

간소화된 윈도우 푸리에 위상을 이용한 계층적 접근기반의 3차원 객체 추출 기법

정희원 민 각*, 한 규 필**, 이 기 수**, 하 영 호*

3D Object Extraction Algorithm Based on Hierarchical Approach Using Reduced Windowed Fourier Phase

Gak Min*, Kyu-Phil Han**, Ky-Soo Lee**, Yeong-Ho Ha* *Regular Members*

요 약

본 논문에서는 두 개의 2차원 영상에서 3차원 객체를 효율적으로 추출하기 위해서 위상 기반의 스테레오 정합 알고리즘을 제시한다. 특히, 윈도우 푸리에 위상을 이용하는 기존의 위상 기반 방법들은 기본적으로 다중-해상도 위상 맵을 사용하기 때문에 계층적인 접근 관점에서 좋은 특성을 가지고 있는 반면 높은 계산량을 요구한다. 그러므로 본 논문에서는 다중-해상도 위상 기반전략과 더불어 위상 계산의 중복성을 제거하는 빠른 계층적 접근기반의 3차원 객체 추출 기법을 제안한다. 또한, 정합 성능을 개선시키기 위해 위상 영역에서 형태학적인 정합 알고리즘을 제시한다. 제안된 알고리즘을 실험한 결과 계산량이 대략 8배 정도로 크게 감소되었으며 안정된 결과 값을 획득할 수 있었다.

ABSTRACT

This paper presents a phase-based stereo matching algorithm in order to efficiently extract 3-dimensional objects from two 2D images. Conventional phase-based methods, especially using windowed Fourier phases, inherit good properties in the case of hierarchical approaches, because they basically use a multi-resolution phase map. On the contrary, their computational costs are very heavy. Therefore, a fast hierarchical approach, using multi-resolution phase-based strategy and reducing the redundancy of phase calculations, is proposed in this paper. In addition, a structural matching algorithm on the phase domain is adopted to improve the matching quality. In experimental results, it is shown that the computation loads are considerably reduced about 8 times and stable outputs are obtained.

1. 서 론

최근 컴퓨터 및 인터넷 환경에서 멀티미디어 데이터의 사용이 급격히 증가하고 있다. 멀티미디어는 문자, 음성, 영상의 각 모노미디어들이 복합된 것으로 현장감을 전달할 수 있다는 장점 때문에 사용자들이 선호하고 있다. 특히, 가상현실의 도래와 더불어 이런 멀티미디어 데이터에 현실감을 추가하는

연구가 여러 분야에서 이루어지고 있다. 지금까지 3차원 정보의 응용은 물체인식 및 검사, 자동항법장치 등과 같은 기계시각(machine vision) 분야^{1,2)}에서 이용되어져 왔지만 최근 많은 컴퓨터 통신환경에서 입체 영상 전송을 위해 영상열 간의 공간적 중복성을 제거하거나 인터넷에서 3차원 객체를 실시간으로 전송하는 분야에 활용되고 있다.^{3,4)}

3차원 깊이정보를 효율적으로 추출하는 가장 대표적인 방법은 스테레오 시각(stereo vision)을 이용

* 경북대학교 전자전기공학부(yha@ee.knu.ac.kr),

논문번호 : 020089-0222, 접수일자 : 2002년 2월 22일

※ 본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문입니다.

** 금오공과대학교 컴퓨터공학부(kphan@kumoh.ac.kr)

하는 것이다.^[2] 스테레오 시각 기법은 인간의 시각 시스템을 모방한 것으로 서로 다른 관점(view point)에서 얻어진 두 장의 영상으로부터 3차원 정보를 복원하는 수동 거리측정 방법의 하나이다.^[11] 스테레오 시각에서의 핵심은 좌·우 영상 사이의 대응점(corresponding point)간의 거리, 즉 변이(disparity)를 찾아내는 정합(matching) 과정이다. 기존의 정합 방법은 영상에서 특징을 추출하여 정합하는 특징 기반법(feature-based method)^[5,6]과 밝기 값을 그대로 사용하는 휘도 기반법(intensity-based method),^[7] 주파수변환에서 추출된 위상을 이용하는 위상 기반법(phase-based method)^[8] 등이 대표적이다. 특징 및 휘도 기반 방법들은 국부적인 공간상에서 주변 값의 정보만을 이용하기 때문에 오정합 가능성이 높다. 그러므로 이 방법들은 성능을 높이기 위해서 계층적인 접근기법을 도입하여 적용되어진다.^[7] 위상 기반 방법은 기본적으로 계층적 구조를 사용하기 때문에 보다 나은 결과를 얻을 수 있다는 장점이 있지만 주파수 변환과정과 계층적 구조를 획득하기 위해 과도한 계산량을 요구한다.^[7]

본 논문에서는 정합 성능을 높이기 위해 계층적 접근 기반의 위상 정합 기법을 사용하면서 기존 위상 기반의 주파수 변환에서 요구되는 과도한 계산의 중복성을 제거하여 계산량을 줄이고 정합 성능을 높이는 위상 기반의 정합 방법을 제시한다. 또한, 이렇게 추출된 3차원 깊이 정보를 컴퓨터 및 인터넷 환경에서 활용할 수 있게 VRML(virtual reality markup language)을 이용하여 3차원 메쉬로 표현한다. 512×512 영상에 대한 실험에서 제안한 방법은 기존의 위상기법에 비해 계산량을 1114배, 수행시간을 8배정도 줄일 수 있었으며 성능 면에서도 우수한 결과를 얻을 수 있었다. 그리고 제안한 방법이 3차원 모델을 그대로 출력하기 때문에 컴퓨터 및 인터넷 환경에서 추출된 모델에 대한 별도의 처리과정이 없이 이용할 수 있다는 장점이 있었다.

II. 푸리에 위상과 윈도우 푸리에 위상(WFP)

1. 푸리에 위상과 전역이동(global shift)

무한의 연속 신호, $f(x)$,와 이 신호의 푸리에 변환 신호를 $F(u)$, 그리고 이동된 신호, $f_s(x)$,가 있다고 가정하면 주파수 영역에서의 이동 신호는

$$F_s(u) = \int_{-\infty}^{\infty} f_s(x) e^{-j2\pi ux} dx = e^{-j2\pi us} F(u) \quad (1)$$

로 표현된다. 식(1)에서 단순이동에 대한 관계를 유도하기 곤란하므로 고정된 특정 주파수, v ,에 대한 푸리에커널, $e^{j2\pi vx}$,을 공간영역에서의 상승적분식을 적용하면 식 (2)로 표현된다.

$$\begin{aligned} a(x) &= e^{j2\pi vx} * f(x) \\ &= \int_{-\infty}^{\infty} f(\tau) e^{j2\pi v(x-\tau)} d\tau \\ &= e^{j2\pi vx} F(u) = |a(x)| e^{\arg[a(x)]} \end{aligned} \quad (2)$$

여기서 $\arg[a(x)] \equiv 2\pi vx + \arg[F(u)]$ 로 위상부분이 처리된다. 이와 같은 방법으로 이동된 신호에 대해서 적용하면 다음과 같이 표현된다.

$$\begin{aligned} a_s(x) &= e^{j2\pi vx} * f_s(x) \\ &= e^{j2\pi v(x-s)} F(u) \\ &= |a_s(x)| e^{\arg[a_s(x)]} \end{aligned} \quad (3)$$

여기서 $\arg[a_s(x)] \equiv 2\pi v(x-s) + \arg[F(u)]$ 이다. 원 신호와 이동된 신호간의 위상차는 식(2)와 식(3) 사이의 위상 정보 차와 같다. 그러므로 위상차는

$$\arg[a_s(x)] - \arg[a(x)] \equiv 2\pi vs \quad (4)$$

로 표현되고 여기서 이동된 변이, s ,는

$$s = \frac{\arg[a_s(x)] - \arg[a(x)]}{2\pi v} \quad (5)$$

로 계산될 수 있다. 식(5)는 임의의 $f(x)$ 에 포함된 특정 주파수, v ,가 전체 파형에서 얼마만큼 이동했는가를 전역이동 정보를 포함한다. 그러므로 푸리에 변환에서 각 주파수 성분이 가지는 위상 값은 신호정합에 사용되는 국부적인 파형의 이동정보, 즉 신호 일부분의 이동된 값을 표현함에는 부족하다.^[7]

2. WFP와 지역이동(local shift)

신호의 지역이동 정보를 처리하기 위해서는 전체 신호의 일부를 표본화하는 윈도우 함수가 필요하다. 그림 1은 윈도우 함수로 표본화된 지역신호가 전체 신호로 처리되어 각 파형의 이동 정보를 계산함으로써 WFP 값이 구해지는 과정을 보여준다.

윈도우 크기가 M 인 윈도우 함수를

$$W_M(x) = \begin{cases} 1 & \text{if } |x| \leq M/2 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

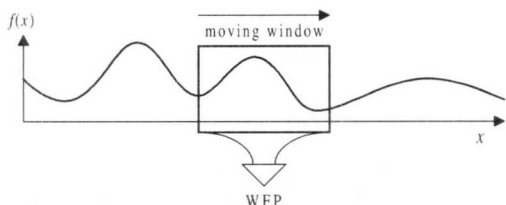


그림 1. 윈도우 중심에서 WFP가 계산되는 영역

라고 하면 윈도우 푸리에 커널(windowed Fourier kernel)은

$$h(x) = W_M(x)e^{j2\pi vx} \quad (7)$$

식 (7)처럼 정의된다. 그러므로 윈도우의 크기, M 이 곧 파형의 주기가 되고 공간주파수는 $1/M$ 이 된다.

그림 2는 “pentagon” 실영상 100번째 줄 정보를 윈도우의 크기 M , 즉 신호 주기를 5에서 32화소까지 변화시키면서 계산된 WFP이다.

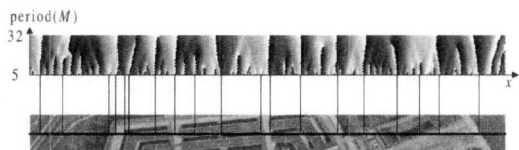


그림 2. 실제 영상에서 계산된 WFP의 예($5 \leq M \leq 32$).

III. 제안한 3차원 객체추출 기법

3차원 객체추출을 위해 제안한 알고리즘의 전체 블록도는 그림 3과 같다.

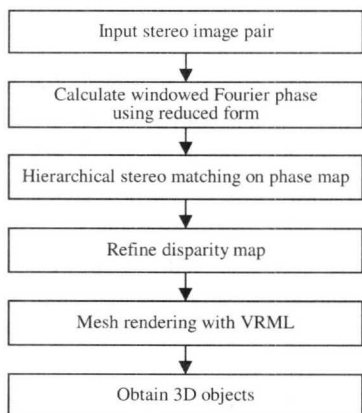


그림 3. 제안한 알고리즘의 전체 블록도

1. 위상 계산의 중복성 제거

일반적으로 푸리에 변환은 여현(cos) 커널에 의한 실수 성분과 정현(sin) 커널에 의한 허수 성분으로 구분되고 각 성분은 tan 함수의 역함수로서 위상이 계산되어진다. 계산 속도가 개선된 고속 푸리에 변환을 WFP 계산에 적용하면 윈도우의 크기가 2의 지수승 형태로 증가되기 때문에 계산결과 신호의 불연속성이라는 단점을 가지게 된다.

이러한 문제점을 해결하기 위한 방법으로 본 논문에서는 윈도우의 크기를 정수 단위로 증가시키면서 정현 커널과 여현 커널의 중복되는 성질을 이용하여 불필요한 삼각함수의 연산을 제거함으로써 계산 속도를 개선하였다.

표 1과 2에서 윈도우의 크기가 5와 6인 경우 10번 및 12번의 커널 계산이 요구되며, 윈도우의 크기가 가변될 경우 새로운 커널값을 계산하여야 한다.

표 1. 윈도우의 크기가 5일 때 각 커널의 값($v=1/M$)

공간변수(x) kernel	0	1	2	3	4
cos(2πx/M)	1	0.309	-0.809	-0.809	0.309
sin(2πx/M)	0	0.951	0.587	-0.587	-0.951

표 2. 윈도우의 크기가 6일 때 각 커널의 값($v=1/M$)

공간변수(x) kernel	0	1	2	3	4	5
cos(2πx/M)	1	0.500	-0.500	-1	-0.500	0.500
sin(2πx/M)	0	0.866	0.866	0	-0.866	-0.866

예를 들면, 512 샘플의 1차원 신호가 윈도우 크기가 5에서 32까지 변화될 경우, 기존 WFP 계산량은 $O(WFP) = 512 \times 2 \times (5 + 6 + 7 + \dots + 32) = 265,216$ 번 연산이 수행된다.

$$\cos(x) = \cos(2\pi - x), \sin(x) = -\sin(2\pi - x) \quad (8)$$

식 (8)은 커널값의 중복성을 나타내며, WFP연산시 각 커널을 메모리 참조하는 간소화된 위상계산의 개략적인 비용은 $O(RWFP) = 2 \times 1/2 \times (5 + 6 + 7 + \dots + 32) = 518$ 번 연산이 수행된다. 여기서 1/2은 각 윈도우의 크기에 절반 정도의 계산이 필요하다는 것이다. 삼각함수의 연산이 필요하지 않는 경우를 고려하면 윈도우의 크기가 M 인 경우 정확한 삼각함수의 계산 횟수, C_M 은

$$C_N = \text{int}\left(\frac{M-1}{2}\right) \quad (9)$$

처럼 표현된다. 따라서, 간소화된 위상계산의 정확한 비용은 $O(RWFP) = 2 \times 1/2 \times (2 + 2 + 3 + 3 + 4 + 4 + \dots + 15 + 15) = 238$ 번 연산이 수행된다.

2. 제안한 계층적 위상기반 정합

위상 정보의 정점(peak)과 영교차점(zero-crossing point)은 밝기 신호의 정점과 골(valley)에 대응되기 때문에 밝기 변화를 그대로 간직하고 있어서 정합에 적합하다.^[8] 그림 4는 원 신호와 이동된 신호의 WFP를 각각 보여준다. 그림 5는 그림 4에서 각각의 위상 특징을 구한 것을 나타낸다. 특징 선에서 검은 부분이 정점들이고 회색선은 영교차점이다.

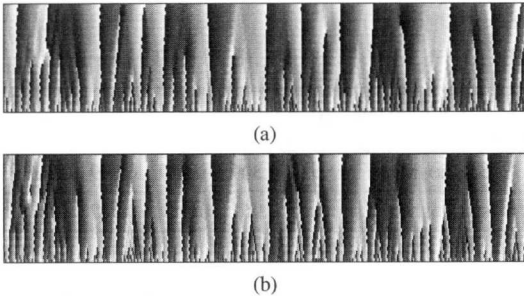


그림 4. 원 신호와 이동된 신호의 WTF. (a)원 신호, (b)이동된 신호

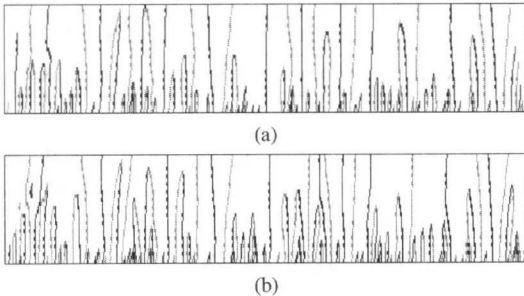


그림 5. 그림 4의 위상 맵에서 추출된 특징 영상. (a)원 신호, (b)이동된 신호

위상 맵은 해상도 방향으로 양호한 계층적인 특성을 가지고 있다. 그러므로 본 논문에서는 강한 밝기 변화를 대표하는 저해상도의 위상 특징에서 정합을 먼저 수행하고, 여기서 정합된 값을 고정시키고 해상도를 높이면 정합된 특징들 사이에 정합되어야 할 새로운 특징들이 발생된다. 이 특징들을 이미 정합된 점들 사이에 존재하기 때문에 탐색 영역 및 오정합 가능성이 감소되고, 순서 제약(ordering

constraint) 조건을 만족하게 된다. 제안한 정합 과정은 1) 고립 특징(isolated feature) 제거와 2) 정합 후보(candidate) 선출 단계를 거쳐서 수행된다.

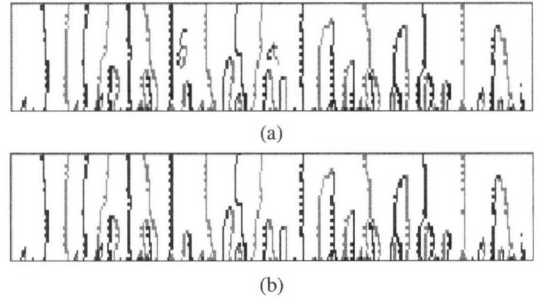


그림 6. 원 특징영상과 비 연결 특징을 제거한 영상. (a)원 특징 영상 (b)비 연결특징을 제거한 영상

1) 고립 특징 제거

WFP 신호에서 특징을 추출하면 그림 5(b)의 좌측부분이나 그림 6(a)의 가운데 부분에 고립된 특징이 존재하며, 이런 점들은 정합과정에서 오류를 유발시키므로 제거된다. 고립 특징을 제거하기 위해 본 논문에서는 space-efficient 2-pass labelling 알고리즘을 사용하였다.^[9] 저해상도의 특징 시작점에서 고해상도로 레이블링을 수행하고 그 끝점이 최고 해상도까지 도달하지 않는 레이블은 제거하였다.

2) 정합후보 선출 및 정합

해상도가 낮아질수록 위상값을 계산하는 윈도우의 크기가 증가하는 특성이 있다. 이는 특징값의 변화를 발생시키는 요인이 되므로 정합후보 선출은 해상도에 따라서 탐색영역의 크기를 변화시키면서 추출한다. 그래서 해상도 R단계에서 탐색영역, S_R 은

$$S_R = s_r + M \quad (10)$$

로 표현되고, 여기서 s_r 은 원 영상의 기본적인 탐색영역이고 M은 윈도우 크기를 나타낸다. 그래서, 해상도 변화, 즉 위상 계산에 사용되는 윈도우의 크기에 따라서 정합 후보가 이동되어지는 편차를 고려할 수 있다. 예를 들면, 기준(source) 측의 정점을 정합할 때는 식 (10)를 이용하여 탐색영역에 존재하는 모든 목표(target) 측의 정점들이 정합 후보로 선출되고 영교차점을 정합할 때는 목표 측에서의 영교차점들이 선출된다.

식 (11)은 여러 개의 정합 후보들로부터 최상의 정합점을 찾기 위한 거리 척도(distance measure)이며, 위상 및 휘도 값을 모두 고려되었다.

$$D(s_i, t_j) = \min_{all j} (D_f(s_i, t_j) + D_i(s_i, t_j)) \quad (11)$$

여기서 D_f 와 D_i 는 특징 공간에서의 오차와 그 특징이 추출된 밝기 공간에서의 오차를 각각 나타낸다. 그러므로 j 개의 모든 정합 후보들 중에서 특징 및 밝기 오차 값의 합이 최소가 되는 특징을 정합점으로 선택한다. 특징 공간에서의 오차는

$$D_f(s_i, t_j) = \frac{d}{D_{max}} (L(s_i) - L(t_j)) \quad (12)$$

과 같이 표현된다. 여기서 d 는 특징 s_i 와 t_j 사이의 거리, D_{max} 는 탐색 범위의 최대 값, 그리고 $L(\cdot)$ 은 특징의 길이를 산출하는 함수이다. 식(12)의 의미는 비교되는 두 특징의 길이 차이가 작고 두 특징간의 거리가 작을수록 오차가 작아진다는 것을 의미한다. 밝기공간에서의 오차는

$$D_i(s_i, t_j) = \frac{d}{D_{max}} \left\{ \frac{1}{N^2} \sum_m \sum_n |f_R(x+m, y+n) - f_T(x+m, y+n+d)| \right\} \quad (13)$$

와 같이 표현된다. 중괄호 안의 수식은 $N \times N$ 윈도우 영역내의 MAD(mean of absolute differences)를 찾는 것이다. 상기 언급된 정합 과정을 설명하면 다음과 같다. 그림 7은 주어진 실영상의 특징 영상이며, 그림 7(a)의 타원으로 표시된 부분이 정합하고자 하는 참조 특징이다. 그림 7(b)에 다각형으로 둘러싸인 특징은 참조 위치로부터 탐색공간 사이의 목표 영상내에 정합후보로써 사용되는 후보 특징들이다.

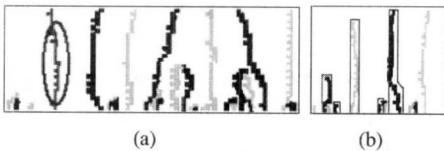


그림 7. 후보 추출과정 예 (a)참조 영상에서의 특징 (b)목표 영상에서의 후보들

추출된 후보 특징들은 식 (11)의 결과에 대한 유사정도에 따라 표 3과 같이 정합 순위가 결정되어, 최종적으로 정합이 이루어진다.

표 4는 한 라인에 대하여 정합이 수행된 후의 정합이 이루어지지 않은 특징을 표시하였다. 표 4에서 중복 플래그는 참조 화소에서 목표 측으로 참조한 중복성을 표시하고 결과 필드에서 오정합은 정합 오차를 최소로 선택한 후 정합되지 않거나 중복 정

합된 특징을 보여준다. 표 5는 “Pentagon” 영상과 RDS(Random Dot Stereogram) 영상에서 추출된 오정합 화소의 비율을 각각 나타낸다. 그림 8은 한 라인에서 오정합된 특징들을 표시한 것이다. 앞서 살펴본 것과 같이 특징의 강도가 큰 부분은 위상 영상에서도 그 값이 크므로 오정합 되지 않는 부분이 존재하지 않지만 위상 영상에서 작은 부분을 차지하는 부분에서 오정합이 발생되므로 2차 보간을 사용한 번이 개선으로 제거될 수 있다.

표 3. 그림 7에 대한 정합 결과.

No.	Candidate position	Disparity	Feature type	Matching results using Eq. (10)	Rank
1	5	-10	ZC	3.4409	4
2	9	-9	ZC	2.3324	3
3	14	-2	PK	0.0000	1
4	23	9	ZC	1.7889	2
5	30	12	ZC	3.7202	5

표 4. 중복 정합이 발생한 예.

Source features		Matched features in target			Results	
Point	Feature type	Point	Feature type	Duplication flag	False matching	Depth
5	PK	14	PK	○		19
7	ZC	9	ZC			13
11	PK	14	PK	○		15
17	PK	14	PK	○		10
29	ZC	30	ZC			24
31	PK	38	PK			15
33	ZC	42	ZC			24
36	PK	44	PK			20
41	ZC	52	ZC	○		24
43	PK	50	PK			16
44	ZC	52	ZC	○		17
49	PK	54	PK			19
51	ZC	52	ZC	○		12
53	PK	56	ZC		○	14

표 5. 전체 영상에서의 중복 및 오정합 화소 비율.

Image	Size	Duplicated pixels	False matched pixels	False matched rate(%)
RDS	256 × 256	8192	4352	6.6
Pantagon	512 × 512	34304	12288	4.7

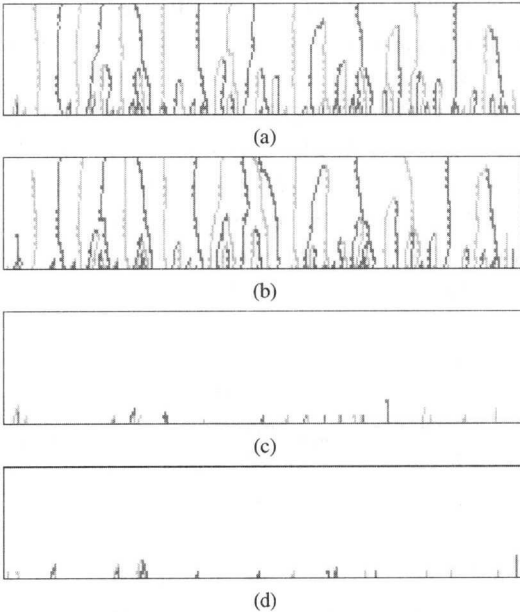


그림 8. 오정합된 특징들 (a)참조 특징 (b)목표 특징 (c)정합 후 참조 측에 남은 특징 (d)정합 후 목표 측에 남은 특징

3. Mesh rendering

계산된 변이로부터 3차원 모델을 생성하기 위해서 삼각법을 적용하였다. 공간적으로 인접한 4개의 정점, $V_{i,j}(i, j, d_{ij})$, $V_{i+1,j}(i+1, j, d_{i+1,j})$, $V_{i,j+1}(i, j+1, d_{i,j+1})$ 그리고 $V_{i+1,j+1}(i+1, j+1, d_{i+1,j+1})$ 을 이용하여 식 (14)처럼 삼각형 T_1 과 T_2 가 생성될 수 있다.

$$T_1 = F(V_{i,j}(i, j, d_{i,j}), V_{i+1,j}(i+1, j, d_{i+1,j}), V_{i,j+1}(i, j+1, d_{i,j+1})) \quad (14)$$

$$T_2 = F(V_{i+1,j}(i+1, j, d_{i+1,j}), V_{i,j+1}(i, j+1, d_{i,j+1}), V_{i+1,j+1}(i+1, j+1, d_{i+1,j+1}))$$

여기서 $F(\cdot)$ 는 삼각형 생성함수이다. 전체 영상에 획득한 변이 값들에 대해서 식 (14)을 적용하여 메쉬를 생성하면 3차원 객체를 얻을 수 있다.

IV. 실험 및 고찰

제안한 3차원 객체추출 기법의 성능분석을 위해 그림 9, 10의 128×128 크기의 50% RDS(random dot stereo gram)와 512×512 크기의 “pentagon” 영상을 사용하였다.

RDS에 대한 오정합율의 비교하기 위해 SSE(sum of squared error)와 MSE(mean of squared error)를

기준으로 표 6의 결과를 얻었다. 이 결과는 이완처리를 포함한 휘도기반의 정합에 비해 오정합을 40% 정도 줄였으며 계산량 또한 7.6배정도 감소되었다.

그림 11은 획득한 변이를 VRML로 표현한 3차원 객체 결과를 나타내며, 그림 12은 원 영상을 획득한 3차원 메쉬에 겹쳐서 표현한 것이다.

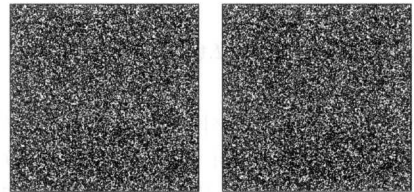


그림 9. 50% RDS(random noise 20%)

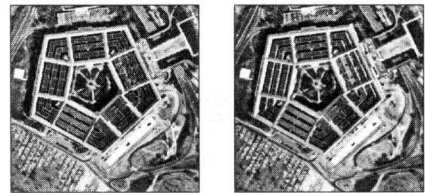


그림 10. “Pentagon” 영상

표 6. 50% RDS에 대한 이완처리를 포함한 휘도기반 방법과 제안한 방법의 정합 오차비교

Error \ Method	Intensity-based method with relaxation scheme	The proposed method
SSE	1375	836
MSE	0.084	0.051

V. 결론

본 논문에서 제안한 3차원 객체 추출 기법은 기존 위상 기반의 계층적 구조를 그대로 이용하면서 주파수 변환에서 요구하는 과도한 계산의 중복성을 제거하는 정합 방법을 제시하였다. 512×512 크기의 영상에서 비교할 때, 기존의 WFP의 위상 계산 연산을 1114배로 줄였고 전체 계산시간은 7.6배 줄일 수 있었다. 또한, 이완처리를 포함한 보편화된 BMA기법에 비해 40% 정도로 오정합율을 줄일 수 있었다. 그러므로 본 논문에서 제안한 기법은 오정합율과 성능면에서 안정된 결과를 제시할 수 있었다. 이렇게 추출된 3차원 깊이 정보를 컴퓨터 및 인터넷 환경에서 활용할 수 있게 VRML(virtual reality markup language)을 이용하여 3차원 메쉬로

