

論文96-33B-2-9

다중 방향성 마스크를 이용한 영역 기반 스테레오 정합 알고리즘

(An Area-Based Stereo Matching Algorithm Using Multiple Directional Masks)

金 洛 賢

(Nak Hyun Kim)

요 약

스테레오 영상으로부터 변위치를 추출하기 위해 사용되는 기존의 영역 기반 정합방식에서는 일반적으로 정사각형 형태의 하나의 상관영역을 사용하여 좌-우 영상에서 상호 상관함수를 계산하고, 상관관계가 최대가 되는 점을 찾음으로써 변위값을 계산한다. 이러한 방식은 깊이 불연속선 (depth discontinuity) 주변에서 상대적으로 많은 정합 오차를 발생하게 되는데, 깊이 불연속선 주변에서는 변위값이 불연속적으로 변하므로 대응점 (corresponding point) 주변에서 좌-우 영상의 상관관계가 낮아지기 때문이다. 본 논문에서는 여러 개의 방향성 상관영역을 동시에 적용하여 여러 방향에서 변위 후보치들을 계산한 다음, 가장 높은 상관함수치를 가지는 점을 선택하여 변위값을 계산함으로써 깊이 불연속선 주위에서의 정합 오차를 줄이는 방법을 제안한다. 이 방식은 깊이 불연속선이 아닌 일반적인 영역에서도 정합 오차를 낮출 수 있음을 실험을 통해 확인하였고, 또한 변위값 외에도 변위 gradient 를 추정하는 수단을 제공한다. 본 논문에서는 다중 방향성 상관영역에 기반을 둔 정합 방식과 변위 gradient 추정 방법을 설명한 다음, 합성 영상과 자연 영상을 이용한 실험 결과를 제시한다.

Abstract

Existing area-based stereo matching algorithms utilize a single rectangular correlation area for computing cross-correlation between corresponding points in stereo images, and compute disparity by finding the peak in the correlation function. This technique results in relatively many matching errors in the vicinity of depth discontinuity, since, because of inconstant disparities around discontinuities, the cross-correlation becomes low in such area. In this paper, a new area-based matching strategy is proposed exploiting multiple directional correlation masks instead of a single one. The proposed technique computes multiple cross-covariance functions using each oriented mask. Peaks are detected from each covariance function and the disparity is computed by choosing the location with the highest covariance value. Proposed approach can also be applied to compute disparity gradients without obtaining dense depth data. A number of examples are presented using synthetic and natural stereo images.

*正會員, 韓國外國語大學校 制御計測工學科
(Dept. of Control & Instrumentation Eng.,
Hankuk University of Foreign Studies)

※ 이 논문은 1994년도 한국학술진흥재단의 공모과제
연구비에 의하여 연구되었음.

接受日字: 1995年11月7日, 수정완료일: 1996年1月30日

I. 서론

2차원 영상으로부터 거리(distance)나 깊이(depth)와 같은 3차원 정보를 복원하는 수단들 중 수동적인 방식으로는 가장 효율적인 스테레오 비전에서는 양 스테레오 영상에서 대응점(corresponding points)들을 찾아 변위값(disparity)을 추출하는 정합과정이 가장 핵심적인 부분이다. 스테레오 정합을 위해 다양한 정합 알고리즘이 제안된 바 있는데, 기존의 방법들은 크게 영역기반방식과 특징기반방식으로 분류할 수 있다.^{[11], [12]} 영역기반방식은 기준 영상에서 정합하려는 점을 포함하는 일정한 넓이의 기준 영역을 설정한 다음, 이 영역을 상대 영상의 탐색범위내에서 이동시키면서 상호연관도(cross correlation)를 구해 이 연관성이 최대가 되는 점을 찾음으로써 변위값을 구하는 방식이다. 이 방식은 영상에서의 명암의 변화가 뚜렷하고 3차원 형상의 변위 함수(disparity function)가 연속적으로 완만하게 변하는 경우 우수한 결과를 얻게 해 주고, 계산과정이 국소적으로 이루어질 수 있으므로 정합의 밀도를 임의로 조절할 수 있다는 장점들을 가진다. 반면 단점으로는 3차원 형상에서 깊이(depth)가 불연속적으로 변하거나 경사가 급격하게 변하는 경우, 또는 영상에서 반복적인 패턴이 나타나는 경우 오차가 많이 나타날 수 있고, 계산량이 비교적 많다는 점들을 들 수 있다.^[13] 특징기반 정합 방식은 상대적으로 나중에 개발된 방법으로서, 정합과정에서 영상의 밝기 대신 양 스테레오 영상에서 추출한 특징점들을 사용하여 정합 가능점(potential matches)들을 찾은 다음, 물리적 현상에 바탕을 둔 몇 가지의 제한조건들(constraints)을 적용하여 정합 가능점중에서 이 조건들을 가장 잘 만족시키는 점을 찾음으로써 양 영상을 정합하는 방식이다.^{[14]-[16]} 이 방식은 계산량 측면에서 대체로 효율적이며, 반복적인 패턴에 대해서도 적용될 수 있고, 3차원 형상에 의한 기하학적 왜곡(geometric distortion)에 상대적으로 강하다는 장점을 가진다. 그러나 이 방식은 스테레오 정합의 정확도가 특징점 추출 과정에 의해 제한된다는 한계를 가지는데, 양 스테레오 영상에서의 에지 위치가 각각 다르게 추출되거나 에지가 한 영상에서만 검출되는 경우와 같은 특징 추출상의 오차는 정합 오차로 나타나기 때문이다.

특징기반방식의 이와 같은 한계 때문에 항공사진분

석이나 로보틱스등 실제 응용분야에서는 영역기반방식이 선호되어왔고,^{[11], [17]-[19]} 다양한 성능 개선 방안이 제시되었다. 예를 들면, 영상에 포함되어있는 물체의 형상이 임의로 변하는 데서 발생하는 기하학적 왜곡에 의한 오차를 줄이기 위해 계층적 warping(hierarchical warping) 방식이 제시되었는데, 이 방식에서는 저해상도(low resolution)에서 추출한 변위값을 이용하여 다음 고해상도 영상을 재표본화(resampling)하여 정합을 수행함으로써 왜곡에 의한 오차를 줄인다.^{[18], [10]} 또한 기존의 상관함수를 사용하는 방식을 변형하여 일정한 초기 변위값으로부터 정확한 변위값을 반복적인 방식으로 계산하는 차등방식(differential approach)도 제안되었는데,^{[11], [11], [12]} 이 방식은 변위 초기값이 어느 정도 정확하지 않으면 반복과정의 수렴을 보장할 수 없는 한계를 가진다.

본 논문에서는 상관함수를 구하는 마스크 영역을 한 개만 사용하는 기존의 영역기반방식과 달리, 방향성을 가지는 다수의 마스크를 동시에 사용하여 스테레오 정합을 수행하는 방법을 제안한다. 기존의 영역기반방식은 스테레오 영상에서 밝기 함수가 대응점 주변에서 변위값 만큼 단순히 천이(translation)되었다는 점, 즉 정사각형의 상관영역내에서 변위값이 상수라는 가정에 기반을 두고 상관함수가 최대가 되는 점을 찾음으로써 변위값을 계산한다. 그러나 깊이 불연속성(depth discontinuity)에서는 변위값이 불연속적으로 변하기 때문에 위의 가정과는 달리 변위값이 상관영역내에서 상수가 아니며, 따라서 정합점에서의 상관함수값도 상대적으로 낮아지게 되어 정합오차가 발생할 가능성이 높아지게 된다. 논의를 단순화하기 위하여 정합하려는 점의 좌우 부분에서 변위값이 서로 다른 수직방향의 계단형 깊이 불연속선을 고려한다면, 상관영역을 정사각형으로 하는 대신 정합점을 중심으로 좌우 부분으로 나눈다면 각각의 영역에서 보다 높은 상관함수값을 얻을 수 있는데, 왜냐 하면 각각의 좌우 정합영역에서 변위값은 상수가 되므로 영역기반방식의 기본 가정을 만족시키기 때문이다. 따라서 이와 같이 분할된 정합 영역을 사용함으로써 깊이 불연속선 주변에서의 정합 오차를 줄일 수 있다.

본 논문에서 제안하는 다중 방향성 마스크 방식은 위와 같은 계단 형태의 깊이 모형(depth model)에 기반을 둔 방식으로서, 깊이 불연속선 주변의 대응점에

서 좌우 영상의 상관관계를 높이기 위해 좌우로 분할된 상관영역을 사용하는 것이 본 방식의 기본 개념이다. 스테레오 영상에서 깊이 불연속선은 임의의 방향(orientation)을 가질 수 있기 때문에 기본 개념을 확장하여, 분할된 영역을 일정한 각도 간격으로 회전시켜 생성된 다수의 마스크를 사용하여, 각각의 영역에서 상관함수를 계산하고 상관함수의 최대점을 찾아 정합 후보점들을 선정한 다음, 가장 높은 상관관계를 가지는 점을 선택함으로써 스테레오 영상을 정합한다. 이러한 정합방식은 깊이 불연속선 주변이 아닌 일반적인 영역에서도 변위값을 추출하는데 사용될 수 있다. 다중 영역 방식은 또한 변위 gradient를 추정하는 수단도 제공한다. 각 방향의 좌우 영역에서 추정된 변위값의 차이는 그 방향에서 변위치 편미분(directional derivative)의 근사치에 해당되므로, 여러 방향에서의 변위치의 차이를 이용하여 변위 gradients를 추정할 수 있다는 것이다. 본 논문에서는 다중영역방식을 사용한 변위 추출 알고리즘과 변위 gradient 추정 알고리즘을 제안하고, 이 방식들을 합성 영상과 자연 영상에 적용하여 얻은 실험 결과를 설명하고자 한다.

본 논문의 구성은 다음과 같다. 우선 II 장에서는 기존의 영역기반 상관함수 정합방식을 설명한 다음, 본 논문에서 제안하는 다중 방향성 상관 영역의 개념과 그 타당성을 설명한다. III 장에서 이 개념에 기반을 둔 스테레오 정합 알고리즘과 변위 gradient 추정 방식을 제안한다. 제안된 알고리즘의 타당성과 유용성을 검증하기 위해 다수의 합성 영상과 자연 영상을 사용한 실험이 수행되었는데, 그 실험 결과를 IV 장에서 소개한다.

II. 영역 기반 방식과 방향성 정합 영역

1. 기존의 영역기반 상관함수 방식

이 절에서는 기존의 영역기반방식에 의한 스테레오 정합을 살펴보기로 한다. 좌우 스테레오 명암도 영상(intensity image)을 각각 $f_l(x, y)$, $f_r(x, y)$ 라 하면 두 영상의 관계는

$$f_l(x, y) = f_r(x + h(x, y), y) + n(x, y) \quad (1)$$

와 같이 표현할 수 있는데, 여기서 $h(x, y)$ 는 (x, y) 에서의 변위값이고, $n(x, y)$ 는 영상에 섞인 잡음 성분을 나타낸다. 변위치 $h(x, y)$ 를 구하기 위하여 영역기반

상관함수 방식에서는 (x, y) 를 중심으로 하는 정사각형 영역 Ω 를 설정한 다음,

$$C(\xi) = \sum_{(a, \beta) \in \Omega} f_l(x + a, y + \beta) f_r(x + \xi + a, y + \beta) \quad (2)$$

를 계산하여 $C(\xi)$ 가 최대가 되는 ξ 를 찾음으로써 변위치를 구한다. 이 방식은 상관 영역내에서는 영상신호가 변형없이 천이되었다는 것, 즉,

$$f_l(x + a, y + \beta) = f_r(x + h_0 + a, y + \beta), (a, \beta) \in \Omega \quad (3)$$

이라는 가정에 기반을 둔 것이다. (1) 식과 (3) 식을 비교해보면 상관관계 방식은 $h(x, y) = h_0$, 즉, 변위 함수가 국소적으로 상수일 때 이 방식은 정확한 변위를 추정할 수 있게 한다는 것을 알 수 있다. 그러나 실제 영상에서는 변위 함수 $h(x, y)$ 가 임의의 형태를 가질 수 있으므로 변위함수(disparity function)의 변동이 심한 부분에서는 정합오차가 크게 나타날 수 있는데, 이러한 현상을 기하학적 왜곡이라 한다.^[3] 기하학적 왜곡은 warping에 의해서 줄일 수 있는데, 이것은 변위치의 초기 추정값이 있을 때 (1) 식에 의해 두 영상 중 하나를 부분적으로 재표본화(resampling)한 다음 상관관계를 구함으로써 변위 추정시의 정확도를 높이는 방법이다.^{[8], [9]} 다해상도(multi-resolution) 방식을 사용한다면 저해상도에서 구한 변위치를 다음 해상도에서의 변위 초기치로 사용하여 warping을 수행할 수 있다.

영역기반방식을 보다 실제적인 구현상의 관점에서 본다면, 양 영상의 밝기와 contrast 차이를 보상하기 위해 (2)의 correlation 관계식 대신 다음의 covariance 식을 더 많이 사용한다.

$$\chi(\xi) = \sum_{(a, \beta) \in \Omega} \frac{[f_l(x + a, y + \beta) - \bar{f}_l][f_r(x + \xi + a, y + \beta) - \bar{f}_r]}{\sigma_l \sigma_r} \quad (4)$$

여기서 \bar{f} 와 σ_r 는 각각 해당 마스크 영역내에서의 영상 밝기의 평균치와 표준편차를 나타낸다. 영역기반 방식의 최종 단계는 변위값의 범위내에서 (2) 또는 (4) 식을 사용하여 구한 상관함수 또는 covariance 함수에서 함수값이 최대가 되는 점의 위치를 구함으로써 변위값을 결정하는 것이다. 최대점의 위치는 화소 단위, 또는 화소이하(subpixel) 단위의 정확도로 구할 수 있는데, 화소이하의 정확도를 얻으려면 상관함수 값을 최대점 주변에서 보간법(interpolation)을 적

용하여 특정한 함수형태로 근사시킨 다음, 근사함수의 최대점을 찾는 방식으로 수행된다. 가장 보편적으로 사용되는 방법은 최대점 부근의 함수를 2차 다항식 (polynomial) 으로 근사시키는 방식인데, 이 근사방식을 간략하게 살펴보기로 한다. 최대점을 h_0 라 하고 h_0-1 , h_0 , h_0+1 등 세 점에서의 covariance 값을 각각 $\gamma_-, \gamma_0, \gamma_+$ 라 하자. h_0 를 원점으로 하고 세 점을 지나는 2차 보간 다항식을 $p(x) = ax^2 + bx + c$ 라 한다면 $p(x)$ 는

$$\begin{aligned} p(-1) &= a - b + c = \gamma_- \\ p(0) &= c = \gamma_0 \\ p(1) &= a + b + c = \gamma_+ \end{aligned} \quad (5)$$

를 만족시켜야 한다. 따라서 (5) 식을 풀어 계수 a, b, c 를 구한 다음 $p(x)$ 의 최대점 $p_{max} = -\frac{b}{2a}$ 를 구하면 보다 정확한 최대점, 즉, 변위값을 $h_0 = h_0 + p_{max}$ 로 계산할 수 있다.

2. 계단형 깊이 모형과 방향성 마스크 방식

이 절에서는 기존의 영역기반방식에서 가정하고 있는 상수 변위 모형 (constant disparity model) 대신 계단형 깊이 모형 (step depth model) 을 사용한 방향성 상관영역 방식의 개념을 설명한다. 이 개념에 기초를 둔 본격적인 정합 알고리즘은 III 장에서 설명하기로 한다. 일반적으로 스테레오 영상에서는 물체의 차폐 (occlusion) 등으로 인해 물체의 경계부분에서 깊이 불연속선 (depth discontinuity) 이 존재하게 되는데, 다음과 같은 수직 방향의 계단 불연속선 (step discontinuity) 을 가지는 변위 함수를 고려하자.

$$h(x, y) = \begin{cases} h_1, & x \leq x_0 \\ h_2, & x > x_0 \end{cases} \quad (6)$$

이러한 변위 구조에서는 (2) 식이나 (4) 식을 사용하여 (x_0, y) 에서 변위값을 구하면 오차가 생길 수 있는데, 상관 영역내에서 양 스테레오 신호의 절반만이 일치할 뿐이고, 나머지 절반은 서로 일치하지 않기 때문이다. 따라서 기존의 상관함수 방법은 깊이가 불연속적으로 변하는 영역에서는 오차가 크게 나타난다.

이제 상관 영역을 그림 1과 같이 Ω_1, Ω_2 의 두 부분으로 나누고 각각의 영역들에 대해 cross-cova-

riance 을 각각 적용하면 다음과 같은 두 개의 covariance 함수를 구할 수 있다.

$$\gamma_1(\xi) = \sum_{(\alpha, \beta) \in \Omega_1} \frac{[f_r(x_0 + \alpha, y + \beta) - \bar{f}_r][f_l(x_0 + \xi + \alpha, y + \beta) - \bar{f}_l]}{\sigma_r \sigma_l} \quad (7)$$

$$\gamma_2(\xi) = \sum_{(\alpha, \beta) \in \Omega_2} \frac{[f_r(x_0 + \alpha, y + \beta) - \bar{f}_r][f_l(x_0 + \xi + \alpha, y + \beta) - \bar{f}_l]}{\sigma_r \sigma_l} \quad (8)$$

γ_1 과 γ_2 는 상관영역이 서로 다르다는 점외에는 차이가 없는데, 잡음이 없는 경우 covariance 함수의 최대점의 위치는 각각 h_1 과 h_2 으로 다르게 나타나게 된다. 따라서 기존의 상관영역의 경우와는 달리 상관영역을 분할함으로써 계단 불연속선의 좌우에서 두 개의 변위값을 추출할 수 있다. 이 것이 본 논문에서 제안하는 방향성 상관영역을 사용한 정합 방식의 기본적인 개념으로서, (2) 식의 경우와는 달리 (7), (8) 식에서는 상관 영역 내에서 두 스테레오 영상신호가 완전히 일치하는 경우가 각각 존재하게 되므로 보다 정확하게 변위값을 추출할 수 있게 된다.

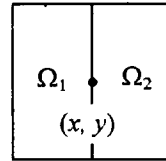


그림 1. 상관영역의 분할
Fig. 1. Decomposition of correlation area.

변위값을 추정하는 본격적인 정합 알고리즘은 III 장에서 설명하기로 하고 여기서는 (6) 식의 계단형 깊이 모형에서의 (7), (8) 식의 의미와 타당성을 검토하기로 한다. 수식표현의 편의를 위해 보다 실용적인 (4) 식 대신 (2) 식을 사용하여 설명하자면, 우선 대응하려는 점 (x, y) 을 원점으로 잡고 (1) 식을 (2) 식에 대입하면

$$\begin{aligned} C(\xi) &= \sum_{\alpha, \beta} f_r(\alpha, \beta) f_l(\xi + \alpha, \beta) \\ &= \sum_{\alpha, \beta} [f_r(h(0, 0) + \alpha, \beta) + n(\alpha, \beta)] f_l(\xi + \alpha, \beta) \\ &= \sum_{\alpha, \beta} f_r(h(0, 0) + \alpha, \beta) f_l(\xi + \alpha, \beta) + \sum_{\alpha, \beta} f_l(h(0, 0) + \alpha, \beta) \\ &\quad + \sum_{\alpha, \beta} n(\alpha, \beta) f_l(\xi + \alpha, \beta) \\ &= C_1(\xi) + C_2(\xi) + N \end{aligned}$$

이 된다. 깊이 불연속선, 즉, (6) 식의 와 같은 점

에서 기존의 correlation 방식을 적용하는 경우 C_1, C_2 중 하나는 두 스테레오 영상신호가 서로 un-correlated 된 상태이므로 correlation 을 취하는 경우 신호성분이 아닌 잡음성분으로 나타나게 된다. 즉, $C_1(\xi)$ 과 $C_2(\xi)$ 의 최대치는 각각 \hat{h}_1 과 \hat{h}_2 에서 나타나므로 예를 들면 $\xi = \hat{h}_1$ 에서 $C(\hat{h}_1) = C_1(\hat{h}_1) + C_2(\hat{h}_1) + N$ 이 되는데, 여기서 $C_2(\hat{h}_1) + N$ 은 잡음 성분을 형성하게 된다. 따라서 깊이불연속선에서 (4) 식의 정규화된 covariance 함수를 사용하는 경우 covariance 값의 이론적 최대치는 잡음이 존재하지 않더라도 0.5 이하에서 머물게 된다. 반면 영역을 Ω_1 과 Ω_2 로 나누어 두 개의 마스크를 각각 적용하는 경우 마스크에 따라 C_1, C_2 중 한 성분만 사용하게 되므로 다른 성분은 0 이 되어, 같은 면적의 영역을 사용하는 경우, 보다 높은 correlation 값을 기대할 수 있고, 정규화된 covariance 함수의 경우 잡음이 없으면 이론적으로는 1의 최대치를 가질 수 있다. 따라서 스테레오 정합시 정합오차를 줄일 수 있는 토대를 제공한다.

영역분할 방식으로 또한 깊이 불연속선을 검출하는 것도 가능한데, (7) 과 (8) 식을 사용하여 얻은 두 개의 변위값 \hat{h}_1, \hat{h}_2 의 차이를 이용하면 깊이 불연속선을 검출할 수 있다. 즉, 두 변위값의 차이가 일정한 임계값 보다 큰지의 여부를 판단함으로써 불연속선의 존재를 검출할 수 있고, 또한 불연속선의 높이도 추정할 수 있다. 이 방법을 확장하면 변위 gradient를 추정할 수 있는데, 그 방법은 다음 장에서 살펴보기로 한다.

III. 스테레오 정합 알고리즘

1. 정합 알고리즘

II 장에서는 논의의 편의를 위해 수직방향의 불연속선에 대해 고려하였으나, 스테레오 영상에 존재하는 깊이 불연속선은 영상이 2차원 함수이므로 임의의 방향을 가질 수 있다. 이러한 방향은 변위값을 추출하기 전에는 알 수 없으므로 (7), (8) 식에서 제시된 방법은 수직방향의 불연속선이 아닌 일반적인 경우에 적용할 수 없다. 따라서 본 연구에서는 변위값을 구하지 않은 상태에서도 임의의 방향을 가지는 깊이 구조를 처리할 수 있도록 여러 방향에서의 다수의 마스크를 동시에 사용하는 방법을 사용하였다. 본 논문의 방법에서는 상관영역 마스크를 그림 2에 보인 바와 같이 일정한 각

도 간격으로 생성한다. 마스크를 생성하는 각도의 간격은 $30^\circ \sim 45^\circ$ 정도가 적절한 것으로 실험적으로 확인하였다. 이렇게 생성된 상관 마스크 영역을 $\Omega_1, \Omega_2, \dots, \Omega_{2m-1}, \Omega_{2m}$ 이라 하자. 여기서 m 은 180° 를 회전 각도의 간격으로 나눈 값인데, 예를 들면 간격이 30° 인 경우 전체 마스크의 갯수는 12개이다. 약간 상황은 다르지만, 본 논문에서 사용한 마스크와 유사하게 방향성을 가지는 필터들은 에지 검출시 많이 사용된 바 있으며, 이러한 방향성 필터들은 신호 대 잡음비 (SNR) 나 에지 검출 특성에서 등방향성 (isotropic) 필터들에 비해 우수한 것으로 알려져 있다. [13], [14]

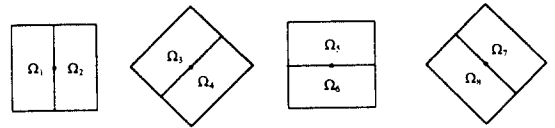


그림 2. 45° 간격으로의 다중 방향성 마스크의 생성
Fig. 2. Generation of directional masks at 45° interval.

정합의 다음 단계는 각 영역에서 다음의 정규화된 covariance 함수 $\gamma_i, i=1,2,\dots,2m$ 를 구하고, 각 함수의 최대점 위치를 찾는 것이다.

$$\gamma_i(\xi) = \sum_{(w, \beta) \in \Omega} \frac{[f_i(x+a, y+\beta) - \bar{f}_i] [f_i(x+\xi+a, y+\beta) - \bar{f}_i]}{\sigma_i \sigma_i} \quad (9)$$

이렇게 구한 위치들을 $\hat{h}_1, \hat{h}_2, \dots, \hat{h}_{2m}$ 이라 하면, 이 값들은 변위 후보치 (potential disparities) 에 해당된다.

변위 후보치에서 실제 변위값을 구하는 방법에는 여러 가지가 있을 수 있으나 본 연구에서는 가장 단순하면서 기존의 영역기반방식과 개념적으로 동일한 방법, 즉, 변위 후보치중 가장 높은 covariance 값을 가지는 점을 선택하는 방식을 사용하였다. 이 방식의 장점은 i) 간단해서 구현하기 용이하고, ii) 추가적인 계산량이 매우 적고, iii) 방법 자체가 국소적이어서 요구되는 정합의 밀도와 관계없이 일정한 정확도를 가질 수 있다는 점등을 들 수 있다.

이제까지 설명한 정합단계들을 전체적으로 요약하면 다음과 같다.

- i) 사각형 형태의 상관마스크를 그림 2에서 보인 바

와 같이 일정한 각도 간격으로 생성한다.

- ii) 각각의 상관 마스크를 사용하여 다음과 같은 정규화된 covariance를 계산한다.

$$\gamma(\xi) = \sum_{(a, \beta) \in \Omega} \frac{[f_i(a, \beta) - \bar{f}_i][f_i(a + \xi, \beta) - \bar{f}_i]}{\sigma_i \sigma_i}$$

여기서 \bar{f}_i , σ_i 은 각각 영상신호의 평균치와 표준편차를 의미한다.

- iii) 각각의 covariance 로 부터 최대점을 찾아 각 마스크에서의 정합 후보치 \hat{h}_i , $i=1, \dots, 2m$ 을 구한다. 최대점의 위치는 II 장에서 설명한 방식을 이용하여 화소이하 단위의 정확도로 계산한다.

- iv) 정합 후보치중 가장 높은 covariance 값을 가지는 값을 변위값으로 정한다. 추출된 변위값을 \hat{h} 이라 하면 변위값의 신뢰성을 높이기 위해

$$|\hat{h} - \hat{h}_i| \leq 1$$

을 만족시키는 \hat{h}_i 의 갯수가 일정한 값 (보통 $2m \times 0.5$ 보다 약간 작은값) 보다 작으면 \hat{h} 값을 버린다.

위 과정 iv) 에서 정합 후보치들의 일관성을 시험하여 정합의 신뢰성을 높이는 방법을 사용했는데, 이 방법은 추정된 변위값이 다른 정합 후보치들과 동떨어진 경우에는 그 변위값이 오차일 가능성이 많다는 관찰에 기반을 둔 것이다.

2. 변위 gradient 추정

2차원 변위함수를 $h(x, y)$ 라 하면 변위 gradient 는 $\nabla h = (\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y})$ 로 정의되는데 일반적으로 변위 gradients 나 surface parameter 들은 추출된 변위 값을 이용하여 계산할 수 있다. 그러나 변위 gradient 를 계산하려면 매우 밀도가 높은 변위값 (dense disparity) 을 추출하여야 하므로, 저밀도의 변위 (sparse disparity) 데이터만 제공하는 정합방식으로는 표면 복원 (surface reconstruction) 절차를 거치지 않고서는 변위 gradient 를 계산할 수 없다. 본 절에서는 변위값을 추출하지 않고서도 다중 방향성 영역 방식을 사용하여 변위 gradient 를 추정하는 기법을 제시하고자 한다. 약간 방법이 다르지만 확장 correlation 방식 (extended correlation method) 을 사

용하여 변위 gradient 와 surface parameter 들을 구하는 방법은 Devenay 와 Faugeras^[15] 에 의해 제시된 바 있다.

본 제안의 기본 개념은 180° 의 위상 차이를 가지는 한 쌍의 방향성 마스크를 통해 검출된 변위값들은 그 방향에서의 변위값의 discontinuity의 크기, 즉, 그 방향에서의 변위 함수의 편미분의 근사치에 해당된다는 관측에 기반을 둔 것이다. 그림 3에서 보인 바와 같이 i 번째 마스크쌍이 수평축과 이루는 각도를 θ_i 라 하고 마스크와 직각방향의 단위 벡터를 m_i 라 하자. 변위 함수 $h(x, y)$ 의 m_i 방향에서의 편미분은 directional derivative의 정의에 의해

$$\frac{\partial h}{\partial m_i} = \nabla h \cdot m_i = \frac{\partial h}{\partial x} \cos \theta_i + \frac{\partial h}{\partial y} \sin \theta_i \quad (10)$$

로 주어진다. 한 쌍의 마스크를 사용하여 얻어진 변위 추정치를 \hat{h}_{2i-1} 와 \hat{h}_{2i} 라 하고 그 차이를 $\hat{d}_i = \hat{h}_{2i-1} - \hat{h}_{2i}$ 라고 정의하면 \hat{d}_i 가 m_i 방향에서의 $h(x, y)$ 의 편미분과 유사하다는 가정에서

$$\hat{d}_i \approx \frac{\partial h}{\partial m_i} = h_x \cos \theta_i + h_y \sin \theta_i$$

라는 근사표현을 얻을 수 있는데, 여기서 $h_x = \partial h / \partial x$, $h_y = \partial h / \partial y$ 이다. 방향성 마스크 쌍의 갯수는 m 개이므로 총 m 개의 차이값 \hat{d}_i , $i=1, \dots, m$ 를 얻을 수 있다.

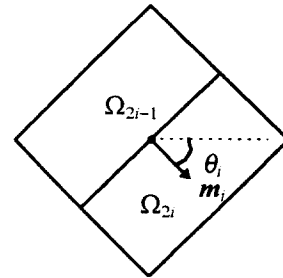


그림 3. i 번째 마스크쌍에서의 방향과 단위 벡터
Fig. 3. Orientation and the unit vector of i -th mask pair.

(10) 식을 이용하면 이론적으로는 m 이 2 이상이면 \hat{d}_i , $i=1, \dots, m$ 로 부터 h_x 와 h_y 를 결정할 수 있으나 추정의 신뢰성을 높이기 위해 보다 많은 \hat{d}_i 를 사용하면 일반적인 least squared error 방법을 사용하여

gradient를 추정할 수 있다. 즉, m 개의 데이터를 사용하는 경우 전체 오차는

$$E = \sum_{i=1}^m [d_i - h_x \cos \theta_i - h_y \sin \theta_i]^2$$

와 같이 되므로, 오차를 최소로 하는 h_x 와 h_y 는

$$\frac{\partial E}{\partial h_x} = 0, \quad \frac{\partial E}{\partial h_y} = 0$$

의 두 식을 만족시키는 근을 구함으로써 얻을 수 있다.

IV. 실험 결과

1. 합성 영상을 사용한 실험

그림 4에 보인 영상은 가우시안 잡음이 섞인 wedding cake 형태의 random-dot stereogram 으로서 크기는 128x128 이며 변위의 범위는 0 ~ 12 화소이다. 이 영상에 기존의 영역기반 방식과 방향성 방식을 적용한 정합 결과가 그림 5에 나와있는데, 이 그림으로부터 방향성 마스크 방식의 결과가 상대적으로 오차가 적음을 확인할 수 있다.

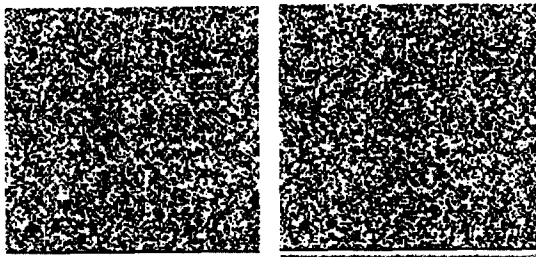
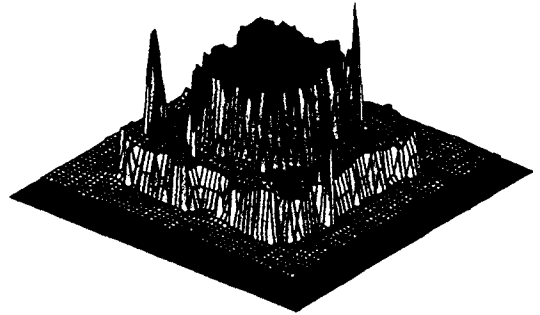


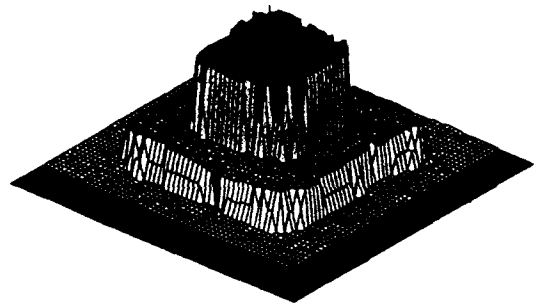
그림 4. Wedding cake 스테레오그램
Fig. 4. Wedding cake stereogram.

정합 결과를 정량적으로 분석하기 위하여 각 방식에 의해 추출한 변위값을 실제 변위값과 비교하여 한 화소당의 평균정합오차를 계산하였는데, 그 결과가 표 1에 요약되어 있다. 표 1에서는 영상의 전체 영역에서의 평균 정합 오차와 불연속선 주위 (2 화소) 에서의 평균정합오차를 보이고 있는데, 각 방식은 불연속선 주위에서 상대적으로 높은 정합오차를 발생시키는 것을 알 수 있다. 일반적으로 영역기반 방식은 넓은 마스크를 사용할수록 평균적인 정합오차가 줄어드는 성향을 보이는데, 표 1에서 6x11 크기의 방향성 영역을 사용하는 경우 마스크의 크기가 정사각형 영역 방식에 비

해 절반 정도임에도 정사각형 방식에 비해 전체 영역에서는 35%, 불연속선 주변에서는 43% 정도의 오차 감소율을 보였으며, 11x11 영역과 마스크 면적이 거의 같은 7x17의 방향성 마스크를 사용하는 경우에는 오차가 더욱 많이 줄었다.



(a)



(b)

그림 5. 그림 4의 영상에 대한 정합 결과 (a) 기존의 영역 기반 방식 (11x11) (b) 방향성 마스크 방식 (6x11)

Fig. 5. Matching results for the images in Fig. 4. (a) Using a single 11x11 mask (b) Using 8 directional 6x11 masks .

표 1. 한 화소당 평균 정합 오차 비교 (단위: pixel)

Table 1. Comparison of the mean matching error per pixel.

	정사각형 영역 (11x11)	방향성 영역 (6x11)	방향성 영역 (7x15)
전체 영역 평균	0.468	0.303	0.261
불연속선 주변에서의 평균	2.051	1.167	1.026

그림 6에 보인 영상은 프로그램에 의해 합성한 250x250 크기의 스테레오그램으로서 반구 (hemisphere)를 위에서 바라본 형태의 변위 구조를 가지고 있고, 변위 범위는 0 ~ 24 화소이다. 이 스테레오 영상에 대해 변위 gradient 추정 알고리즘을 적용하였는데, 이 경우 회전각의 간격은 30° 이며, 각 마스크의 크기는 15x7이다. 이 영상에 대한 변위 gradient 값을 그림 7에 보였는데, (a)는 영상에서의 실제 gradient를 나타냈고, (b)에서는 gradient 추정 알고리즘에 의해 계산된 gradient를 표시하였다. 그림에서는 각 점에서의 h_x 와 h_y 를 벡터 형태로 나타냈는데, 선의 방향과 길이로 gradient의 orientation과 크기를 표현한 것이다. 그림 7(a)와 (b)를 비교해보면 추정된 gradient는 변위치의 변화가 매우 심한 경계 부분의 일부를 제외하고는 대체로 정확하다는 것을 알 수 있다.

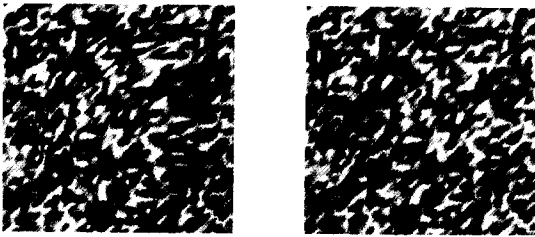


그림 6. 반구 변위 구조를 가진 합성 스테레오 영상
Fig. 6. Synthetic stereo images with hemisphere disparity structure.



(a) (b)

그림 7. 그림 6의 영상에서의 변위 gradient.
(a) 실제 gradient (b) 추정 알고리즘의 결과
Fig. 7. Disparity gradients of images in Fig. 6.
(a) Actual gradient. (b) Estimated gradient.

2. 자연 영상을 사용한 실험

그림 8은 항공 영상으로서 펜타곤을 보인 것이다.

영상의 크기는 256x256이며, 변위의 범위는 -4 ~ 4 화소이다. 그림 9에서는 기존의 영역기반 방식과 방향성 다중 영역방식을 사용한 결과를 비교하였다. 이 결과들은 모든 화소에서 정합을 수행하여 얻어진 것인데, 그림 9의 3차원 그래프를 작성하기 위해 depth map을 단순한 local averaging을 사용하여 85x85의 크기로 줄인 것이다. 단일 영역 방식에 비해 방향성 영역의 결과가 정합오차가 적음을 확인할 수 있다.

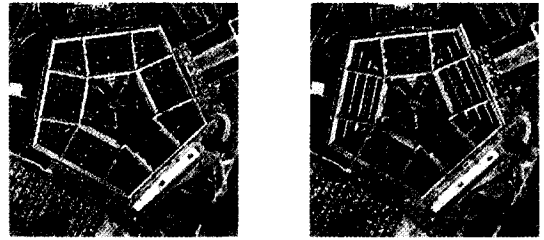


그림 8. '펜타곤' 스테레오 영상
Fig. 8. 'Pentagon' stereo images.

영역 기반 방식은 global 한 제어 구조를 사용하지 않고 정합이 완전히 국소적으로, 그리고 각 정합 대상 점에 대해 개별적으로 이루어지므로 정합의 정확도는 정합 대상점의 밀도에 영향을 받지 않는다. 따라서 응용분야에 따라서는 정합을 전체 화소에 대해 수행하지 않고 interest operator를 사용하여 추출한 일부 대상점들에 대해 적용하는 것도 가능한데, 이러한 가능성을 보여주는 결과가 그림 10에 나와있다. 이 결과는 다음과 같은 방법으로 얻어졌다. 우선 좌측 영상에 Marr-Hildreth edge detector^[16]를 적용하여 에지를 추출한 다음, 이들 에지점들을 정합 대상점으로 하여 방향성 영역 방식을 사용하여 대응점들을 구했다. 이것은 특징기반방식과는 차이가 있는데, 특징기반방식에서는 양 영상에서 특징점들을 동시에 추출한 다음 특징점끼리만 정합을 수행하게 된다. 즉, 정합가능점 (potential matches) 들이 특징점이 있는 위치로만 한정이 되는데, 그림 10의 정합과정에서는 edge detection은 단지 좌측 영상에서 정합 대상점들의 숫자를 제한하는 용도로만 사용되었고, 우측 영상은 좌측 에지 영상과 정합 결과로 얻어진 변위 데이터로부터 합성된 것이다. 이 영상들을 cross-eye view 방식으로 관찰하면 스테레오 정합의 결과를 확인할 수 있는데, 차폐 현상이 나타나는 펜타곤의 지붕 부분외에는 정합이 대체로 정확하게 이루어졌음을 볼 수 있다.

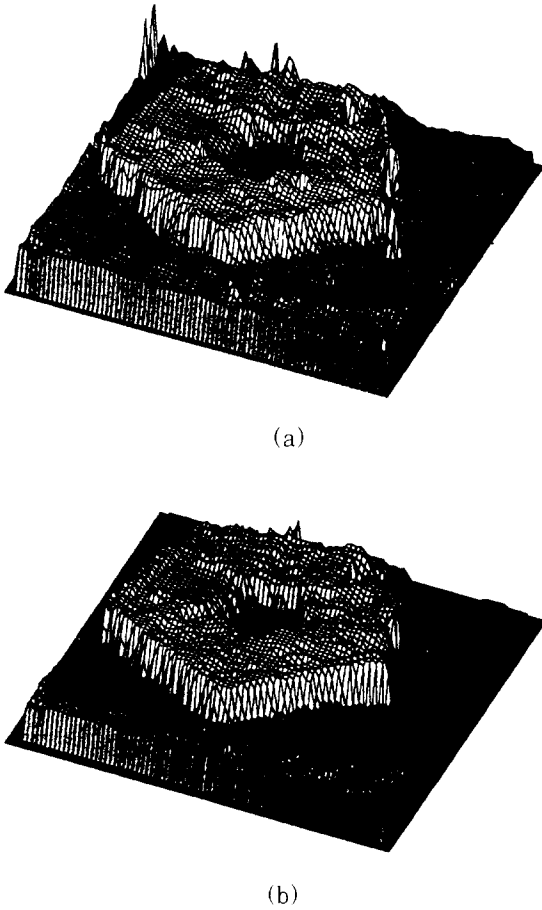


그림 9. 펜타곤 영상에 대한 정합 결과. (a) 기존의 영역기반방식 (11x11) (b) 방향성 마스크 방식 (7x17)

Fig. 9. Matching results for 'Pentagon'. (a) Using a single 11x11 mask. (b) Using 8 directional 7x17 masks.

펜타곤 영상에 대해 변위 gradient 추정 알고리즘을 적용한 실험도 수행하였는데, gradient 추정 알고리즘도 좌측 영상의 에지점들에 대해서만 적용하였다. 펜타곤 영상은 그림 6의 합성영상과 비교하면 매우 복잡한 변위 구조를 가지고 있으므로 gradient 추정 결과를 그림 7과 같은 needle diagram으로 표시하기는 어려워서 gradient의 방향은 생략하고 크기만을 보인 것이 그림 11 인데, (a) 에서는 magnitude가 영상의 밝기에 비례하도록 나타난 것이다. 또한 (b) - (d) 는 magnitude를 여러 단계에서 thresholding 한 이진 영상을 보인 것인데, 여기서 각 영상은 에지점들중 gradient magnitude 가 일정한 값 이상인 점들만 표

시한 것이다. 그림 11을 통해 gradient 추정 알고리즘이 펜타곤 영상의 건물 부분을 명확히 분리해냈고, 우측의 다리 부분에서도 상대적으로 변위 gradient의 크기가 크게 나타난 것을 볼 수 있는데, 이는 실제 지형과 매우 유사함을 알 수 있다.

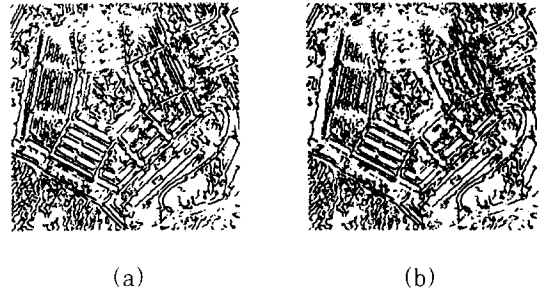


그림 10. 에지점들에 대해서만 정합을 수행한 경우의 결과 (a) 좌측 에지 영상 (b) 좌측 에지 영상과 변위 데이터로부터 합성한 우측 에지 영상

Fig. 10. Matching results obtained by applying the algorithm to the edge points in the left image (a) Edge points in the left image (b) Right edge images composed from the left edge images and the disparity data.

V. 결 론

본 논문에서는 기존의 영역기반방식을 확장하여 다수의 방향성 마스크를 동시에 사용하여 정합오차를 줄이는 정합방식과 변위 gradient 를 추정하는 방법을 제안하였다. 방향성 마스크 방식을 깊이 불연속선 부근에서의 정합 오차를 줄이는 것을 일차적인 목표로 하여 탐구되었으나, 합성 영상과 자연 영상을 이용한 실험을 통해 방향성 영역 방식은 기존의 단일 영역 방식에 비해 깊이 불연속선 부근외에 일반적인 영역에서도 보다 개선된 성능을 보이고 있음을 확인하였다. 본 논문에서 제안한 정합 방식과 변위 gradient 추정 방식은 기존의 영역기반방식과 마찬가지로 계산과정이 순수하게 국소적으로 이루어지므로 변위 또는 변위 gradient 의 측정 밀도를 임의로 조절할 수 있다는 장점을 가진다.

따라서 지도 제작등에서와 같이 높은 밀도의 정합 결과가 요구되는 분야나 obstacle avoidance 분야와

같이 저밀도의 공간 점유 정보만으로도 충분한 분야 등 다양한 응용 분야에서 적용될 수 있을 것으로 기대한다.

템을 구현하는 경우에는 계층적 제어 구조를 사용함으로써 계산량을 줄이거나 변위 측정의 정밀도를 조절할 수 있을 것이다.

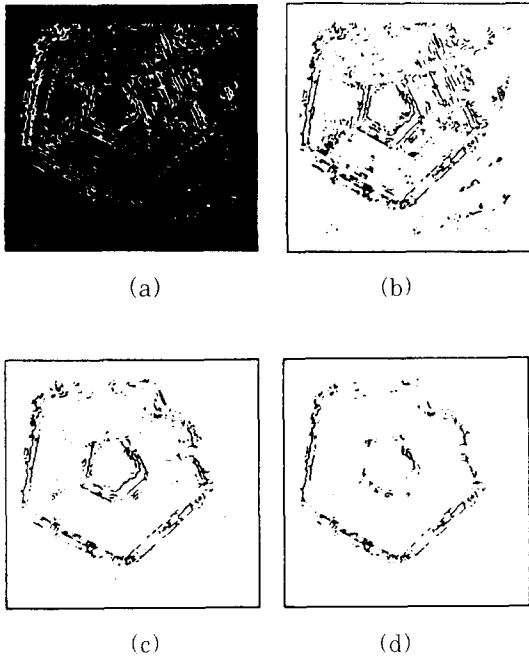


그림 11. 좌측 영상의 각 에지점에서의 gradient 추정 결과 (a) Gradient 크기를 밝기에 비례하도록 나타낸 영상 (b) - (d) Gradient 크기 영상을 0.5, 1, 1.5 화소 단위로 thresholding 한 이진 영상

Fig. 11. Gradient magnitude estimated at each edge point in the left image. (a) Gradient magnitude coded by image intensity. (b) - (d) Binary images obtained by thresholding based on gradient magnitude. Thresholds are (b) 0.5, (c) 1, (d) 1.5 pixels.

방향성 영역 방식은 기존의 방식에 비해 보다 많은 갯수의 마스크 영역을 동시에 사용하므로 계산량 측면에서는 기존의 방식 보다 불리하다고 할 수 있다. 그러나 컴퓨터 시스템의 성능이 지속적으로 개선되고 있으므로 추가 계산량에 대한 부담은 점점 더 낮아질 것으로 기대된다.

본 논문에서 제안된 정합방식은 다해상도 (multi-resolution) 방식과 같은 계층적 제어 구조 (hierarchical control structure) 로 확장하는 것도 가능하다. 따라서 특정 응용 분야에서 사용되는 실제 시스

참 고 문 헌

- [1] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. II, Addison-Wesley, 1993.
- [2] U. R. Dhond and J. K. Aggarwal, "Structure from stereo - A review," *IEEE Trans. Systems, Man, and Cybern.*, vol. 19, pp. 1489-1510, 1989.
- [3] R. D. Arnold and T. O. Binford, "Geometric constraints in stereo vision," *Proc. SPIE*, vol. 238, pp. 281-292, 1980.
- [4] H. H. Baker and T. O. Binford, "Depth from edge and intensity-based stereo," *Proc. 7th Int. Joint Conf. Artif. Intell.*, pp. 631-636, 1981.
- [5] W.E.L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, pp. 17-34, 1985.
- [6] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, pp. 139-154, 1985.
- [7] H. K. Nishihara, "PRISM: A practical realtime imaging stereo matcher," *AI Memo 780*, MIT, 1984.
- [8] L. H. Quam, "Hierarchical warp stereo," *Proc. DARPA Image Understanding Workshop*, pp. 149-155, 1984.
- [9] 강민석, 김준식, 박래홍, 이쾌희, "연속 항공영상의 스테레오 모델링에 의한 지형 복원," *전자공학회논문지*, 제30권 B편 제9호, pp. 64-75, 1993
- [10] A. Witkin, D. Terzopoulos, and M. Kass, "Signal matching through scale space," *Int. J. Computer Vision*, vol. 1, pp. 133-144, 1987.
- [11] B. D. Lucas and T. Kanade, "An ite-

- rative image registration technique with an application to stereo vision," *Proc. Int. Joint Conf. Artif. Intell.*, pp. 674-679, 1981.
- [12] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-16, pp. 920-932, 1994.
- [13] R. M. Haralick, "Digital step edges from zero crossings of second directional derivatives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 58-68, 1984.
- [14] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, pp. 697-698, 1986.
- [15] F. Devernay and O. D. Faugeras, "Computing differential properties of 3-D shapes from stereoscopic images without 3-D models," *Proc. 1994 IEEE Conf. Computer Vision and Pattern Recognition*, pp. 208-213, Jun. 1994.
- [16] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B* 207, pp. 186-217, 1980.

 저 자 소 개



金 洛 賢(正會員)

1959년 12월 1일생. 1982년 서울대학교 전자공학과(공학사). 1984년 서울대학교 전자공학과(공학석사). 1989년 University of Texas at Austin(공학박사). 1989년 ~ 1993년 삼성전자 컴퓨터부문 연구원. 1993년 3월 ~ 현재 한국외국어대학교 제어계측공학과 조교수. 주관심분야는 컴퓨터 비전, 영상처리, OA software 설계 등임