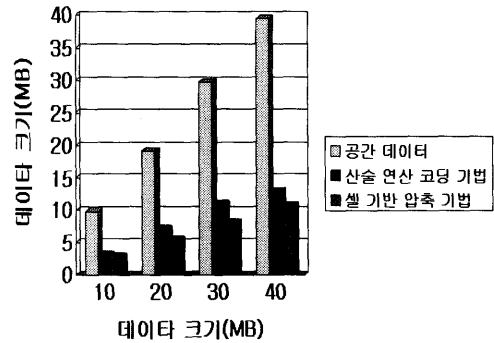


그림 4. 공간 데이터 크기에 따른 압축률 비교
Fig. 4. Comparison of Compression Rate by Spatial Data's Size

그림 4에서 보여주는 바와 공간 데이터의 압축 효율이 기존의 산술 연산 코딩 기법보다 압축률이 약 10% 정도 향상됨을 알 수 있다. 또한, 산술 연산 코딩 기법에 비해 셀 기반 압축 기법은 크기가 고정적이므로 압축 크기가 데이터에 크기에 명확히 비례함을 알 수 있다.

그림 5는 공간 데이터를 압축하였을 경우 산술 연산 코딩 기법, 셀 기반 압축 기법의 압축 속도를 비교한 것이다.

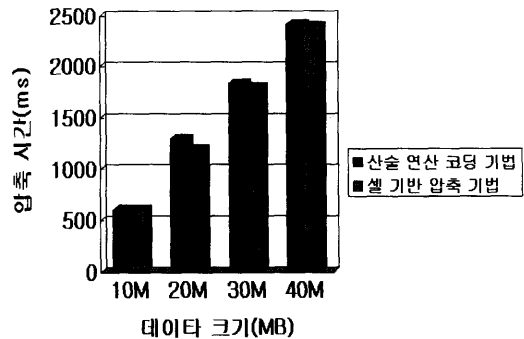


그림 5. 공간 데이터 크기에 따른 압축 시간 비교
Fig. 5. Comparison of efficiently Compression Time by Spatial Data's Size

그림 5에서 보여주는 바와 같이 셀 기반 압축 기법의 공간 데이터의 압축 시간은 기존의 산술 연산 코딩 기법과 거의 유사함을 알 수 있다. 이는 산술 연산 코딩 기법이 계산식이 단순하지만 비트 압축 시간이 걸리는 반면 셀 기반 압축 기법은 계산식이 더 복잡하기 때문이다.

그림 6은 40M의 공간 데이터를 저장한 경우 질의 영역이 각각 전체의 1%, 5%, 10%, 일 때 산술 연산 코딩 기법, 셀 기반 압축 기법에 대해서 데이터 검색 시간을 비교한 것을 보여준 것이다.

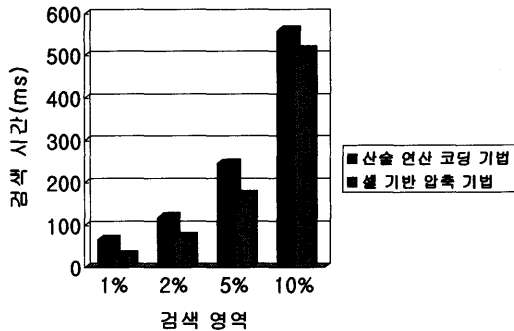


그림 6. 질의 영역에 따른 데이터 검색 시간 비교
Fig. 6. Comparison of Data Search Time by Query Region

그림 6에서 보여주는 바와 같이 각각의 질의 영역이 작을수록 셀 기반 압축 기법이 좀 더 효율적임을 알 수 있다. 이는 셀 기반 압축 기법의 경우 크기가 고정이므로 랜덤 접근에 그만큼 효과적임을 알 수 있다.

V. 결론

최근에 무선 통신 기술과 지리정보시스템의 발달과 더불어 모바일 GIS에 대한 관심이 높아지고 있다. 이러한 모바일 GIS는 기존의 서버 컴퓨팅 환경과는 다르게 자원이 매우 제한된 모바일 환경에 적합하여야 한다. 이를 위해 특히 대용량인 공간 데이터를 모바일 장치에 적합하게 줄이는 것이 매우 중요하다. 그러나 기존의 공간 데이터 압축 기법은 데이터를 가변적으로 처리하는 기법으로써 데이터의 크기를 예측하기가 매우 힘들며, 순차적으로 처리해야만 하는 단점이 존재한다.

따라서, 본 논문에서는 모바일 GIS를 위한 셀 기반 공

간 데이터 압축 기법을 제시하였다. 이 방법은 공간 데이터를 데이터 분포에 적합한 셀로 구성하며 각 셀의 기준점을 중심으로 상대 좌표를 이용하여 공간 데이터를 표현한다. 마지막으로, 본 논문에서는 실제 데이터를 통한 성능 평가를 수행함으로써 여 본 논문에서 제시한 셀 기반 공간 데이터 압축 기법의 우수성을 입증하였다.

참고 문헌

- [1] Bar-Noy, A., Kessler, I., "Tracking Mobile Users in Wireless Communications Networks", *IEEE Transactions on Information Theory*, Vol.39, pp. 1877-1886, 1993
- [2] Bar-Noy, A., Kessler, I., Sidi, M., "Mobile Users: To Update or not to Update?", *ACM/Baltzer Wireless Networks Journal*, Vol.1, No.2, pp. 175-185, 1995
- [3] Lee, K.Y., Kim, D.O., Yun, J.K., and Han, K.J., "A Real-time Mobile GIS based on the HBR-tree," *Proc. of the 33rd Int. Conf. on Computers & Industrial Engineering*, 2004
- [4] Yun, J.K., Kim, D.O., and Han, K.J., "Development of a Real-Time Mobile GIS supporting the Open Location Service," *Proc. of Geotec Event Conference, Canada*, 2003
- [5] 윤재관, 장영승, 한기준, "모바일 GIS를 위한 위치 기반 서비스," *한국정보과학회 데이터베이스 연구*, 18권 1호, pp.3-15, 2002
- [6] Lee, D.L., Zhu, M., H. Hu, "When Location-Based Services Meet Database", *Mobile Information System*, Vol.1, No.2, pp. 1-21, 2003
- [7] 양영규, "위치기반 서비스(LBS: Location Based Service)기술 현황 및 전망," *한국정보처리학회 정보처리학회지*, 8권 6호, pp.4-5, 2001
- [8] Kang, H.Y., Hwang, S.Y., Han, D.C., Li, K.J., "Framework Design for Ubiquitous GIS", *Proceedings of Korea Information Processing Society Conference*, Vol.11, No.1, pp. 1-4, 2004
- [9] Greene, D., "An Implementation and Performance Analysis of Spatial Data Access Methods," *IEEE*

Transaction on Knowledge and Data Engineering, Vol.5, No.1, pp.606-615, 1989

[10] Gueting, R.H., "An Introduction to Spatial Database Systems," The VLDB Journal, Vol.6, No.1, pp.357-399, 1994

[11] 윤재관, 김동오, 한기준, "LBS를 위한 실시간 GIS 엔진의 설계 및 구현," 한국정보과학회 학술발표 논문집, Vol.29, No.2, pp.244-246, 2002

[12] 윤재관, 장유정, 한기준, "대용량 위치 데이터를 위한 분산 위치 저장 컴포넌트의 개발," 한국정보과학회 데이터베이스 연구, Vol.18, No.4, pp.67-80, 2002

[13] 이계영, 권동섭, 송병호, 이석호, "이동 객체의 궤적 검색을 위한 인덱스 구조," 한국정보과학회 학술발표논문집, Vol.28, No.2, pp.217-219, 2001

[14] 조형주, 정진완, "다차원 색인 구조를 위한 효율적인 압축 방법," 한국정보과학회 논문지, 30권 5호, pp.429-437, 2003

저자 소개

이 기 영(정회원)



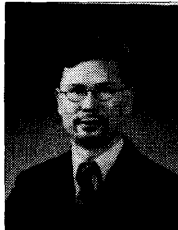
- 1984년 숭실대학교 전자계산학과 학사 졸업.
 - 1988년 건국대학교 대학원 컴퓨터공학과 석사 졸업.
 - 2005년 건국대학교 대학원 컴퓨터공학과 박사 졸업
 - 1984년~1991년 한국해양연구원 연구원
 - 1996년~1998년 한국컴퓨터정보학회 이사 및 서울동부지회장
 - 1991년~현재 을지대학교 의료산업학부 교수
- <주관심분야 : 공간 데이터베이스, GIS, LBS, USN, 텔레매틱스>

임 근(정회원)



- 1998년 중앙대학교 대학원 컴퓨터공학과(공학박사)
 - 1992년 3월 - 현재 을지대학교 의료산업학 부교수
- <주관심분야 : S/W공학, CBD 방법론, HCI 등> cchlrbtjt

최 규 석(중신회원)



- 1982년 연세대학교(전기전자) 공학사
 - 1987년 연세대학교(전기전자) 공학석사
 - 1997년 연세대학교(전기전자) 공학박사
 - 1987년 1월~1997년 1월 (주)데이콤 정보통신연구소 연구원 및 (주)SK텔레콤 중앙연구원 책임연구원 근무, 1997년 ~현재 청운대학교 컴퓨터학과 교수
- <관심분야> : 이동통신, 인공지능, 인공생명, 지능형 교통체계(ITS), 이동 컴퓨팅