

스테레오 비전을 이용한 가상환경구현

이 희 만*

The Implementation of Virtual Environment by Using Stereo Vision

Heeran Lee*

요 약

본 논문에서는 스테레오 정합을 위한 수렴반복 알고리즘을 제안하였으며, 이 알고리즘은 가상 스튜디오를 위한 가상환경 구축하는 데 적용되었다. 스테레오 정합을 위하여 2차원 에러함수를 정의하여 사용하였으며 3차원 정보를 얻을 수 있는 변위정보는 정의된 에러함수의 값이 반복 수렴에 의해 최소화될 때 얻어진다. 스테레오비전에서의 투영에러와 폐색문제에 대처하기 위하여 변화하는 크기의 윈도우를 사용한다. 2쌍의 스테레오 이미지로부터 얻은 심도정보는 Z-Mixing방법에 의해 가상환경을 구축하는 데 이용하였다. 본 논문에서 제안한 스테레오 알고리즘의 실험 결과는 가상 스튜디오 구축에 잘 적용될 수 있음을 알 수 있다.

Abstract

In this paper, an iterative algorithm for stereo registration is proposed. The proposed algorithm is used for reconstructing virtual environment in a virtual studio. The second order error function is employed for stereo registration. The disparity information is obtained by minimizing the error function in an iterative manner. The variable window sizes are used to cope with the projection error and occlusion problem in the stereo vision. The depth information obtained from two pairs of stereo images is used for creating virtual environment by Z-Mixing. The experiment results proves the possibility of applying the proposed algorithm to virtual studio.

▶ Keyword : stereo vision(스테레오비전), image registration(영상정합), virtual studio(가상 스튜디오)

• 제1저자 : 이희만
• 접수일 : 2004.08.28, 심사완료일 : 2004.09.08
* 서원대학교 컴퓨터정보통신공학부 교수

I. 서론

컴퓨터 비전에 있어 복수개의 영상정보로부터 심도 정보의 추출은 그 동안 주요 연구대상이었다[1]. 심도의 정보는 로보틱스, 물체의 인식 등에 있어서 중요한 정보로 활용되며 최근에는 가상현실의 구현에도 점차 적용이 되고 있다[2][3]. 최근 컴퓨터 비전의 3차원 정보구현 기술발전은 새로운 개념의 스튜디오, 즉 가상 스튜디오구축을 가능하게 하고 있다[4][5]. 가상 스튜디오는 기존 스튜디오와 컴퓨터의 처리기능을 상호 긴밀하게 결합한 새로운 개념의 스튜디오이다[6]. 가상 스튜디오에서는 물체 또는 연기자의 3차원 모양이나 재질을 포착하여 가상 화면으로 재구성한다. 이런 새로운 응용분야에서의 어려움은 심도정보 추출에 에러가 발생되며, 이 에러가 합성화면에서 쉽게 눈이 보인다는 점이다. 비디오 이미지를 그래픽 화면에 삽입하거나 삭제하는 방법, 또는 역으로 그래픽 물체를 비디오 화면에 삽입 및 삭제를 하는 방법은 가상 스튜디오에서 어려운 기술이다[7][8][3]. 3차원 화면을 복수개의 카메라를 이용하여 재구현하는 방법은 능동방식 구현시스템과 비 능동방식의 2가지가 있다[3]. 능동방식은 Cyberware(미국), Vitronic(독일)과 같이 구조화된 패턴조명을 이용하는 방법으로 이런 조명을 사용하지 않는 비 능동형 방법보다 신뢰성 있는 정보를 추출할 수 있으나 동적으로 움직이는 물체의 3차원 정보추출에는 어려움이 많다. 구조화된 조명을 사용하지 않는 비 능동형 방식으로 스테레오 비전 방식이 있다. 본 논문에서는 가상환경의 구축을 위하여 비 능동형 방식의 스테레오 비전을 사용한 새로운 수렴반복 방식의 스테레오 정합방법을 제안한다.

II. 선행연구

영상정합은 스테레오비전 또는 복수 영상을 사용하여 3차원 정보를 추출하는 방법에 있어 매우 중요한 역할을 한

다. 영상 정합은 3차원 정보의 추출에 사용될 뿐만 아니라 또한 모션의 분석 또는 패턴의 인식에도 사용되는 주요기술이다[9]. 전통적인 정합방식은 주어진 2개 영상의 밝기 함수에서 불일치(Disparity) 벡터를 구하는 것이다[10]. 도출한 불일치 벡터에서 3차원 정보의 추출과 기하학적 관계 정보의 추출이 가능하다.

영상정합에 있어 가장 어려운 점은 서로 일치하는 점이나 또는 영역을 알아내는 것이다. 스테레오 비전에는 전통적으로 2가지 접근 방법이 있다. 첫 번째 방법은 특징정보를 이용하는 방법이고[11][12], 두 번째 방법은 영상의 영역정보를 이용하는 것이다[4][13][14]. 특징정보를 이용하는 방법은 특징정보가 위치한 지점의 불일치 벡터를 구하게 됨으로서 부분부분 정보가 없는 지점이 발생하는 단점이 있는 반면 영역기반 방법은 조밀한 불일치 벡터를 얻을 수 있는 장점이 있다. Mar와 Poggio[15]는 스테레오비전 연구의 선구자로 영상의 모서리 정보를 이용하여 불일치 벡터를 추출함으로써 특징정보를 이용하는 방법에 많은 영향을 주었다.

가상스튜디오를 위한 가상환경 구축에 있어서는 조밀한 불일치 벡터정보가 선호된다. 영역기반의 정합방식은 주위 픽셀의 정보를 복수개의 영상에서 상호 비교하여 가장 정합이 잘되는 위치를 찾아내는 방법이다. 영역기반 정합에 많은 영향을 미치는 것은 얼마나 큰 또는 작은 영역을 상호 비교를 하여야 하는지 즉, 윈도우의 크기를 어떻게 결정하여야 하는가에 달려있다. Kanade[4]는 윈도우의 직경의 윈도우 크기를 정하는 방법을 집중 연구 및 실험을 하였다. 영역 기반의 정합에 있어 또 다른 어려운 점은 복수개의 영상을 얻는 과정에서 각각 카메라의 위치가 상호 다르게 됨으로 프로젝션 에러가 존재한다는 점이며 또한 같은 지점의 밝기정보도 보는 각도에 따라 달라짐에 따라 윈도우에서의 영역정보가 상호 다르다는 점이다. 이러한 문제점을 보완하기 위하여 윈도우 크기를 적절히 가변시키는 적응형 윈도우 방식이 제안되었으며[4], 또는 윈도우의 모양을 표면의 방향에 따라 변형시키는 방법도 연구 발표되었다[16]. 또한 카메라의 사각지역에 따라 물체의 표면이 한쪽 방향의 카메라에만 보이게 되어 좌우 영상에서 영역이 서로 매칭되지 않는 영역이 존재하는 폐색현상(Occlusion)이 있다.

Fua는 이런 문제를 왼쪽/오른쪽 영상을 교차로 기준으로 하여 정합을 하고 서로 일치하지 않는 부분을 기각하는 쌍방향 방법을 제안하였다[17]. Lawrence와 Zitnick[18]는 반복 방식을 제안하여 위 문제를 해결하고자 시도하였다. 이러한 방법 등은 다소 개선된 결과를 보여주나 컴퓨터 코

스트가 많이 든다. 본 논문은 특정 정보와 영역기반의 정보를 함께 사용하여 영상정합을 하며 정합된 위치를 통하여 불일치 벡터를 추출하고 심도정보를 얻음으로서 가상환경에 적용 가능하도록 한다.

가상스튜디오는 제3자 즉 방송을 보는 사람에게 가상환경에서 사람 또는 물체가 어색하지 않고 자연스럽게 상호작용하는 연속적인 비디오 동영상으로 가상현실을 제공한다. 가상 스튜디오 기술을 처음으로 시도한 곳은 1991년 일본 NHK로 당시 과학 도구멘터리 영상물을 제작하였다[6]. NHK는 전경화면(foreground)에 트래킹 시스템을 부착하고 배경화면은 컴퓨터 그래픽스로 렌더링을 하였다. 그 이후 Plaform, IMP, ELSET 및 VAP 등 많은 시스템이 등장하였고 많은 상업시스템도 등장하고 있다. 가상스튜디오의 가장 큰 어려움은 정확한 심도정보를 사용하여 가상환경에서 어색하지 않도록 영상을 합성하는 데 있다. 가상스튜디오에서 심도 정보를 얻는 것은 일반적으로 매우 어렵다. 일반적으로 전경에 있는 물체는 일정한 상수의 심도정보를 사용하여 합성하는 경우가 대부분이며 이 기술은 동적인 응용분야에 적용하기에는 어려움이 많다. 본 논문은 스테레오 비전 기법을 이용하여 정확한 심도정보를 추출하여 영상을 합성하도록 함으로서 추후 동적인 응용분야에도 적용 가능하도록 할 수 있다.

III. 수렴반복 알고리즘

본 논문의 스테레오 영상정합은 임의의 함수 $f(x)$ 와 이의 변이된 위치의 함수 $f(x-D_T)$ 에서 불일치 정도 D_T 를 찾아내는 방법에서 출발한다. 스테레오 좌/우측 영상 에피폴라 선상의 영상의 밝기를 함수로 표현하고 2개의 함수가 불일치 D_T 만을 갖는 같은 함수라고 가정한다. 실제 2개의 함수는 같은 함수가 아니지만 전체적으로 매우 유사함으로 좋은 출발 가정이 될 수 있다. 본 논문은 이런 가정 하에 알려지지 않은 불일치 D_T 를 구한다. 불일치 D_T 값을 구하기 위하여 2차 정규 에러함수를 다음과 같이 정의한다.

$$e(d) = \int_{-\infty}^{\infty} (f(x-D_T+d) - f(x))^2 dx \dots\dots\dots (1)$$

이 수식에서 정합된 위치 $d=D_T$ 인 지점은 $e(d)$ 함수가 최소가 되는 곳이다. 위 수식 (1)에서 최소가 되는 지점을 수렴반복을 통하여 알아내기 위하여 다음과 같이 수정한다.

$$e_k(d) = \int_{-\infty}^{\infty} (f(x-D_T+d+D_k) - f(x))^2 dx \dots (2)$$

여기에서

$$D_k = D_{k-1} - \alpha \nabla e_{k-1}(0) \dots\dots\dots (3)$$

알고리즘은 $D_0 = 0$ 에서 시작하며, 수식(2)와 (3)을 수렴값을 얻을 때 까지 반복한다. 즉 $e_k(d)$ 함수가 $e_k(0)$ 에서 최저치가 되도록 한 지점 D_{k-1} 에서 D_k 로 함수를 전이시킨다. D_k 값은 매 반복 계산 때마다 $-\nabla e_{k-1}(0)$ 기울기 방향으로 스텝과 α 배 만큼 이동된다. 만약 $e_k(0)=0$ 가 된다면 이 지점은 $D_k = D_T$ 가 되는 지점으로 2개의 함수가 정합되는 위치이다. 이 알고리즘은 $e_k(0)$ 에서 미분값을 얻을 수 있다는 전제 조건에서 출발한다. 만약 $e_k(d)$ 함수가 최저 값 근처에서 가파른 경사를 이룬다면 최소치의 접근은 매우 빠르게 수렴할 것이다. 그러나 본 논문에서는 이산적 데이터를 취급하므로 $-\nabla e(0)$ 값에 따라 정규화된 값 즉 ± 1 만을 사용하여 D_k 을 변경한다.

IV. 스테레오 이미지에 적용

3장에서 언급한 수렴반복 알고리즘을 스테레오 비전에 적용한다. 스테레오비전의 목표는 좌/우 2개의 이미지로부터 심도 정보를 얻는 데 있다. 대부분의 스테레오비전 알고리즘은 4단계로 이루어져 있다. 즉 정합 예정 물체를 선정하는 단계, 2개의 영상으로부터 정합하는 단계, 실제 카메라 또는 가상의 카메라의 파라미터를 결정하는 단계, 마지막으로 카메라로부터 정합된 물체의 거리를 계산하는 단계로 구성된다[10]. 이중 가장 어려운 단계는 정합단계로 주어진 2개의 영상으로부터 서로 대응되는 위치를 발견하는 것이 매우 어려운 과제이다. 스테레오비전에서 정합의 위치를 찾는 경우 에피폴라(Epipolar) 제한을 이용하는 것이

일반적이다. 좌우 카메라의 기하학적 위치 때문에 대응될 상호 위치는 E선상으로 제한되기 때문이다.

본 논문에서는 정의한 2차 에러함수의 최저 값을 수렴반복에 의해 찾아내어 정합되는 위치를 알아낸다. 또한 실험에서는 카메라의 광축이 평행한 경우만을 고려하였다. 에피폴라 선상의 이미지의 밝기 정보를 함수로 생각할 때 좌우 영상의 함수는 같지는 않지만 매우 유사한 함수이다. 즉 전체적 관점(global point of view)에서 함수는 근사적으로 같다고 볼 수 있지만 국부적(local point of view)으로는 조금씩 다르다. 스테레오비전 정합에서 오류정합 문제가 발생하는 데 이는 좌우 카메라의 사각지대가 각기 다르기 때문에 발생한다. 또한 물체의 표면의 방향이 카메라에 따라 다르므로 프로젝션할 때 국부적으로 각각 서로 다른 함수가 될 수 있으며 또한 CCD의 감도가 카메라 마다 서로 다르므로 빛의 밝기 또한 많이 다르다. 그러므로 스테레오 정합에 있어 전체적 관점에서 정합을 하고 국부적 관점에서 정합을 할 필요가 있다. 3장에서 제시한 수렴 반복 알고리즘은 이를 위해 다소 수정 보완하여야 한다. 본 연구에서는 윈도우 함수를 도입하고 윈도우의 크기를 가변시킴에 따라 전체적 관점과 국부적 관점의 정합을 가능하도록 한다. 즉 윈도우 크기를 크게 하면 전체적 관점에서 정합을 할 수 있으며 윈도우 사이즈를 적게 만들면 국부적 정합을 할 수 있다. 그러므로 우선 전체적 정합을 하고 세세한 부분을 국부적 정합과정을 하도록 한다.

스테레오 비전을 위한 알고리즘은 다음과 같다. 먼저 왼쪽영상 에피폴라 선상의 영상 밝기정보를 함수 $F(x)$ 로 설정하고 오른쪽 영상을 $G(x)$ 라고 하자. 반복수렴 알고리즘을 위해 2차 정규 에러함수를 수식(4)와 같이 표현한다.

$$E_k(d) = \int_{-\infty}^{\infty} W(x)(F(x) - G(x + d + D_k))^2 dx \dots (4)$$

여기에서 $W(x)$ 는 윈도우 함수이다. D_k 는 디지털 이미지 이므로 정수 값만을 갖도록 한다. 알고리즘은 다음의 6단계로 구성된다.

단계 1 : 초기화 $D_k = 0$

WindowSize = MaxSize

(예상되는 최대 변이 값보다 최소한 2배 이상 크기로 설정한다.)

단계 2 : 왼쪽 영상에서 기준위치를 설정하고 그 위치에서 기울기 $\nabla E(0)$ 을 계산한다. 만일 $|\nabla E(0)| < \epsilon$ 이

면 단계 6을 한다.

단계 3 : $\nabla E(0) < 0$ 이면 $p = 1$

$\nabla E(0) > 0$ 또는

WindowSize = MaxSize이면 $p = -1$

단계 4 : D_k 값을 다음과 같이 변경한다.

$$D_k = D_{k-1} + p$$

단계 5 : $G(x)$ 를 이동하고 단계 2를 수행한다.

$$G(x) \leftarrow G(x + D_{k+1})$$

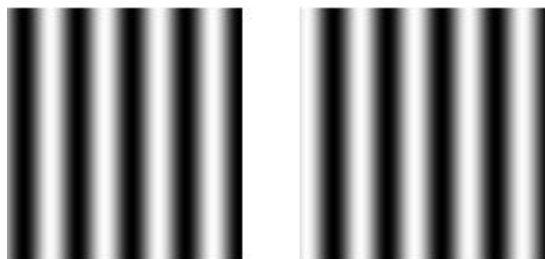
단계 6 : 윈도우 사이즈를 변경한다.

$$\text{WindowSize} = \text{WindowSize} - \beta$$

만일 WindowSize가 MinSize보다 적게되면 알고리즘을 종료하며 아직 MinSize보다 크게되면 단계 2를 수행한다. (MinSize는 예상되는 최소 변이 값보다 작거나 같도록 설정한다.)

V. 실험결과

본 논문에서 제안한 방법은 일종의 영역기반의 정합방법과 유사한 방법이지만 실험에서는 단계 2의 기준위치의 선정에 있어 특징정보를 이용하였다.



(a) 기준 영상

(b) 변이된 영상

그림 1. 단일 변이 값을 갖는 정현파 영상
Fig 1. Sine wave images with single disparity

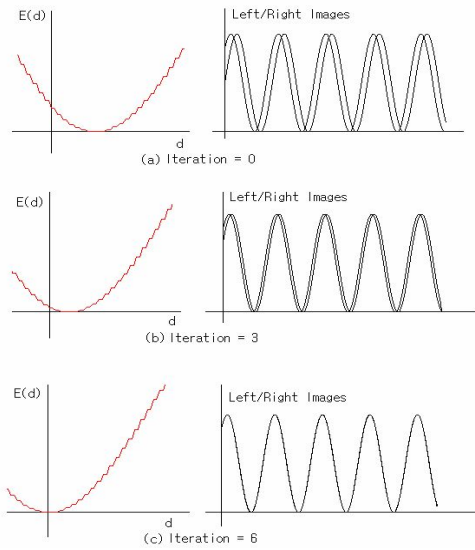


그림 2. 정현파 영상의 실험결과
Fig 2. Experiment results of sine wave images



그림 3. 직사각형 스테레오 영상
Fig 3. rectangular stereo images

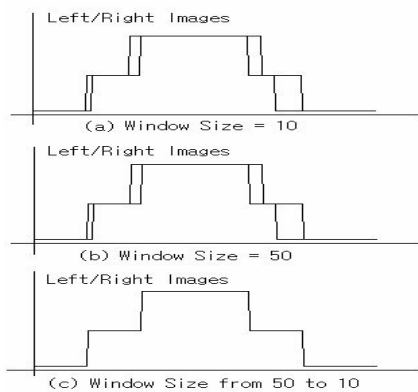


그림 4. 윈도우 사이즈 실험
Fig 4. window size experiment



(a) 중첩영상 (b) 정합영상 (c) 전경 영상

그림 5. 전경 스테레오 영상의 처리결과
Fig 5. Experiment results of foreground images



(a) 중첩 영상 (b) 정합 영상



(c) 심도정보 (d) 합성영상

그림 6. 배경 영상과 합성영상 실험결과
Fig 6. Experiment results of background images and image synthesis

즉, 왼쪽 영상에서 모서리가 검출된 위치만을 기준위치로 선정하여 정합과정을 하였으며 중간 지점은 보간법으로 구한다. 본 알고리즘은 특징지점을 먼저 정합하고 중간부분을 지역적으로 정합하는 방법에도 적용가능하며 또는 낮은 해상도에서 높은 해상도로 단계별로 적용할 수도 있다. (그림 1) 은 알고리즘의 이해를 위해 단일 변이를 갖는 정현곡선 영상을 보였으며 (그림 2)는 이를 대상으로 실험한 것이다. 왼쪽 영상의 임의의 점을 기준점으로 선정하고 알고리즘이 진행되면서 $E(0)$ 값이 최소화 되는 과정을 보인 것으로 최소화된 위치가 정합된 위치임을 보여준다.

(그림 3)은 스테레오비전에서 자주 볼 수 있듯이 복수개의 변이를 갖으며 좌우 영상에서의 상호 크기의 상이함을 실험하기 위하여 인위적으로 제작된 영상이다. 일반적으로 영역기반의 스테레오 알고리즘들은 비교하는 윈도우의 크기의 선정에 따라 결과가 달라진다[4]. 본 알고리즘은 크기가 큰 윈도우에서 먼저 정합을 하고 그 결과를 바탕으로 점차 윈도우의 크기를 감소시키면서 세부 정합을 하도록 되어 있다. (그림 4)의 (a)(b)에서 알 수 있듯이 단일 윈도우의 크기만으로는 정확한 결과를 얻을 수 없다. 그러나 윈도우 크기를 크게 하여 전체적 관점에서 정합을 먼저 시도하고 윈도우 크기를 점차 축소하면서 국부적 관점에서 정합을 시도한 (c)는 정확한 정합 결과를 얻고 있음을 알 수 있다. 이 실험에서 왼쪽 영상의 정합 기준위치는 모서리 정보가 있는 위치만을 사용하였다. 그 외의 지점은 보간법으로 구하거나 또는 일정 경험 법칙에 따라 추출할 수 있다. (그림 5)의 (a)는 일반 스튜디오에서 얻은 스테레오 영상을 단순 중첩한 것이며, (b)는 정합과정을 통하여 얻은 영상과 기준영상(좌측)을 중첩한 것이며 (c)는 정합과정에서 얻어진 변이정보에서 특정 변이 정보를 갖는 픽셀 값을 선택한 영상으로 가상 스튜디오의 전경 영상으로 추출한 것이다. (그림 6)의 (a)(b)는 컴퓨터 그래픽을 통하여 구현된 스테레오 이미지[19]를 (그림 5) 같은 순서로 보인 것이며, (c)는 정합과정에서 추출된 변이 정보를 영상으로 보였으며, (d)는 (그림 5)의 변이정보와 (그림 6)의 변이정보를 이용하여 Z-Mixing 방법으로 합성한 영상으로서 가상환경을 구현한 영상을 보인 것이다. 본 연구에서 제안한 방법을 통하여 비디오 동영상의 변이정보 즉 심도정보를 이용하여 가상환경을 구축할 수 있음을 알 수 있다.

VI. 결론

본 논문에서는 가상환경 구축을 위한 스테레오 영상의 정합방법을 제안하였다. 스테레오 정합방법은 수렴 반복법으로 본 연구에서 제안한 2차 에리함수를 이용하며, 이 함수의 최저치를 구하는 과정에서 변이정보를 얻게 된다. 실험에서는 2개의 카메라의 광축이 평행한 페러렐 스테레오 모델을 사용하였으며, 윈도우의 크기를 가변함으로써 전체적 관점과 국부적 관점에서의 정합이 되도록 하였다. 전체

적 관점에 의한 정합은 스테레오 비전의 정합을 어렵게 만드는 투영변환에러와 폐색 문제를 효과적으로 대처하게 할 수 있다. 두 쌍의 스테레오 영상으로부터 추출한 심도정보를 Z-Mixing하여 가상환경 구축을 가능하게 한다.

참고문헌

- [1] 신장호, 김계국, "TV영상의 3차원 변환을 위한 공간 분석 알고리즘에 관한 연구," 한국컴퓨터정보학회, vol.7, no.4, 2002.
- [2] Richard Szeliski, "Stereo Matching with transparency and Matting," International Journal of Computer Vision, pp1-17, 1998.
- [3] Oliver Grau, "A Combined Studio Production System for 3D Capturing of Live Action and Immersive Actor Feedback," IEEE Trans. Circuits and System for Video Technology, vol.14, no.3, 2004.
- [4] T. Kanade, "A Stereo matching algorithm with an adaptive window: Theory and Experiment", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.16, no.9, pp.920-932, 1994.
- [5] Takeo Kanade, "Virtualized Reality : Constructing Virtual Worlds from Real Scenes," IEEE Multimedia, pp.34-47, Jan, 1997.
- [6] Simon Gibbs, "Virtual Studios: An Overview," IEEE Multimedia, pp18-35, Jan, 1998.
- [7] Paker, Y, "Image Processing for Broadcast and Video Production," Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, pp.22-24, Nov, 1994.
- [8] Blonde, L, "A Virtual studio for Live Broadcasting: The Mona Lisa Project," IEEE Multimedia, 3(2), pp.18-19, 1996.

- [9] 서정만, 정순기, "모션벡터를 이용한 가상현실 체험 시스템의 구현," 한국컴퓨터 정보학회, vol. 7, no.3, 2003.
- [10] Bruce D. Lucas, "An Iterative image registration Technique with an Application to Stereo Vision," Proceedings of Imaging Understanding Workshop. pp.121-130, 1981.
- [11] W.E.L. Grimson, "Computational Experiments with a feature based Algorithm," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.7, no.1, pp.17-34, 1985.
- [12] H.H. Baker, "Depth from edge and Intensity based Stereo," In Proc. Of the 7th International Joint Conference on Artificial Intelligence, pp.631-636, 1981.
- [13] M. Levine, "Computer Determination of Depth Maps," Computer Graphics and Image Processing, vol.2, no.2, pp. 131-150, 1973.
- [14] K. Mori, "An Iterative prediction and Correction Method for Automatic Stereo Comparison," Computer Graphics and Image Processing, vol.12, pp.393-401, 1973.
- [15] D. Marr, Poggio, "A Computational Experiments Theory of Human Stereo Vision," Proc. Roy. Soc. London, B204, pp.301 - 328, 1979.
- [16] D.J. Panton, "A Flexible Approach to Digital Stereo Mapping," Photogram. Eng. Remote Sensing, vol.44, no.12, pp.1499-1512, 1978.
- [17] P. Fua, "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features," Machine Vision and Applications, vol.6, pp.34-49, 1993.
- [18] C. Lawrence Zitnick, "A volumetric Iterative Approach to Stereo Matching and Occlusion Detection," CMU-RI-TR-98-30, Carnegie Mellon University, 1998.
- [19] <http://www-student.cs.uni-bonn.de/~gerdes/MRTStereo/>

저자 소개



이희만

1984년 2월 고려대학교 전자공학과 졸업(공학사)
 1986년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사)
 1986년 ~ 1990년 산업연구원
 1990년 ~ 1994년 Texas A&M Electrical Eng., Ph.D.
 1994년 ~ 1996년 삼성중공업 중앙연구소 선임연구원
 1996년 ~ 현재 서울대학교 컴퓨터정보통신공학부 교수
 <관심분야> 가상현실, 애니메이션, 컴퓨터 그래픽스

