변형된 Census 변환과 적응적 윈도우를 이용한 다해상도 스테레오 정합

홍석근*, 조석제**
*한국해양대학교 제어계측공학과
**한국해양대학교 IT공학부
e-mail:bassoflord@gmail.com

A Multiresolution Stereo Matching with Modified Census Transform and Adaptive Window

Seok-Keun Hong*, Seok-Je Cho**
*Dept of Control & Instrumentation Engineering, Korea Maritime University
**Division Information Technology Engineering, Korea Maritime University

요 약

본 논문은 스테레오 시각에서 3차원 정보를 얻기 위해 변형된 Census 변환과 적응적 윈도우를 이용한 다해상도 스테레오 영상 정합 방법을 제안하고자 한다. 영역 기반 기법으로 스테레오 영상의 대응점을 찾기 위해 탐색 화소 주변 영역의 기울기를 계산하여 윈도우의 크기 및 모양을 적응적으로 결정하고, 윈도우 영역의 정합 비용 함수는 변형된 Census 변환과 AD 연산 결과의 가중합으로 구한다. 정합 과정에서 계산복잡도를 줄이기 위해 계층적 다해상도 구조를 적용하여 영상 피라미드를 만들고, 정합의 정확성을 향상시키기 위해 정합 영역에 변위 평활성과 같은 제약 조건을 적용하여 변위를 전파하는 방법을 사용한다. 실험을 통해 제안한 방법이 변위 탐색 시간을 감소시킬 뿐만 아니라 정합의 타당성이 보장됨을 확인하고자 한다.

1. 서론

스테레오 시각(Stereo vision)은 한 장면을 담은 스테레오 영상에서 물체의 위치 차이를 계산하여 3차원 정보를 추출하는 과정이다. 스테레오 시각 연구에서 두 영상간의 대응점을 찾아내는 정합 과정(stereo matching)이 가장 어렵다. 그래서 대부분의 스테레오 시각에 대한 연구는 대응성 문제, 즉 정합점을 정확히 찾는데 초점이 맞춰져 있다[1].

스테레오 정합은 크게 두 부류로 나눌 수 있다. 하나는 정해진 크기의 영역에 스테레오 영상의 특성에 따른 제약 조건들을 적용하여 변위를 구하는 지역적 정합 기법(local matching method)이고, 다른 하나는 신경회로망(neural network), MRF(Markov Random Field), 신뢰전파기법(belief propagation) 및 진화 연산 알고리즘(evolutionary computation algorithm)과 같은 최적화 기법을 적용하여 전체 영상의 에너지 함수를 계산하여 최적의 변위 분포를 구하는 전역적 정합 기법(global matching method)이다[1]. 이들 중 전역적 정합 기법이 지역적 정합 기법에 비해 정합의 정확도가 높다고 알려져 있으나, 지역적 정합 기법보다 적용하기 어렵고 계산복잡도가 높아 효율성을 향상시키는 방향으로 연구가 주로 이뤄지고 있다. 반대로 지역적 정합 기법은 비교적 계산시간이 짧고 하드웨어적으로 설계하여 적용하기 쉽지만 윈도우의 크기 및 모양을 정합 영역에 알맞게 적응적으로 결정할 수 있어야 정확도가 보장된다.

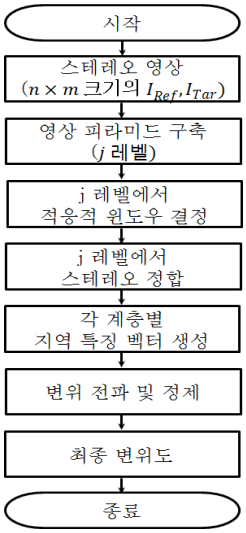
최근 지역적 기법의 성능을 향상시키기 위해 고정된 크기의 영역에 적응적으로 가중치를 두는 방법과 정합 영역 자체의 모양과 크기를 적응적으로 결정하는 방법으로 나뉘어 연구가 진행되고 있다. 적응적으로 가중치를 두는 방법은 각 화소의 칼라와 기하학적인 관계에 의해 결정되는데, 칼라 세그먼트 정보를 이용하는 방법과 정신물리학(psychophisics)에 기반을 둔 칼라 유사도 및 기하학적 인접성과 연속성을 이용하여 가중치를 설정하는 방법이 대표적이다. 윈도우를 적응적으로 결정하는 방법은 정합 윈도우의 크기와 모양을 결정하는데 중점을 둔다. T.Kanade는 가우시안 모델을 이용하여 윈도우 내의 변위 분포를 구하고 이를 이용해 윈도우의 크기와 모양을 결정하는 방법을 제안하였고, O.Veksterㄴ는 영역의 성장(regional growth)개념을 이용하여 윈도우를 결정하는 방법을 제안하였다[2]. A.Fusiello는 한 개의 화소에 9개의 서로 다른 크기의 창을 적용하는 다중 윈도우 알고리즘을 제안하였다. 이러한 기법들은 지역적 기법의 정확도를 전역적 기법만큼 높이는 결과를 가져왔으나 높은 계산복잡도를 요구하기 때문에 지역적 기법의 장점이 퇴색되는 결과를 초래하였다. 따라서 계산복잡도의 증가를 최소화하면서 정확도를 향상시킬 수 있는 방법의 연구가 필요하다. 이는 적절한 윈도우의 종류를 정의하거나 계층적 다해상도 구조를 적용하여 해결할 수 있다[1,3].

본 논문에서는 변위도를 효율적으로 얻기 위한 방법으로 변형된 Census 변환과 적응적 윈도우(adaptive window)를

이용한 스테레오 정합 방법을 제안한다. 대응점을 찾을 화소를 중심으로 하는 영역의 기울기(gradient)를 계산하여 알맞은 크기와 모양의 윈도우를 결정하여 정합비용함수를 구한다. 정합 과정에서 증가하는 계산 복잡도를 줄이기 위해 다해상도 기법을 적용하여 스테레오 영상을 정해진 단계까지 축소하여 영상 피라미드를 만든다. 이 때 근사영상, 예측차 영상 및 에지 영상에 대한 계층구조를 만든다. 따라서 최상위 계층에서 적응적 윈도우를 이용하여 정합을 하고, 정합비용함수는 기존의 SAD(Sum of Absolute Differences)나 NCC(Normalized cross-correlation) 대신 변형된 Census 변환과 AD연산의 가중합을 이용해 구한다. 이러한 방법으로 최상위 계층에서 초기 변위도(initial disparity map)를 얻고, 각 계층별 변위도는 영상의 에지 정보와 변위 평활성(disparity smoothness)과 같은 제약조건(constraint)을 이용하여 전파(propagation)하여 얻는다. 실험을 통하여 제안한 방법의 성능을 확인하였다.

2. 제안한 방법의 개요

제안한 방법의 블록 다이어그램은 그림 1과 같다. 처음에 주어진 스테레오 영상으로 피라미드 구조를 만든다. 원영상 크기를 레벨 0로 두고 단계별로 2배씩 축소하여 j+1 계층의 피라미드를 만든다. 이 때 기준영상의 각 단계별 근사영상에서 에지 영상을 생성한다. 이는 원래 크기로 복원하는 과정에서 각 단계별로 복원될 변위도의 정확한 에지 위치를 보장하기 위함이다. 레벨 j의 에지 영상을 이용하여 각 화소에서의 윈도우 크기 및 모양을 결정하고, 스테레오 근사 영상에 적응적 윈도우를 적용하여 초기 변위도를 얻는다. 그리고 이 변위도를 단계별로 확대하며 전파 및 보간을 수행한다. 이 때 각 영역에 변위 제약 조건을 적용하여 변위를 정확히 전파하고 각 단계별로 얻어진 에지의 위치 정보를 참고하여 해당 단계의 변위도를 얻는다. 이러한 과정을 레벨 0까지 수행하여 원영상 크기의 최종 변위도를 얻는다.



(그림 1) 제안한 방법의 블록 다이어그램

3. 적응적 윈도우를 이용한 스테레오 정합

정합에 사용할 윈도우의 형태는 화소 주변의 질감에 따라 결정된다. 질감의 정도는 화소의 기울기를 계산하여 측정할 수 있고, 기울기의 방향을 계산하여 윈도우의 확장 방향을 결정할 수 있다. 따라서 본 논문에서는 해당 화소의 수직, 수평 기울기를 계산하여 적응적으로 윈도우의 크기를 결정한다.

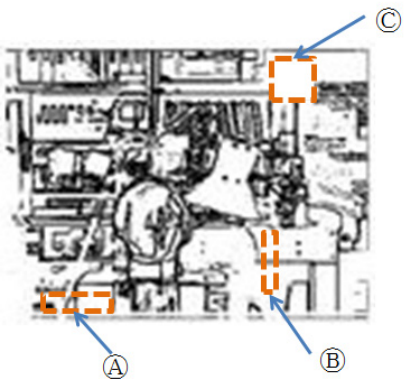
영상의 각 화소에 대한 기울기값을 계산하여 윈도우의 모양과 크기를 결정하는 원리는 그림 2와 같다. 수직방향의 기울기값이 두드러지는 화소는 Ⓐ와 같이 수평방향으로 긴 윈도우를 적용하고, 수평방향으로 기울기값이 두드러지는 화소는 Ⓑ와 같이 수직방향으로 긴 위도우를 적용한다. 마지막으로 질감이 밋밋한 영역은 Ⓒ와 같이 적당히 큰 정사각형의 윈도우를 적용한다.

이와 같이 윈도우를 적용하기 위해서는 수직, 수평방향 각각의 기울기값을 구하고 두 기울기값의 절대치를 구해야한다. 기준영상 $I_{Ref}(x,y)$의 한 화소 (x_p,y_p)에서의 기울기값에 따라 윈도우의 크기를 결정하는 규칙은 다음과 같다.

1. $\nabla f(x_p,y_p) > \alpha$ 이고 $|G_x(x_p,y_p)| - |G_y(x_p,y_p)| > \beta$ 일 때, 윈도우의 크기를 $n_1 \times m_1$로 결정한다.
2. $\nabla f(x_p,y_p) > \alpha$ 이고 $|G_y(x_p,y_p)| - |G_x(x_p,y_p)| > \beta$ 일 때, 윈도우의 크기를 $n_2 \times m_2$로 결정한다.
3. $\nabla f(x_p,y_p) \le \alpha$ 일 때, 윈도우의 크기를 $n_3 \times m_3$로 결정한다.
4. $\nabla f(x_p,y_p) > \alpha$이고 $|G_x(x_p,y_p) - G_y(x_p,y_p)| \le \beta$ 일 때, 윈도우의 크기를 $n_4 \times m_4$로 결정한다.

여기서, $G_x(x_p,y_p)$와 $G_y(x_p,y_p)$는 각각 화소 (x_p,y_p)에서의 수평, 수직 기울기값이고, $\nabla f(x_p,y_p)$는 식 (1)과 같이 수평, 수직 기울기값의 절대값을 조합하여 얻은 근사치이다.

$$\nabla f(x_p,y_p) \approx |G_x(x_p,y_p)| + |G_y(x_p,y_p)| \qquad (1)$$



(그림 2) 기울기 특성에 따른 윈도우의 적용의 예

4. 변형된 Census 변환을 이용한 정합비용함수

제안한 방법은 일반적인 AD-Census 비용척도에서 Census 변환을 변형한 방법을 사용한다. AD-Census 비용척도는 Census 변환에 의한 비용항과 AD(absolute difference)에 의한 비용항의 가중합으로 표현된다[4]. Census변환은 잡음과 같은 요인에 비교적 강인하고 변위가 급격히 변하는 영역에서도 성능이 우수하다. 하지만 중심화소값의 역할이 절대적이기 때문에 중심화소에 큰 잡음이 영향을 미치면 오정합 발생빈도가 증가하고 중심화소와 주변화소의 명암도가 같을 경우 명확히 변환할 수 없다는 문제점을 가지고 있다. 이러한 문제점은 Census 변환시 중심화소와 주변화소들과의 평균값을 계산하여 중심화소값 대신 비교할 때 사용하면 개선시킬 수 있다. 하지만 평균값을 취해도 주변화소값이 평균과 동일해지는 경우가 있을 수 있기 때문에 작은 offset을 설정하여 평균값에 더한 후 비교하는 방법을 사용한다. $n\times m$ 영역에 대해, 중심화소 (x_p,y_p)에서의 변형된 Census 변환 $C_c(x_p,y_p)$은 식 (2)와 같다.

$$C_c(x_p,y_p)=\bigotimes_{i=n}\bigotimes_{j=m}(\xi(\bar{I}(x_p,y_p)+\Delta,I(x_p+i,y_p+j))) \quad (2)$$

여기서, $\bar{I}(x_p,y_p)$는 영역의 평균값이고 Δ는 offset값이다. 그리고 비교함수 $\xi(\alpha,\beta)$는 식 (3)과 같이 정의된다.

$$\xi(\alpha,\beta)=\begin{cases}0,\alpha\le\beta\\1,otherwise\end{cases} \quad (3)$$

선택된 윈도우가 3×3일 때, 중심 화소의 Census 변환 비용을 계산하기 위해 영역의 평균값과 주변 화소값의 크기를 비교하여 Census 변환을 수행한다. 평균값이 더 크면 해당 화소의 위치에 0을 부여하고, 반대의 경우엔 1을 부여한다. 그 비트들을 순서대로 나열한 비트열이 중심화소의 변형된 Census 변환이다. 이러한 과정을 스테레오 영상에 적용하여 비트열을 얻고, 두 비트열의 해밍거리를 계산하면 변형된 Census 변환에 의한 정합 비용을 얻을 수 있다. 이러한 방법은 잡음과 변위 불연속 영역에서 강인하다는 장점이 있지만 화소들의 2진 비트를 만드는 과정에서 화소 정보를 잃게되고, 성능을 높이기 위해 창의 크기를 증가시키면 비트열의 길이가 급격히 늘어나는 단점이 있다. 이러한 단점은 AD 연산을 조합하여 개선할 수 있다.

기준영상의 화소 $p=(x_p,y_p)$에서 변위가 d일 때, AD-Census 비용은 다음과 같이 얻을 수 있다.

$$C(p,d)=\rho(C_c(p,d),\lambda_c)+\rho(C_{AD}(p,d),\lambda_{AD}) \quad (3)$$

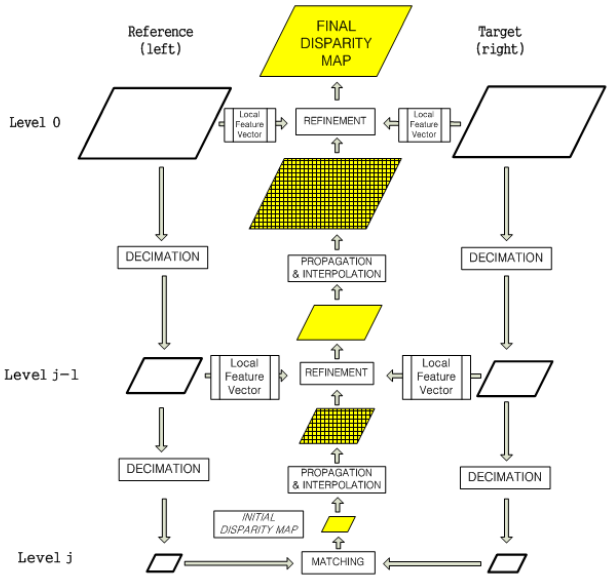
여기서 $C_c(p,d)$는 Census 변환에 의한 정합비용이고 $C_{AD}(p,d)$는 AD 연산에 의한 스테레오 영상의 정합비용이고 $C_{AD}(p,d)=I_L(p)-I_R(p,d)$와 같다. 그리고 $\rho(c,\lambda)$는 $\rho(c,\lambda)=1-e^{(-\frac{c}{\lambda})}$과 같다. 여기서 λ는 ρ의 동적범위를 [0,1]로 해주는 파라미터이다.

5. 영상 피라미드 구조와 변위 전파

본 논문은 영상 피라미드 기법을 이용한 효율적인 스테레오 정합 알고리즘을 제안한다. 제안한 다해상도 기반의 스테레오 정합을 그림 3에 나타내었다. 레벨 0의 스테레오 영상에 피라미드 기법을 적용하여 각각 레벨 j 단계의 피라미드 영상들을 얻는다. 레벨 j의 스테레오 영상을 정합하고 다음 단계인 레벨 j-1에 해당하는 크기로 확대하면서 지역적 특징 정보를 이용하여 이전 단계 변위를 정확한 위치로 전파하고 주변 화소들의 변위를 동시에 구한다. 그리고 레벨 j-1 단계의 스테레오 영상과 에지 영상을 이용하여 변위를 정제함으로써 현 단계 변위도의 정확도를 높인다. 이러한 일련의 과정을 거쳐 원래 크기인 레벨 0의 최종 변위도를 얻는다.

계층적 접근을 통해 스테레오 영상 각각에 대한 근사 영상 피라미드와 예측차 영상 피라미드를 얻을 수 있다. 레벨 j-1의 영상을 근사화 필터(approximation filter)처리하여 2배 축소하면 레벨 j의 근사 영상을 얻을 수 있다. 2배 확대한 영상과 레벨 j-1의 예측차 영상을 이용하여 레벨 j-1의 영상을 복원할 수 있다. 이런 과정을 반복하여 레벨 0의 원영상을 복원할 수 있다.

제안한 방법은 가우시안 저역통과 필터를 이용하는 가우시안 피라미드 기법을 이용하여 계층화된 영상들을 얻고 동시에 각 단계별 근사영상에 대한 에지 영상을 얻는다. 레벨 j까지 계층화가 이뤄지면 최상단 근사 영상에 대한 스테레오 정합을 실시하여 최초 변위도(initial disparity map)을 얻는다.



(그림 3) 제안한 다해상도 기반 스테레오 정합

레벨 j의 최초 변위도를 얻은 후 변위도를 단계적으로 확대하며 변위를 전파한다. 최초 영상에서 화소 (x,y)가 2배 확대되면 이 화소는 2배 확대된 영상, 즉 레벨 j-1의 $(2x,2y)$, $(2x+1,2y)$, $(2x,2y+1)$, $(2x+1,2y+1)$에 해당한다. 그리고 해상도가 두배씩 증가함에 따라 대응점의 상대거리도 2배 가까이 증가할 수 있다. 먼저 레벨 j의 변위는 레벨 j-1의 4개의 화소에 저장된다. 정확한 변위를 전파하기 위해 그림 4와 같은 Haar 웨이블렛 필터를 이용한 지역적 특징 정보를 사용한다. 레벨 j-1 스테레오 영상에 $2^2 \times 2^2$ 크기의 Haar 웨이블렛 필터를 이용하여 4차원의 특징 벡터 V를 식(3)과 같이 구성한다. 이때 필터는 단계마다 승수를 1씩 증가시켜 레벨 0에서 $2^{j+1} \times 2^{j+1}$의 크기로 처리한다.

$$V=[\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|] \qquad (4)$$

여기서, d_x, d_y는 수평방향과 수직방향 Haar 웨이블렛 필터로 얻은 응답을 나타낸다. 각 위치에 전파된 변위를 이용하여 식 (4)과 같은 특징차 $Dist(x,y)$를 구한다.

$$Dist(x,y)=|V_R(x,y)-(V_T(x-d_i(x,y),y)| \qquad (5)$$

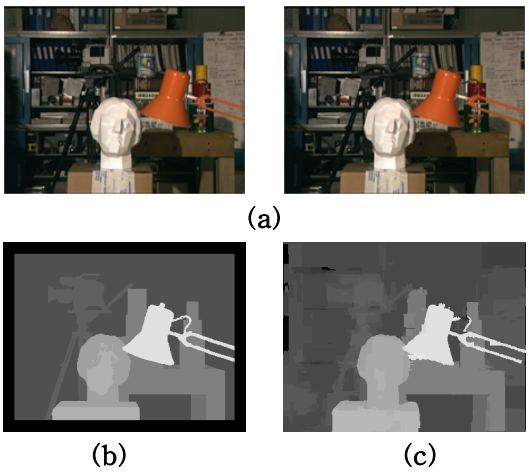
여기서, $V_{Ref}(x,y)$와 $V_{Tar}(x,y)$는 각각 기준 영상과 대상 영상의 좌표 (x,y)의 Haar 웨이블렛 특징 벡터이고 $d_i(x,y)$는 변위이다. 이 때 변위 $d_i(x,y)$는 레벨 j에서 전파된 변위 $d_p(x,y)$부터 $(2d_p(x,y)+1)$까지 순차적으로 1씩 증가시켜 특징차를 가장 작게 하는 변위를 해당 좌표의 변위로 결정한다.

변위도의 정확도를 위해 정제(refinement)과정을 거쳐야한다. 현 단계의 예측차 영상을 이용하여 복원한 스테레오 영상과 에지 영상을 이용한다. 레벨 j-1의 보간된 변위도는 에지 화소 사이 변위들을 차례로 적용하여 SAD가 작은 변위가 그 영역의 변위가 된다.

5. 실험 결과

제안한 방법의 성능을 검증하기 위해 Tsukuba 영상으로 실험했다. 실험에 이용한 Tsukuba 좌, 우 영상은 그림 4의 (a)와 같다. 사실 변위도는 그림 4의 (b)이다. 제안한 방법으로 얻은 변위도는 그림 4의 (c)이다. 적응적 윈도우의 크기는 각각 3×9, 9×3, 9×9, 3×3으로 설정하였다.

제안한 방법으로 정확한 변위도를 얻을 수 있음을 확인할 수 있다. 에지를 활용하여 정합을 시도하기 때문에 경계선을 명확히 보전하면서 조밀한 변위를 얻을 수 있었다. 이는 변위를 전파하고 정제하는 과정에서 변위의 제약조건이 한번 더 적용되는 것이 변위의 오차를 줄이는데 효과가 있음을 의미한다. 비폐색영역과 불연속경계 근처에서의 오정합화소 비율(%)은 각각 3.93과 10.62로 나타났다. 계산시간은 0.5(sec) 소요되었다



(그림 4) (a) Tsukuba 영상
(b) Tsukuba 사실 변위도
(c) 제안한 방법으로 얻은 변위도

6. 결론

본 논문은 변형된 Census 변환과 적응적 윈도우를 이용한 다해상도 스테레오 정합을 제안하였다. 영상 피라미드 기법을 이용하여 저해상도 영상과 에지 정보를 이용하여 적응적으로 정합 윈도우의 크기와 형태를 정의하였고, 변형된 Census 변환으로 정합비용함수를 설계하여 정합 성능을 향상시킬 수 있었다. 최초 변위도를 얻고 영상을 단계적으로 확대하며 변위를 전파할 때 지역적 특징 정보를 이용하여 전파위치를 찾고 주변 변위를 추정하였다. 실험을 통해 제안한 방법이 변위도를 계산하는데 좋은 성능을 보임을 확인할 수 있었고, 다해상도 기법을 결합함으로써 적응적 윈도우 방법의 계산 복잡도를 감소시킬 수 있음을 알 수 있었다.

후 기

본 연구는 한국해양대학교의 지원을 받아 수행된 2012년도 대외경쟁력 강화를 위한 연구강화지원사업의 연구결과입니다.

참고문헌

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," International Journal of Computer Vision, pp. 7-42, 2002.

[2] 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, 1994.

[3] B. Wang, J. Wang, Y. He, and C. Shen. "A novel stereo matching algorithm," Computer Enginnering, Vol.31. pp. 24-26, May 2005.

[4] X. Sun, X. Mei, Jiao, M. Zhou, and H. Wang. "Stereo Matching with Reliable Disparity Propagation," 3DIMPVT 2011.