

여러 대의 카메라를 이용한 계층적 깊이정보 추출 알고리즘

박종일, 이노우에 세이키
ATR지능영상통신연구소

— 요약 —

이 논문에서는 여러대의 카메라를 이용하여, 화면내의 모든 화소에 대해 정확한 깊이정보를 추출하는 알고리즘을 제안한다. 제안방법은 스테레오정합에 기반하고 있다. 스테레오정합에 의해 깊이정보를 추정하는데 있어서는, 오클루전(occlusion)이 추정에 미치는 악영향을 어떻게 극복하는가가 매우 중요한 문제이다. 오클루전문제는 2대의 카메라로는 근본적으로 해결이 불가능하다. 따라서 이 논문에서는 5대의 카메라를 사용하여, 각 카메라로부터의 정보를 선택적으로 사용함으로써, 이 문제를 상당히 개선한다. 한편, 스테레오 정합법에서는 정합창의 크기가 추정성능에 영향을 미치는데, 큰 창을 이용하면 잡음내성은 우수하나 깊이가 불연속인 곳에서 오차가 발생하고, 작은 창을 이용하면 잡음내성은 저하되나 불연속부에서의 창크기에 의한 오차가 줄어드는, 대차관계가 존재한다. 이 논문에서는 처음에 큰 창으로 추정하고 차츰 창의 크기를 줄여나가는 계층적 방법을 제안하여, 잡음내성을 강화하고 불연속부의 오차를 줄이고 있다. 실험결과를 통하여 제안방법의 성능을 제시하고 타당성을 확인한다.

1. 머리말

컴퓨터비전은 과거 로봇비전, 공장자동화 등 과학기술연구나 특수산업에나 응용되는 기술이었으나, 멀티미디어시대를 맞이하여, 고도의 영상처리기술에의 수요가 생겨나면서, 영상제작, 가상현실 등 인간에게 보여주는 것을 목적으로 하는 미디어기술로서 이용되기 시작하였다. 이제는 영상표현의 자유도, 영상제작의 효율을 높여주는 요소기술로서 확고한 자리를 차지하고 있다. 이 논문에서는, 컴퓨터비전 중 스테레오정합기술을 영상미디어제작에 성공적으로 이용하기 위한 방법을 다루고자 한다.

다양한 영상을 더욱 자유롭게 제작할 수 있는 틀을 만들어내고자, 우리는 현재 영상부품데이터베이스의 개발을 추진하고 있다(1). 이것은 여러 개의 영상부품이 시공간적으로 정렬된 것이 영상을 이룬다고 하는 개념에 근거하고 있다(2). 데이터베이스에는 영상, 소리 뿐만 아니라 키워드, 카메라동작, 깊이, 스크립트 등과 같은 속성정보도 함께 저장한다. 영상을 제작할 때에 이들 보조정보를 적절히 이용함으로써, 원하는 영상을 만들어낸다.

여러 가지 보조정보 중에서 자유로운 영상표현을 달성하기 위해 빼놓을 수 없는 중요한 정보가 깊이정보라고 할 수 있다. 여기에서 깊이란 카메라의 시점을 원점으로 하는 카메라 중심좌표계에서 광축을 Z축으로 설정했을 때, 화면상의 매 화소별 Z좌표값을 일컫는다. 깊이정보는, 영상을 시공간기술(記述)할 경우에, 공간에 해당하는 정보로서, 다양한 영상을 자연스럽게 자유롭게 표현, 제작하기 위해서는 꼭 필요한 정보이다. 최근에는 이를 이용하는 연구가 활발히 진행되어, 가상화된 현실감(3), Z-키 영상합성(4), 3차원특수효과(5), 영상기술(記述)의 자동화(5) 등 다채로운 응용이 시도되고 있다.

기존의 컴퓨터비전에서는와는 달리 이와 같은 응용에서는 밀(密)하고 정확한 깊이정보가 요구된다. 여기에서 밀하는 것은 화면전체의 모든 화소에 대해 구할 필요가 있다는 것을 의미한다. 정확하다는 의미는 구한 깊이맵(depth map)이 실제의 물체의 모양을 충실히 반영한다는 것을 의미한다. 경우에 따라서는, 깊이값의 정밀도 보다도 깊이맵의 기하학적 정확도에 더욱 비중을 둔다는 의미이다. 따라서, 이 논문에서는 밀하고 정확한 깊이맵을 구하는 것에 초점을

맞추고자 한다. 제안방법은 5대의 카메라로부터 들어오는 영상정보를 적절히 취사선택하여 오클루전의 영향을 극복하고, 이를 새로운 계층적 추정기 위에 구현함으로써, 밀하고 정확한 깊이맵을 구하는 것이다. 하드웨어의 실현성을 고려하여 모든 처리는 신호수준의 비반복적 처리로 한정한다.

이 논문의 구성은 다음과 같다. 우선 2장에서는 스테레오 정합에 관한 관련연구에 대해 간략히 살펴보고, 3장에서는 제안하는 방법의 구체적인 내용을 실험 예를 곁들이면서 설명한다. 4장에서는 실험결과를 제시하여 제안방법의 가능성과 한계를 논한다.

2 관련연구

스테레오정합은 영상으로부터 깊이맵을 얻어내는 매우 유용한 방법이다. 어떤 점이 위치가 다른 카메라에 비쳤을 때 그 상의 위치의 차이(視差: disparity)에 의해, 삼각측량의 원리로 깊이값을 추출한다. 시차a는 깊이Z에 역비례하므로(3장에서 설명), 시차를 구하면 깊이정보를 얻을 수 있다.

스테레오정합에는 크게 2개의 접근방법이 있다. 하나는 영역에 기반한 방법(area-based method, 이하 영역법)이고 다른 하나는 특징에 기반한 방법(feature-based method)이다. 일반적으로 밀한 깊이맵을 구하는 데에는 영역법이 주로 이용된다(6). 그러나, 널리 알려진 바와 같이 영역법은 많은 문제점을 안고 있다. 즉, 텍스처가 부족한 영역에서의 추정민감도 문제, 오클루전 문제, 시점이동에 따른 상의 변형 문제, 반복적 패턴이 있는 곳에서의 오(誤)대응 문제 등이 대표적 예라고 할 수 있다(6)(7).

컴퓨터비전에서는 이 문제를 해결하기 위해 많은 노력이 경주되었다(7)(8). 그러나, 밀한 깊이정보를 추출하기 위한 대부분의 방법은, 엄청난 양의 계산을 필요로 하거나, 반복적 계산을 요구하고 있어, 실용화에 어려움이 많았다. 그중 몇 안되는 예외라고 할 수 있는 것이 다기준(多基線: multiple-baseline) 스테레오정합이다(9)(10). 일반적인 스테레오정합법이 2대의 카메라를 사용하는데 반해, 이 방법은 더 많은 카메라를 사용하여, 계산량의 커다란 증가 없이도 텍스처 부족의 문제, 반복적 패턴의 문제 등을 상당히 개선한다. 최근에는 이를 이용한 실시간 깊이맵 추출기도 개발되었다(4).

텍스처 관련 문제와는 대조적으로 오클루전 문제는 비교

적 별다른 주목을 받지 못했다. 일반적으로 오클루전영역이 화면전체에서 차지하고 있는 비율은 상대적으로 작기 때문에, 로봇비전이 주된 응용일 경우에는 무시해도 별문제가 없었다(11). 그러나, 인간의 눈에 호소하는 영상제작에의 응용에서는 많은 문제를 야기한다. 일부분의 화질열화가 화면전체의 인상을 망쳐버리는 수가 있는 것이다. 3장에서 자세히 살펴보겠지만, 오클루전이야말로 물체의 모양을 제대로 반영하는 정확한 깊이맵을 얻는데 있어 가장 큰 방해요인이다.

카메라 2대를 사용한 스테레오정합에서는 오클루전문제는 파할 수도 없거나, 해결할 수도 없는 것이 사실이다. 단지, 일정한 가정 하에 몇 가지 지식을 동원하여, 보간할 수밖에 없다(6).

올바른 정합을 달성한다는 의미에서는 2보다 큰 다수의 카메라를 사용하는 것이 해결의 실마리를 제공할 수 있다. 어떤 영역이 한 카메라에서는 관측되지 않더라도 다른 위치에 놓인 카메라에서는 관측될 수 있어, 올바른 정합을 이룰 수 있는 것이다. Kanade 등은 오클루전문제의 해결가능성에 대해서는 언급하고 있으나 구체적으로 어떻게 해결할 것인지에 대해서는 다루지 않고 있다(4). 최근에 오클루전 문제를 본격적으로 다루고자 하는 시도가 보고되었다(12). 동일평면상에 배열된 많은 시점으로부터 취득한 영상을 이용하여, 오클루전의 패턴을 분류하고, 적절한 마스크를 사용한 정합방법을 제시하였다. 그러나 이 방법은 최소한 9대의 카메라를 사용해야 하며, 정합창의 크기에 따른 문제를 극복하려는 노력이 결여되었다.

영역법의 성능을 좌우하는 매우 중요한 요소가 정합창의 크기이다(13)(8). 정합창은 그 영역의 특징을 나타내는 충분한 휘도변화를 포함할 수 있을 정도로 커야한다. 반면에 시점이동에 따른 모양변화를 피하고, 깊이가 불연속인 곳에서의 정합의 혼동을 막기 위해서는(7) 작으면 작을수록 좋다. 이러한 딜레마를 해결하기 위해서 지금까지 2가지의 접근방법이 시도되었다. 하나는 국소적으로 적응하는 창을 이용하는 것이다(13). 이 방법은 각 화소마다 추정의 신뢰성을 가장 높일 수 있도록 창 크기 및 위치를 변화시킴으로써 부드러운 곡면과 예리한 에지를 갖는 깊이맵을 얻고 있다. 그러나, 이 방법의 문제는 반복적 처리를 필요로 한다는 것이다. 이와는 다른 접근방법으로 계층적 방법을 들 수 있다(7)(8). 전통적인 계층적 방법은, 거친 해상도에서 출발하여, 한 단계 해상도를 높이고 전단계 해상도에서의 추정결과를 토대로 그 값 부근만을 탐색하도록 하는 처리

를 단계적으로 실행함으로써, 정합창의 크기에 따른 문제를 해결하고자 하는 것이다. 이러한 개념은 인접한 화소의 시차는 크게 다르지 않다는 가정에 근거하고 있다. 이것은 깊이가 연속인 곳에서는 잘 들어맞는다. 그러나, 우리가 추정하고자 하는 깊이라고 하는 정보는 일반적으로 많은 불연속영역을 포함하고 있으며, 특히 우리가 주목하고 있는 물체의 경계선은 대체로 불연속이라고 보아야 옳다. 따라서, 이러한 영역을 올바르게 처리하려면 기존의 계층적 방법을 그대로 적용해서는 좋은 성과를 기대하기 어렵다. 이 논문의 접근방법도 기본적으로는 계층적 방법이나, 기존 방법과는 달리, 인접화소의 시차가 비슷하다는 가정을 하지 않는다. 대신에 어떤 화소의 올바른 깊이값은 큰 창을 이용하여 취득한 깊이맵 속의 그 화소 부근에 반드시 존재한다는 가정을 하고 있다.

3. 깊이정보 추출

3.1 카메라 배치

제안하는 스테레오 카메라시스템의 카메라 배치와 투영의 개념을 그림 1에 나타내었다. 가운데에 기준 카메라를 배치하고, 그 주위의 일정한 거리 L에 4개의 참조카메라를 배치한다. 모든 카메라는 광축이 평행하고, 같은 사양을 갖는 것으로 가정한다. 사양의 차이나 광축의 방향 등이 가정과 일치하지 않는 경우에는, 카메라교정(calibration)을 통해 보정하여, 그림에 나타난 것과 같은 기하구조로 고친 다음 스테레오정합을 실행한다. 동영상에 적용하기 위해, 모든 카메라는 동기시킨다(1).

우리가 구하고자 하는 것은 가운데의 기준카메라의 영상에 대한 깊이맵이므로 다른 참조카메라는 센서로서의 역할을 담당하는 셈이 된다.

그림 1의 배치에서는 3차원공간의 점 $P=(X, Y, Z)$ 는 기준화면 위의 점 $p_0=(x_0, y_0)$ 에 투영된다. 여기서 $x_0 = FX/Z$, $y_0 = FY/Z$, F는 초점거리를 나타낸다. 점 P는 또한 참조카메라 $C_i(i=1, 2, 3, 4)$ 의 화면상의 점 $p_i=(x_i, y_i)$ 에도 투영된다. 여기서,

$$x_i = F \frac{X - D_{i,x}}{Z}, \quad y_i = F \frac{Y - D_{i,y}}{Z}$$

단, 기선장(基線長) 벡터 $D_i=(D_{i,x}, D_{i,y})$ 는 각각 $D_1=(L, 0)$, $D_2=(-L, 0)$, $D_3=(0, L)$, $D_4=(0, -L)$ 이다. 이러한 구조에서, 점 P의 시차의 참값 d_i 는

$$d_i = \frac{FL}{Z} = |p_i - p_0| \text{ -----(1)}$$

의 관계가 있으므로, 시차를 구함으로써 깊이값을 얻을 수 있다.

이상과 같은 카메라구조는, 어떤 화소에 대해 어느 하나의 카메라로부터 오를루전 때문에 올바른 정합을 얻을 수 없는 경우에, 대칭되는 위치에 놓인 다른 카메라로부터는 올바른 정합을 얻을 수 있다는 가정에 근거하고 있다. 실제로 이 구조는 아주 작은 구멍, 좁고 긴 틈새 등의 특수한 경우를 제외한 대부분의 오를루전에 대해 올바르게 대응할 수 있다.

이러한 관점에서 본다면, 카메라의 대수는 많을수록 좋고 할 수 있다. 그러나 실용적인 면을 고려했을 때에는 적을수록 유리하다. 따라서 이 논문에서는 계산량, 카메라교정의 복잡도 등을 고려하여, 카메라의 대수를 오를루전에 올바르게 대응할 수 있는 필요최소한이라고 할 수 있는 5대로 설정하고 있다. 다음 절에 오를루전영역 부근에서의 오(誤)정합을 억제하기 위한 방법을 설명한다.

3.2 오를루전 극복형 스테레오정합

정합의 척도로서 일반적으로 가장 많이 사용되고 있는 SSD(sum of squared-difference)를 사용한다. 기준화면의 한점 x에 대해, 각 참조화면의 에피폴라선(epipolar line: 이 논문의 카메라배치에서는 수평 및 수직선)에 따라, 정합창을 1화소씩 이동시키면서, SSD

$$e_i(x, d) = \sum_{b \in W} [I_0(x + b) - I_i(x + b + d)]^2$$

를 계산한다. 여기서 b 는 기준화면의, I_i 는 참조화면의 휘도값, W 는 정합창을 나타낸다.

이 논문의 카메라배치에서는 모든 참조카메라에서 관측 가능한 점에 대해서는 시차의 크기 $d=|d_i|$ 가 일정하다. 따라서, 다기선 스테레오정합법(10)을 그대로 적용하면 다

1) 동영상에 대응할 수 있다는 것이 스테레오정합의 큰 장점이라고 할 수 있다. 현재의 기술수준에서 동영상에 대응할 정도의 스캐닝속도를 갖는 능동센서는 존재하지 않으며, 설사 있더라도 능동센서의 위치와 기준카메라의 시점이 일치하지 않기 때문에 정확한 깊이맵을 얻는데는 많은 어려움이 있다.

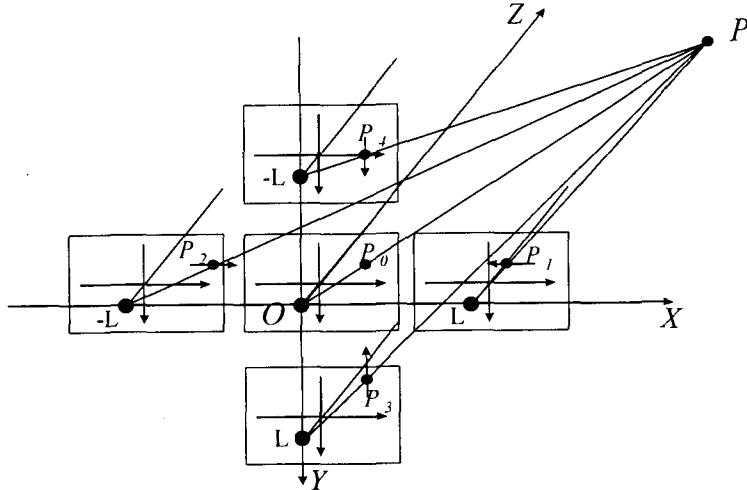


그림 1. 카메라시스템의 투영개념도.

Fig 1. Projection geometry of camera system.

음과 같은 식으로 나타낼 수 있다.

$$\hat{d}(x) = \arg \min \sum_{i=1}^4 e_i(x, d). \quad (2)$$

깊이가 연속적으로 변하는 영역에서는 이 방법에 의해 양호한 결과가 얻어진다. 그러나, 깊이가 불연속적인 영역 부근에서는 그렇지 못하다. 4방향으로부터의 정합데이터 모두가 유용한 정보를 포함한다고 볼 수 없기 때문이다. 오히려, 일부의 데이터, 주로 오클루전영역이 생기는 방향으로부터의 데이터는 추정에 악영향을 끼친다.

이해를 돕기 위하여, 여기서부터는 Tsukuba대학 다시점 화상데이터베이스의 "City"화상을 사용한 실험결과를 제시하면서 설명하기로 한다.

그림 2를 통해 오클루전이 어떻게 악영향을 끼치는지를 설명한다. 물체의 경계선 부근의 어떤 점에 대해 4방향 각각의 e_i 곡선(정합창의 크기는 7×7 화소)을 나타내었다. 식 (2)를 이용하여 추정하면, 오클루전의 영향, 구체적으로 왼쪽 및 아래의 카메라로부터의 SSD가 시차의 참값 ($d \approx 6$)에서 큰 값을 가짐으로써 올바른 정합을 방해하는 것, 때문에 바람직하지 못한 추정결과($d \approx 19$)를 내고 만다. 이와 같은 이유로 식(2)에 의한 추정은 깊이가 불연속인 곳에서 실패하게 된다. 따라서, 올바른 정합을 방해하는 데이터는 제거하여야 한다.

정합을 행할 때, 나쁜 관측데이터를 제대로 제거할 수 있다면 좋은 추정결과를 기대할 수 있다. 이 논문에서는 e_i 곡선의 특성을 관찰한 결과를 토대로 간단한 방법을 제안한다. 많은 경우에 대해, 오클루전영역 부근의 e_i 곡선을 살펴본 결과, 대부분의 경우 4개의 관측데이터 중에 2개 이상은 올바른 정합정보를 갖고 있다는 사실을 확인할 수 있었다. 위의 그림 2도 그러한 예의 하나이다. 따라서, 4방향으로부터의 관측데이터 중에 2개의 데이터는 올바른 정보를 갖고 있다고 가정하여 다음과 같이 추정을 행한다.

어느 화소를 둘러싼 정합창을 a 만큼 이동시켰을 때, 4방향으로부터의 $e_i(x, d)$ 가 얻어진다. 그 가운데 큰 것 2개는 버리고, 나머지 2개만을 가산한다. 이러한 처리를 $d=1, 2, \dots, SR$ 에 대해 수행한 후, 그 값이 최소가 되는 a 를 그 점의 시차로 정한다. 이상을 식으로 나타내면

$$\hat{d}(x) = \arg \min \sum_{i=1}^2 \tilde{e}_i(x, d) \quad (3)$$

이 된다. 여기서 e_i 는 e_i 를 소트한 것으로 모든 $i < j$ 에 대해 $e_i \leq e_j$ 이다.

이상의 방법을 이용하여 구한 깊이맵을 그림 3에 나타내었다. 왼쪽의 맵은 대기선 스테레오정합을 그대로 적용한 경우인데, 깊이가 불연속적인 곳 근처에서 오차가 상당히 많이 눈에 띈다. 오른쪽의 맵은 제안수법에 의한 것으로,

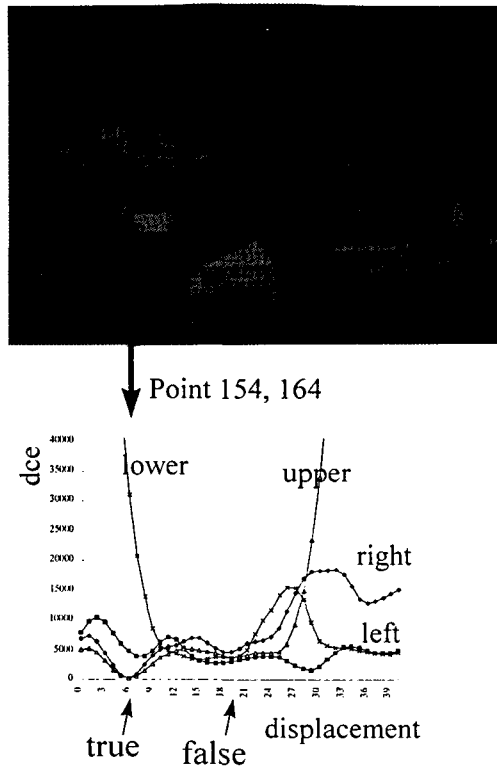


그림 2. 다수의 카메라를 이용한 스테레오정합에서의 오클루 전의 영향.
 Fig 2. Illustration of the influence of occlusion in stereo matching from multiple cameras.

물체간의 경계선 부근의 오차가 거의 사라졌음을 확인할 수 있다. 그러나 4개의 정보 중 2개의 정보만을 사용하기 때문에 예상했던 대로 잡음내성이 약간 저하되었다(10). 이러한 열화는 후술하는 계층적 수법에 의해 보상하게 된다.

그림 3의 깊이맵에서는 2종류의 오차가 관측된다. 하나는 오클루전에 의한 것으로서, 왼쪽의 맵에는 많이 존재하지만, 오른쪽의 맵에는 거의 존재하지 않는다. 다른 하나는 물체간의 경계선이 연장되는 현상(boundary overreach, 이하 BO)으로서, 양쪽 모두 존재한다. 이것은 Cochran 등이 지적하는 바와 같이(6), 텍스처가 많은 영역의 깊이값이 적은 영역으로 스며드는 현상이다. 이러한 현상은 정합창을 사용하는 방법에서는 피할 수 없는 것이다. 그런데 한 가지 재미있는 것은 오클루전 극복형 정합에 있어서는 이와 같은 깊이맵의 변화가 급격하게 일어나고, BO의 크기도 대략 정합창의 크기의 1/2이하로 제한된다는 사실이다. 그림 4에 이러한 현상을 설명하고 있다. 여기에서 정합창의 크기는 15x15화소인데, BO의 크기는 7화소 이하이고 급격하게 변화한다는 사실을 확인할 수 있다. 이러한 관찰은 다음 절에 제시하는 계층적 수법을 구성하는데 있어 매우 중요한 힌트가 되었다.

물체의 경계선과 깊이맵의 불연속선이 일치하도록 하기 위해서는 BO의 크기를 가능한 작게 해야하는데, BO의 크기는 정합창의 크기에 비례하기 때문에 정합창의 크기가 작으면 작을수록 좋다. 그러나 정합창의 크기가 작으면 잡음내성이 저하되어 추정의 신뢰성이 떨어지므로 오차가 많

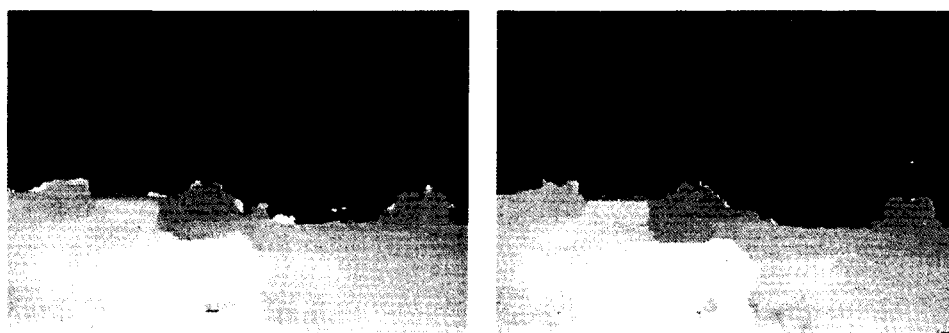


그림 3. 다수카메라를 이용한 스테레오정합으로 얻은 깊이맵. 왼쪽은 오클루 전대책이 없는 것이고 오른쪽은 있는 것임. 정합창의 크기는 15x15 화소.
 Fig 3. Depth map from multi-camera matching. Left map is obtained without occlusion-overcoming strategy and right one is with the strategy. Matching window size is 15x15.

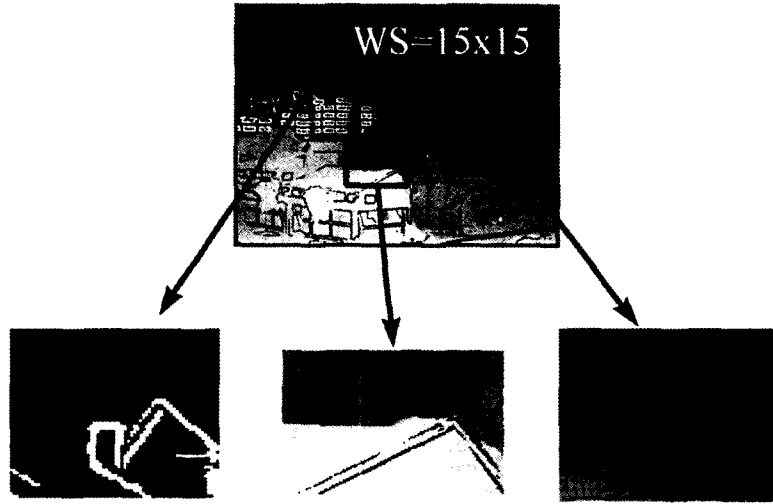


그림 4. 경계선연장 현상.
Fig 4. Boundary overrun.

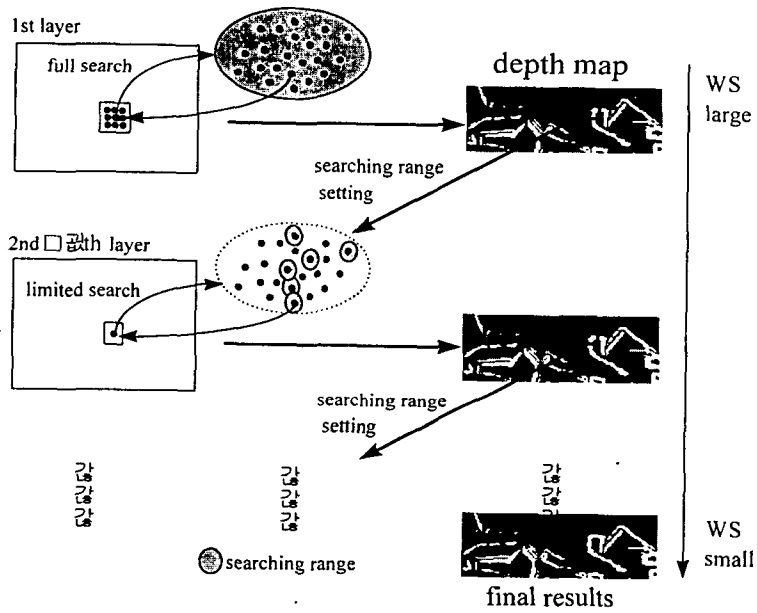


그림 5. 계층적 추정법의 개념도. 제2층부터는 탐색영역이 제한됨.

Fig 5. Illustration of the hierarchical scheme. Searching range is restricted to the estimates in the matching window of the previous layer except for the 1st layer.

이 발생한다. 정합의 신뢰성과 깊이맵의 기하학적 정확성 사이에는 트레이드오프가 존재하는 것이다. 따라서 이 논문에서는 이 문제를 해결하기 위해 계층적 방법을 제안한다.

3.3 계층적 추정법

제안하는 계층적 방법은, 오클루전 극복형 깊이추정을 하여 얻은 깊이맵에 있어, 임의의 점에 대한 올바른 깊이값은 그 점을 둘러싼 정합창의 범위 내에 거의 대부분 존재한다는 관찰에 근거하고 있다. 우선, 큰 정합창으로부터 시작한다. 유용한 깊이정보는 제일 큰 정합창을 사용한 깊이맵 안에 전부 포함되어 있다고 가정한다. 점차 정합창의 크기를 줄이면서 깊이맵을 갱신한다.

그림 5에 제안하는 계층적 방법의 개념을 설명하고 있다. 먼저 정합창의 크기(이하 WS)를 가장 큰 값(예를 들어 31x31화소)으로 설정한 다음, 오클루전 극복형 수법으로 전탐색을 행하여 제1층의 깊이맵을 구한다. WS를 절반으로(예를 들어, 상위층이 31x31화소였다면 15x15화소) 줄이고, 오클루전 극복형 수법으로 깊이맵을 구한다. 이때, 제2층부터는 전탐색이 아니라 부분탐색을 행한다. 부분탐색의 범위는 상위층의 정합창내에 존재하는 깊이값만으로 구성된 집합으로 제한한다.

이것은 앞서 서술한 BO에 대한 고찰에 근거한 것이다. 제2층부터는 매층 WS를 절반으로 줄이면서 WS가 3x3화소가 될 때까지 동일한 처리를 반복한다.

어떤 화소 근처에 깊이값의 불연속부가 존재하지 않는 경우에는 그 화소에 대해 더 이상 추정 갱신을 하는 의미가



그림 6. 계층적 추정법으로 얻은 깊이맵. 오른쪽에는 에지맵도 함께 나타냄.

Fig 6. Depth map obtained by the hierarchical method.
Edge map is overwritten in the right image.



그림 7. 계층적 추정법에 후처리를 행하여 얻은 깊이맵. 오른쪽에는 에지맵도 함께 나타냄.

Fig 7. Depth map obtained by the hierarchical method with post-processing.
Edge map is overwritten in the right image.

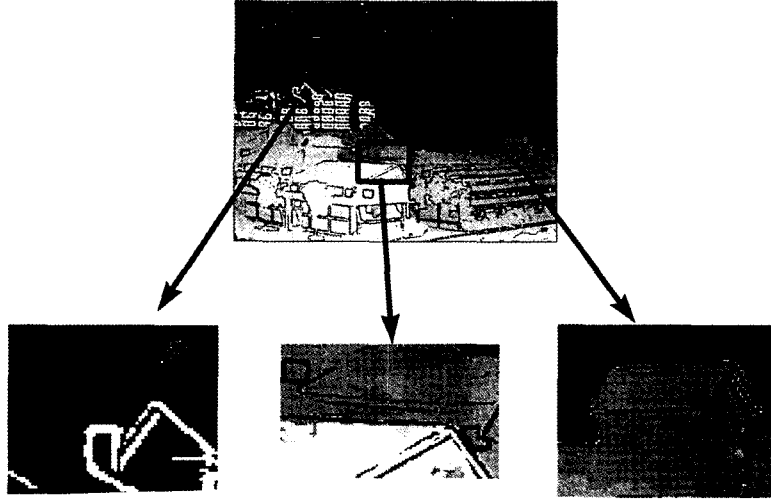


그림 8. 부분확대한 깊이맵.

Fig 8. Magnified depth maps showing how well the proposed method can overcome the occlusion problem and the boundary overreach.

없다. 그 대신에 추정된 깊이값 부근의 SSD함수를 2차 근사하여 화소단위 이하의 정밀한 값을 계산한다. 그 이하의 층에서는 더 이상 추정값을 갱신하지 않는다.

그림 6에 오클루전대책이 있는 계층적 추정법으로 구한 깊이맵을 제시하였다. 이것은 3층처리(15x15 → 7x7 → 3x3화소)를 수행한 결과이다. 깊이맵에서 물체의 경계선이 분명하게 살아있고, 또 그것이 실제의 물체의 경계선과 일치함을 확인할 수 있다. 그러나 물체간의 경계선 부근에서 약간의 추정잡음이 관측된다. 이것은 마지막층(WS=3x3화소)의 추정에서 WS가 너무 작기 때문에 잡음내성이 약화되어 발생한 것이다. 이러한 잡음은 적절한 후처리필터에 의해 제거 가능하다. 이 논문에서는 에지보존성이 뛰어난 적응비선형필터(14)를 사용하여 잡음을 제거하였다. 그림 7이 후처리한 결과이고 그림 8은 이를 부분적으로 확대한 것이다. 잡음이 효과적으로 제거된 정확한 깊이맵이 얻어졌음을 확인할 수 있다. 이 논문에서는 이상과 같은 방법으로 물체의 경계선과 깊이맵의 에지가 일치하는 바람직한 깊이맵을 추출하고 있다.

4 실험결과

4.1 알고리즘의 성능

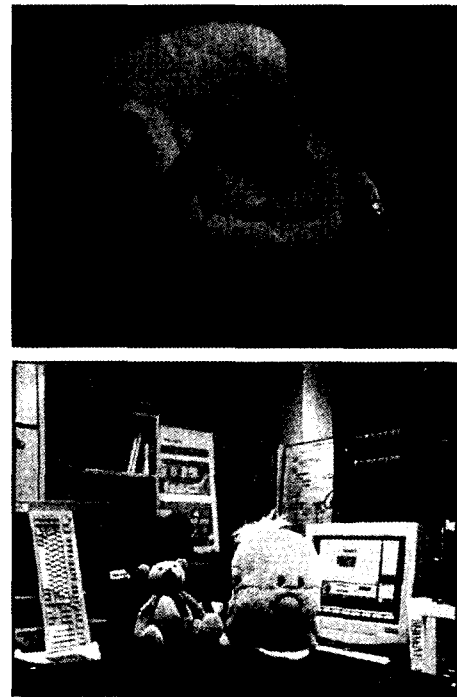


그림 9. 실험에 사용한 영상. "Santa"(위)와 "Lab"(아래).

Fig 9. Images used in the experiments, "Santa"(upper) and "Lab"(lower).

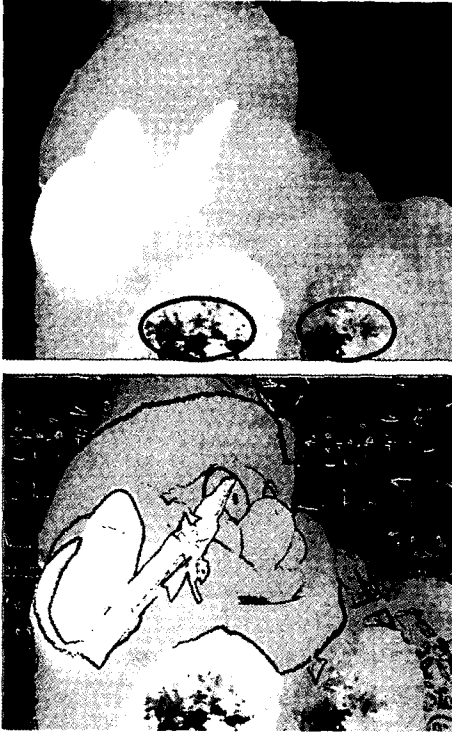


그림 10. "Santa"영상의 깊이맵
Fig 10. Depth map of "Santa" image.
Edge map is overwritten in the lower map

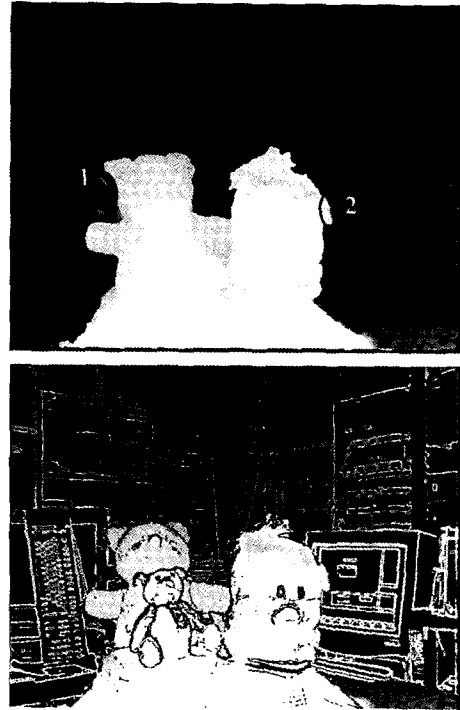


그림 11. "Lab"영상의 깊이맵
Fig 11. Depth map of "Lab" image.
Edge map is overwritten in the lower map

다양한 실험을 통하여 제안알고리즘의 성능을 검토하였다. Tsukuba대학의 다시점 화상데이터베이스 및 실험실에서 촬영한 영상을 사용한 실험결과중의 일부를 제시한다.

그림 9에 실험에 사용한 원영상 중에 기준카메라로부터의 영상을 제시하였다. 해상도 640x480화소의 8비트 그레이 영상이다. 그림 10은 다시점 화상데이터베이스의 "Santa"화상을 사용한 실험결과이다. 제1층의 탐색영역을 0화소에서 50화소로 설정하고, 4계층의 추정(31x31 → 15x15 → 7x7 → 3x3화소)을 행한 것이다. 물체의 경계선의 에지와 깊이맵의 에지가 정확하게 일치하면서도 평탄부가 부드럽게 변화하는 것을 확인할 수 있다. 그러나 텍스처가 전혀 없는 영역, 예를 들어 Santa의 발부분(왼쪽으로 표시)에서는 올바른 정합이 이루어지지 않았다. 이는 영역법의 근본적인 한계라고 할 수 있어, 신호수준의 처리가 아닌 보다 고도의 지식 및 인식에 기반한 처리를 행하기 전에는 해결불가능이라고 볼 수 있다. 그림 11은 SONY 3CCD

DXC-930카메라를 사용하여 실험실에서 촬영한 "Lab"화상의 깊이맵이다. 실험조건은 전과 동일하다. 예리한 경계선을 갖는 정확한 깊이맵이 얻어졌음을 확인할 수 있다. 곰인형의 왼편에 약간의 오차가 보이는데, 이는 복잡한 오를루전(왼쪽으로 표시한 번호 1 부분. 곰인형, 모니터, 책꽂이가 복잡하게 겹치고 있다)의 영향이다. 이 또한 5대의 카메라를 사용한 신호수준의 처리로는 해결하기 어려운 문제라고 생각된다. 병이리인형의 오른쪽 위부분의 오차는 휘도치의 포화로 인한 것(왼쪽으로 표시한 번호 2 부분)으로서 촬영시 조리개를 너무 많이 열어서 발생한 것이므로 피할 수 있는 문제이다.

많은 실험을 통하여 제안방법의 타당성을 확인하였다. 제안방법에서 설정한 가정이 상당히 넓은 범위의 영상에 대해 잘 들어맞았다. 그럼으로써, 우리가 목표로 했던, 물체의 모양을 정확히 나타내는 깊이맵을 얻을 수 있었다.

4.2 응용

제안방법으로 취득한 깊이맵을 사용한 응용례 몇 가지를 소개하고자 한다. 그림 12는 깊이맵을 사용하여 영상을 합성하는, 이른바 Z-key법에 응용한 예이다. 물체의 경계선

이 정확하게 살아있기 때문에 자연스런 합성영상이 얻어짐을 확인할 수 있다. 그림 13은 깊이맵을 사용하여 임의의 시점에서의 영상을 생성한 결과이다. 일단 제작된 영상에 대하여, 시점을 원하는 위치로 마음대로 이동하는 것이 가

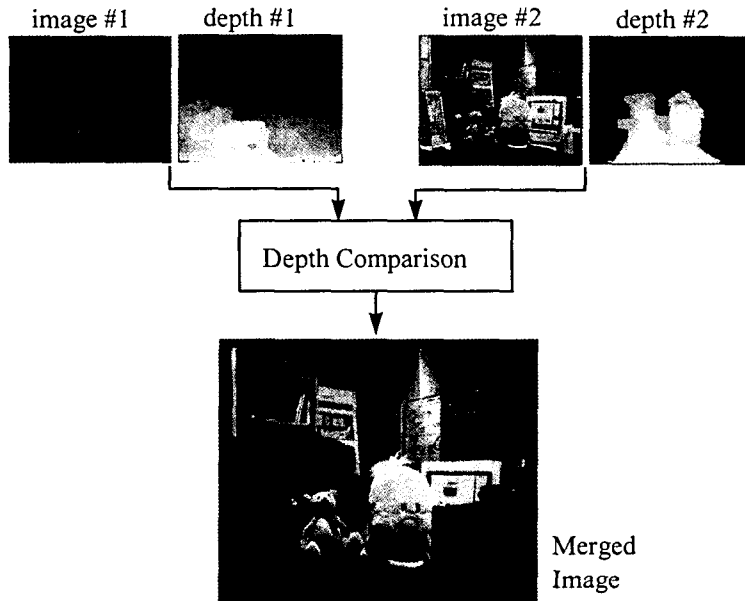


그림 12. Z-key법에 의한 3차원 영상합성의 예
 Fig 12. An example of 3D composition based on Z-key method

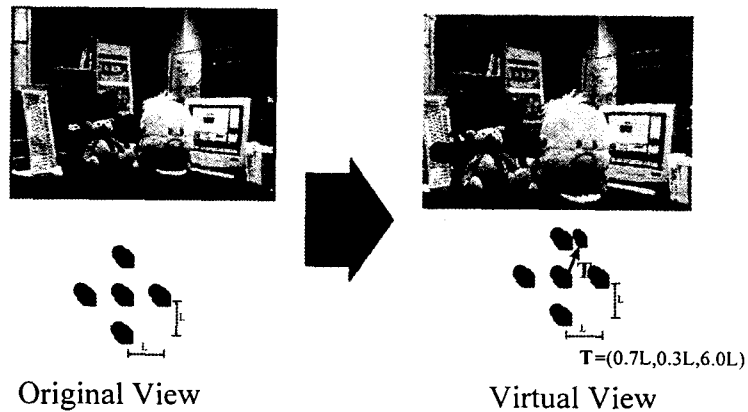


그림 13. 깊이정보를 이용하여 임의시점영상을 제작한 예.
 Fig 13. An example of new view generation using depth information.
 An example of new view generation using depth information.

능하여, 확장된 현실감(augmented reality), 영상제작의 사후제작(post-production) 등의 응용이 기대된다. 제안 방법은 이 밖에도 영상의 자동기술(記述), 3차원 특수효과 등 많은 분야에 이용될 수 있을 것으로 전망된다.

5. 맺음말

이 논문에서는 여러 대의 카메라를 이용하여, 화면내의 모든 화소에 대해 정확한 깊이정보를 추출하는 알고리즘을 제안하고 그 유효성을 확인하였다.

제안방법은, 스테레오정합에 있어서 오클루전(occlusion)이 추정에 미치는 악영향을 극복하기 위해 5대의 카메라를 사용한 후, 각 카메라로부터의 정보를 선택적으로 이용함으로써, 이 문제를 상당히 개선하는 한편, 정합창의 크기에 따른 추정성능의 트레이드오프 문제를 해결하기 위해, 처음에 큰 창으로 추정하고 차츰 창의 크기를 줄여나가는 계층적 방법을 제안하여, 잡음내성을 강화하고 불연속부의 오차를 줄이고 있다.

제안방법은 시간축상의 정보를 이용하고 있지 않으나, 앞으로는 이를 적극적으로 이용하는 다시점 영상으로부터의 구조 및 운동 해석에 대한 연구를 추진할 계획이다. 또, 깊이맵을 활용하여 영상에 대한 보다 높은 차원의 기술(記述)을 자동화하는 방법, 이렇게 추출된 고차원정보를 되먹임해서 보다 정확한 깊이맵을 얻는 데에 활용하는 방법, 색이나 에지정보를 활용하는 방법 등을 검토해 나아갈 예정이다.

참고문헌

- (1) S.Inoue, "Mental image expression by media integration - COMICS," Proc. 1st International Workshop on New Video Media Technology, pp.47-52, Seoul, Korea, March 1994.
- (2) M.Shibata et al., "Scene describing method for video production," ITEJ Technical Report, vol.16, no.10, pp.19-24, Jan. 1992 (in Japanese).
- (3) T.Kanade et al., "Virtualized reality: Concepts and early results," Proc. IEEE Workshop on Representation of Visual Scenes, pp.69-76, June 1995.
- (4) T.Kanade et al., "A stereo machine for video-rate dense depth mapping and its new applications," Proc. IEEE CVPR'96, pp.196-202, San Francisco, June 1996.
- (5) J.Park et al. "Extraction of depth information for scene description and its applications," Proc. ITE'96, pp.112-113, Nagoya, Japan, July 1996 (in Japanese).
- (6) S.D.Cochran and G.Medioni, "3-D surface description from binocular stereo," IEEE Trans. PAMI, vol.14, no.10, pp.981-994, Oct. 1992.
- (7) U.Dhond and J.Aggarwal, "Structure from stereo: A review," IEEE Trans. System, Man, and Cybernetics, vol.19, no.6, pp.1489-1510, Nov./Dec. 1989.
- (8) V.S.Nalwa, A Guided Tour of Computer Vision, Addison-Wesley, 1993.
- (9) R.Tsai, "Multiframe image point matching and 3-D surface reconstruction," IEEE Trans. PAMI, vol.5, no.2, pp.159-174, March 1983.
- (10) M.Okutomi and T.Kanade, "A multiple-baseline stereo," IEEE Trans PAMI, vol.15, no.4, pp.353-363, April 1993.
- (11) M.Okutomi, "CVCV-WG Special Report: Technical review and view in computer vision(X) - Stereo vision -, " IPSJ SIG Notes, Computer Vision and Image Media 102-8, pp.57-64, Nov. 1996 (in Japanese).
- (12) Y.Nakamura et al., "Occlusion-detectable stereo - Occlusion patterns in camera matrix," Proc. IEEE CVPR'96, pp.371-378, San Francisco, June 1996.
- (13) T.Kanade and M.Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," IEEE Trans. PAMI, vol.16, no.9, pp.920-932, Sept. 1994.
- (14) K.H.Yang et al., "Image restoration of noisy images using OS filters with adaptive windows," J. Korean Inst. of Telematics and Electronics, vol.27, no.1, pp.112-119, Jan. 1990.

필자소개



박 종 일

1987 서울대학교 전자공학과 학사
1989 서울대학교 전자공학과 석사
1995 서울대학교 전자공학과 박사
1992. 9 ~ 1994. 3 NHK방송기술연구소 체재연구원
1995. 3 ~ 1996. 3 한국방송개발원 선임연구원
1996. 3 ~ 현 재 ATR지능영상통신연구소 객원연구원
주관심분야 : 영상부호화, 영상처리, 컴퓨터비전



이노우에 세이키

1978 도쿄대학 전기공학과 학사
1980 도쿄대학 전기공학과 석사
1991 도쿄대학 전기공학과 박사
1980. 4 ~ 1995. 3 NHK방송기술연구소 연구원
1995. 3 ~ 현 재 ATR지능영상통신연구소 제3연구실장
주관심분야 영상처리, 영상데이터베이스, 감성신호처리