

유비쿼터스 환경에서의 시각문맥정보인식에 대한 연구

A Study on Visual Contextual Awareness in Ubiquitous Computing

한동주* · 김종복* · 이상훈** · 서일홍*

(DongJu Han · JongBok Kim · SangHoon Lee · IlHong Suh)

Abstract - In many cases, human's visual recognition depends on contextual information. We need to use effective feature information for performing vigorous place recognition to illumination, noise, etc. In the existing cases that use edge and color, etc., visual recognition doesn't cope effectively with real environment. To solve this problem, using natural marker, we improve the efficiency of place recognition.

Key Words : Ubiquitous, Steerable pyramid, PCA, global feature, place recognition

1. 서론

유비쿼터스의 어원은 라틴어로 '언제 어디서나 존재하는'이란 사전적 의미를 가지고 있다. 유비쿼터스 컴퓨팅 환경은 언제 어디서나 자신이 필요로 하는 정보를 얻을 수 있는 환경이다. 유비쿼터스 컴퓨팅 환경에서는 사용자가 필요로 하는 정보를 물리공간과 전자공간을 통해서 검색할 수 있다.

에이전트는 사전적 의미로 '대행자, 대리인'이란 의미를 가지고 있다. 에이전트는 컴퓨터 분야에서는 작업을 대신해 주며 분산 환경에서 작업을 수행하는 자율성을 갖는 응용 프로그램이라 정의한다.

유비쿼터스 환경 하에서 사용자에게 정보를 제공하기 위해 여러 정보보조시스템[1][2]들이 제안되어 왔다. 예를 들어 일정 관리, 상황 인식, 의료 정보 제공 등의 실생활에 도움을 줄 수 있는 것들이 많이 존재한다. 다가오는 미래 사회에 도시, 교육, 국방, 의료 등의 다양한 분야에서 유비쿼터스 환경에서의 정보보조시스템의 활용이 기대되고 있다.

인간은 시각, 청각, 후각, 미각, 촉각 등으로 정보를 취득하며 이 중 시각을 통하여 가장 많은 정보를 취득하고 있다. 시각 정보의 중요성과 영상을 취득하는 장비의 발달 및 처리 기술의 향상으로 인하여 항공, 지질, 해양, 국방, 토목, 환경 기상 등의 분야에서 영상처리에 대한 활용도가 증가하고 있고 연구가 활발히 진행 중에 있다.

위와 같이 활용도가 높은 시각 센서 분야에서 사용자에게 가장 도움을 줄 수 있는 것이 위치 정보이다. 만일 사용자가 낮은 환경에 있을 때 위치 정보를 알고 있다면 작업 행동을 쉽게 선택할 수 있을 것이다. 이를 기반으로 위치 인식을 수

행할 수 있는 시각문맥정보 시스템을 연구하게 되었다.

2. 정보보조시스템[3]의 구성

정보보조시스템(Information Assistant System)은 입력으로는 사용자의 환경 정보를 받고, 출력으로는 서비스 정보를 사용자에게 제공하도록 구성되었다. IAS의 구성은 아래 그림과 같다.

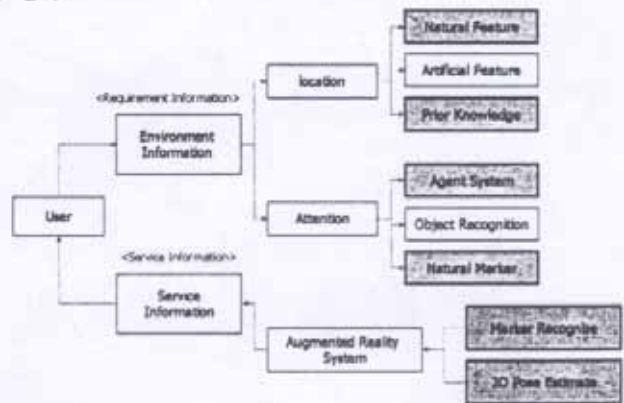


그림1. 정보보조시스템의 구성도

환경 정보는 현재 사용자의 위치(Localization)와 사용자가 주목하는 정보(Attention Information)로 구분할 수 있다. 위치와 사용자가 주목하는 정보는 Natural Marker와 Artificial Marker를 통해 알아내며, 위치 정보에서는 사전 지식을 사용하여 신뢰할 수 있는 정보를 얻는다. 주목하는 정보를 얻기 위해서는 환경 요소에 대한 물체 인식이 필요하다.

서비스 정보는 Marker 인식과 3차원 위치 계산을 통해서 증강 현실 장면을 구성해 사용자에게 제공한다. 사용자에게 제공할 정보는 3차원 객체형태로 변환된다.

[3]에서는 Topology Map과 Artificial Marker를 사용하여

저자 소개

- * 한 동 주 : 漢陽大學 情報通信工學科 碩士課程
- * 김 중 복 : 漢陽大學 情報通信工學科 博士課程
- ** 이 상 훈 : 漢陽大學 電子電氣制御計測工學科 博士課程
- * 서 일 홍 : 漢陽大學 情報通信工學科 教授 · 工博

장소 인식을 수행하였는데, Artificial Marker를 사용하는 경우 2차원 영상에서 Visual Marker를 찾아내서 인식하기 때문에 인식 범위가 Marker가 있는 영상으로 제약이 된다. 반면 Natural Marker를 사용하는 경우 영상이 가지고 있는 무늬의 형태를 보고 판단하기 때문에 비슷한 무늬를 갖는 영상은 같은 장소로 판단할 수 있으며, 영상의 제약 없이 새로운 환경에 대해서도 인식이 가능하다.

3. Natural Marker를 사용한 장소 인식

실세계 장면들의 규칙성은 일반적으로 장면 내에 존재하는 물체들을 확인하지 않고 scene properties와 관련된 feature들을 정의하는 것으로 제시한다. 어떤 scene feature들은 특정 장소를 인식하는 데 탁월한 성능을 발휘하지만 새로운 장소에 대해서 일반화할 수는 없다. 우리는 실험 환경에 대한 유연성을 배제하고 어떤 함수의 제약 조건으로써 표현될 수 있는 feature들을 사용할 것이다.

우리는 Steerable pyramid[4][6][7]를 적용하여 각각의 영상에 대하여 4가지 크기, 6가지 방향 성분에 대한 결과, 즉 하나의 영상에 대하여 총 24개의 흑백 영상을 얻을 수 있다. 이 영상들을 적절히 down-sampling하여 steerable pyramid에 의해서 생성된 24개의 영상에 대해서 4*4 해상도의 영상을 구성할 수 있으며 총 384바이트의 정보로 하나의 영상을 표현할 수 있다. 여기에 Principal Component Analysis (PCA)[4][5]를 적용하여 80개의 PCs를 이용함으로써 불필요한 redundancy를 없애 처리해야 할 데이터의 양을 줄여 영상을 효과적으로 표현할 수 있도록 하였다.

3.1 Steerable Pyramid

Steerable pyramid를 구성하기 위한 첫 단계로 steerable filter[8][9]를 구한다. steerable filter는 방향을 나타내는 기저 필터들의 선형조합으로 표현된 필터들로 구성된다. 2차원의 경우를 예로 들면 다음과 같다.

2차원 상의 Gaussian 함수 G 를

$$G(x, y) = e^{-(x^2+y^2)}$$

로 정의한다. Gaussian의 n 차 도함수를 G_n 이라 하고,

$f^\theta(x, y)$ 는 $f(x, y)$ 를 원점에 대해 각 θ 만큼 회전한 함수라 하자. Gaussian 함수의 x 축, y 축 방향 도함수는 각각 다음과 같다.

$$G_1^0 = \frac{\partial}{\partial x} e^{-(x^2+y^2)} = -2xe^{-(x^2+y^2)}$$

$$G_1^{90} = \frac{\partial}{\partial y} e^{-(x^2+y^2)} = -2ye^{-(x^2+y^2)}$$

이 함수들을 그림 2. (a),(b)에 나타내었다.

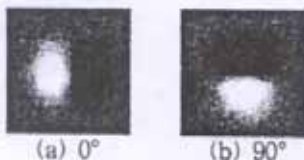


그림 2. Steerable basis filters

임의의 각 θ 에 대한 G_1 필터는 다음과 같이 G_1^0 , G_1^{90} 의 선형 조합으로 표현할 수 있다. :

$$G_1^\theta = \cos(\theta)G_1^0 + \sin(\theta)G_1^{90}$$

여기서 G_1^0 , G_1^{90} 는 G_1^θ 의 기저필터이고, $\cos(\theta)$ 와 $\sin(\theta)$ 는 기저필터들의 interpolation 함수이다. 본 연구에서는 $\theta=0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$ 를 위의 함수에 대입하여 6가지 방향 성분에 대한 steerable filter를 만들었다.

이 steerable filter를 4가지 크기 성분을 갖는 Laplacian pyramid[10]에 적용하면 4가지 크기, 6가지 방향 성분을 갖는 steerable pyramid를 얻을 수 있다.

3.2 Principal Component Analysis (PCA)

Steerable Pyramid에 의해서 생성된 24개의 영상들을 down-sampling하여 모두 24개의 4*4 픽셀 크기의 영상을 구성할 수 있는데 이를 이용하여 하나의 영상에 대하여 총 384 바이트 정보로 표현할 수 있다. 이러한 영상의 경우 global feature를 효과적으로 표현할 수 있지만 데이터 사이즈가 크고 redundancy가 여전히 존재하므로 이를 효과적으로 줄이기 위해서 PCA를 사용할 필요가 있다. 본 연구에서는 384바이트 정보를 80개의 PC, 즉 80바이트 정보로 줄여서 영상을 효과적으로 표현할 수 있도록 하였다.

Steerable pyramid로 구해진 24개의 영상 데이터에 대해 down-sampling한 후 PCA를 적용한 결과 이미지는 다음과 같다.



그림 3. Principal Component Analysis

3.3 Place Recognition

위치 인식 시스템의 목표는 시간 t 까지 얻은 global image feature를 이용하여 현재의 위치를 가능한 한 정확하게 예측하는 것이다. N_p 를 인식의 대상이 되는 장소의 개수라 하고 임의의 위치를 Q_t 라고 하면 global image feature를 이용하여 특정 장소 Q_t 를 판단하도록 한다.[11]

이를 위해서는 주어진 $v_{1:t}^G$ (시간 t 까지 계산된 global feature 값)에 대하여 현재의 위치 $Q_t = q$ 인 확률 $P(Q_t = q | v_{1:t}^G)$ 을 계산함으로써 얻을 수 있는데 다음과 같이 계산할 수 있다.

$$\begin{aligned} P(Q_t = q | v_{1:t}^G) &\propto p(v_t^G | Q_t = q) P(Q_t = q | v_{1:t-1}^G) \\ &= p(v_t^G | Q_t = q) \sum_{q'} A(q', q) P(Q_{t-1} = q' | v_{1:t-1}^G) \end{aligned}$$

여기서 $A(q', q) = P(Q_t = q' | Q_{t-1} = q)$ 이며 transition matrix로 표현된다.

정확한 장소 인식을 위해서는 모델 설정이 매우 중요하다. 각 장소에 대하여 비슷한 형태를 갖는 영상끼리 하나의 view를 구성하여 K 개의 view를 만들 수 있다. 각 view에 대한 Gaussian 모델을 설정하여 위치 인식 알고리즘에 적용한다.

4. 실험

4.1 장소 인식 수행

본 연구에서는 6개의 장소에 대한 인식을 수행하기 위해 각 장소에 대하여 200장의 영상을 추출, 총 1200장의 영상 DB를 구축하였다. 위치 인식 알고리즘의 첫 번째 항을 계산하기 위하여 먼저 PCA 벡터를 이용하여 각 장소의 view에 대한 모델을 구성하면 다음의 표1과 같다.

표1. 각 장소별 모델의 개수

장소	p1	p2	p3	p4	p5	p6	합계
모델	32	24	25	15	17	13	126

위 모델들에 대한 Gaussian matching 확률을 각각 계산하여 그 중 가장 큰 값을 첫째 항(observation likelihood)의 값으로 사용하였다.

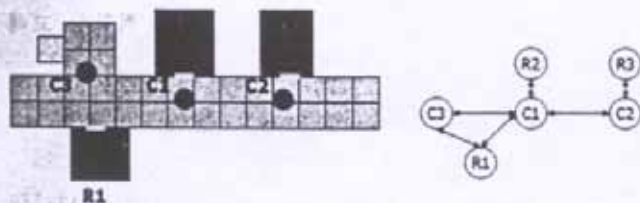


그림4. 6개의 장소에 대한 도면과 위상 맵

위 그림과 같은 위상 맵(topological map)을 사용하여 1차적으로 각 장소에 대한 observation likelihood를 구하고, 위치 인식 시스템의 성능을 테스트해 보았다.

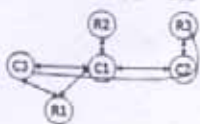
4.2 결과

각 장소별로 10장씩의 영상 data를 추출하여 위상 맵을 따라 장소를 이동할 시에 데이터를 보고 현재 위치가 어디인지를 판단하도록 하는 실험을 수행했다.

(1) C3→R1→C1 : 인식성공률 100 %



(2) C3→C1→C2→R3 : 인식성공률 98 %



결과를 보면 평균 99%로 높은 인식률을 보였다. 다른 경로에 대해서도 비슷한 결과를 얻을 수 있다.

5. 결론

본 논문에서 제안한 시각문맥정보 인식시스템은 Natural Marker를 사용하여 새로운 환경에 접했을 때 그 장소가 어디인지를 인식해 낼 수 있도록 하였다. 또한 위상 맵을 사용하여 여러 장소를 이동하면서 위치가 바뀔 때마다 실시간으로 장소를 인식할 수 있다.

현재 구현한 위치 인식 시스템은 영상의 global feature를 사용하여 top-down approach 방식으로 수행된다. 추후에 bottom-up approach 방식과 통합하여 object를 보고 현재 위치가 어디인지를 판단해 낼 수 있는 시스템을 연구해 볼 필요가 있다.

참고 문헌

- [1] Hao Yan, "Context-Aware Office Assistant."
- [2] Jason I. Hong, "A Context-based Information Assistant Human-Centered Computing Course, Fall 1999," UC Berkeley Computer Science Division.
- [3] 전홍준, 이상훈, 서일홍, "유비쿼터스 컴퓨팅 환경에서의 정보보조시스템의 설계와 구현," 대한전기학회 하계학술대회 논문집 2004.7.14-16.
- [4] Forsyth and Ponce, "Computer Vision A Modern Approach," Prentice Hall 2003.
- [5] Gonzalez, Rafael C "Digital Image Processing," Prentice Hall 2002.
- [6] Anestis Karasaris and Eero P. Simoncelli, "A Filter Design Technique for Steerable Pyramid Image Transforms," The 5th International Conference on Computer Vision, 1995, Boston, MA.
- [7] Eero P. Simoncelli and W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation,," 2nd IEEE International Conference on Image Processing, Washington, DC, vol III, pp 444-447, October, 1995.
- [8] W. T. Freeman and E. H. Adelson, "The Design and Use of Steerable Filters," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9 September 1991.
- [9] E. P. Simoncelli, H. Farid, "Steerable Wedge Filters," The 5th International Conference on Computer Vision, 1995, Boston, MA.
- [10] P. J. Burt, E. H. Adelson, "The Laplacian Pyramid and as a Compact Image Code," IEEE Transactions on Communications, Vol. Com-31, No. 4, April 1983.
- [11] A. Torralba, K. P. Murphy, W. T. Freeman, M. A. Rubin, MIT AI Lab. "Context-based vision system for place and object recognition."