

## 고속 푸리에 변환을 이용한 계층적 위상기반 3차원 객체 추출 기법

한 규 필, \*이 채 수, \*박 양 우, \*\*엄 태 역  
금오공과대학교 컴퓨터공학부, \*경운대학교 멀티미디어정보학부, \*\*구마기농대학 전자과  
전화 : 054-467-4280

### 3D Object Extraction Algorithm Based on Hierarchical Phase Using Fast Fourier Transform

Kyu-Phil Han, \*Chae-Soo Lee, \*Yang-Woo Park, \*\*Tae-Uk Uam  
School of Computer Eng., Kumoh Nat'l Univ. of Tech.  
E-mail : kphan@knut.kumoh.ac.kr

#### 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 cost is too heavy. Therefore, a fast hierarchical approach, using multi-resolution phase-based strategy and reducing redundancies of phase calculations based on FFT concept, is proposed in this paper. In addition, a structural matching algorithm on the phase domain is presented 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 from various images.

#### I. 서론

최근 컴퓨터 및 인터넷 환경에서 멀티미디어 데이터의 사용이 급격히 증가하고 있다. 멀티미디어는 문자, 음성, 영상의 각 모노미디어들이 복합된 것으로 현장감을 전달할 수 있다는 장점으로 인해 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 기계 시각(machine vision) 분야[1]에서 이용되어져 왔지만 최근 많은 컴퓨터 통신환경에서 입체 영상 전송을 위해 영상열 간의 공간적 중복성을 제거하거나 인터넷에서 3차원 객체를 실시간으로 전송하는 분야에 활용되고 있다.

3차원 깊이정보를 효율적으로 추출하는 가장 대표적인 방법은 스테레오 시각을 이용하는 것이다. 스테레오 시각 기법은 인간의 시각 시스템과 같이 서로 다

른 관점(view point)에서 얻어진 두 장의 영상으로부터 3차원 정보를 복원하는 수동 거리측정 방법의 하나이다[1]. 스테레오 시각에서 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)[2]과 밝기 값을 그대로 사용하는 휘도 기반법(intensity-based method)[3], 주파수변환에서 추출된 위상을 이용하는 위상 기반법(phase-based method)[4] 등이 대표적이다. 특징 및 휘도 기반 방법들은 국부적인 공간상에서 주변 값의 정보만을 이용하기 때문에 오정합 가능성이 높다. 그러므로 이 방법들은 성능을 높이기 위해 계층적인 접근기법을 도입하여 적용되어진다. 위상 기반법은 기본적으로 계층적 구조를 사용하기 때문에 보다 나은 결과를 얻을 수 있지만 주파수 변환과정과 계층적 구조를 획득하기 위해 과도한 계산량을 요구한다는 단점이 있다[4].

본 논문에서는 기존 위상 기반의 계층적 구조를 그대로 사용하기 위해서 주파수 변환에서 요구하는 과도한 계산의 중복성을 제거하여 계산량을 줄이고 위상의 형태에 기반한 정합 방법을 제시한다. 또한, 이렇게 추출된 3차원 깊이 정보를 컴퓨터 및 인터넷 환경에서 활용할 수 있게 VRML(virtual reality mark up language)로 3차원 메쉬로 표현한다. 실험을 통하여 제안된 방법은 기존의 위상기법에 비해 8배정도의 계산량을 줄일 수 있었으며 성능 면에서도 우수한 결과를 얻을 수 있었다.

#### II. Fourier phase와 windowed Fourier phase

##### 2.1 Fourier phase와 전역이동(global shift)

무한의 연속 신호,  $f(x)$ ,와 이 신호의 푸리에 변환신호를  $F(u)$ , 그리고 이동된 신호,  $f_s(x)$ ,가 있다고 가정하

면 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 푸리에 kernel,  $e^{j2\pi vx}$ ,을 공간영역에서의 상승적분 식을 적용하여 표현하면

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

이 된다. 여기서  $arg[a(x)] \equiv 2\pi vx + arg[F(u)]$ 이다. 같은 방법을 적용하면 이동된 신호는 다음과 같다.

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

여기서  $arg[a_s(x)] \equiv 2\pi v(x-s) + arg[F(u)]$ 이다. 식(2)와 식(3)의 위상 정보의 차이는

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

이고 여기에서 이동된 변이,  $s$ ,는

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

계산될 수 있다. 식(5)가 내포하고 있는 것은 하나의 파형  $f(x)$ 에 포함된 특정 주파수,  $v$ ,가 전체 파형에서 얼마만큼 이동했는가하는 전역 이동 정보(global shift information)를 포함한다. 그러므로 푸리에 변환에서 각 주파수 성분이 가지는 위상 값은 해당 주파수의 전역 이동정보를 포함하고 있으므로 신호 정합(signal matching)에 사용되는 국부적인 파형의 이동정보를 표현하기에는 부족한 면이 있다[4].

### 2.2 WFP와 지역이동(local shift)

신호의 지역적 이동을 다루기 위해서는 전 파형의 일부를 표본화하는 윈도우 함수가 필요하다. WFP에서는 윈도우 함수에 의해 표본화된 신호가 전체 파형으로 간주되어서 그 파형의 이동 정보를 계산하기 때문에 윈도우 함수의 크기를 조절함으로써 각 국부영역의 이동성분을 계산할 수 있다. 크기가  $M$ 인 윈도우 함수를

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

라고 하면 windowed Fourier kernel은

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

이 되고 이것을 식(2)와 식(3)과 같이 원 신호 및 이동

된 신호에 상승적분하여 각각의 위상을 계산할 수 있다. 그림 1은 Windowed 푸리에 위상을 계산하기 위한 신호에서 표본화되는 과정을 보여준다.

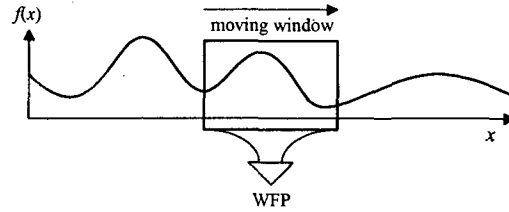


그림 1. 창의 중심 위치가  $x$ 일 때 WFP가 계산되는 영역

일반적으로 창의 크기가 고정되면 그 창의 크기를 주기로 설정하여 창의 크기를 가변시킨다. 그러므로 창의 크기  $M$ 이 곧 파형의 주기(period)가 되고 공간주파수는  $1/M$ 이 된다.

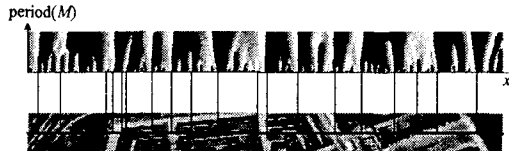


그림 2. 실제 파형에서 계산된 WFP( $5 \leq M \leq 32$ )

그림 2는 이런 위상정보가 실제 파형에 적용되어진 예를 보여주는데, 영상에서 공간 영역의 밝기가 위상 정보에 그대로 반영된 것을 알 수 있다. 그리고 파형의 변화가 강할수록 창의 크기 즉, 주기가 큰 영역에서 존재함을 볼 수 있다. 또한, 주기가 큰 위상 공간에서는 강한 변화만 나타나므로 드문 정함 후보를 가지게 된다. 그러므로 창의 크기가 큰 곳에서 작은 곳으로 계층적인 정함을 수행하면 정확한 변이를 계산할 수 있다.

### III. 제안한 3차원 객체추출 기법

제안한 시스템의 전체 블록도는 그림 3에서와 같이 위상 계산의 중복성을 제거한 다음 계층적 정함이 이루어지고 위상의 특징이 없는 부분에 대해서 선형 보간을 한 후, 위상의 특징점 주변으로 변이를 refine한다. 그리고 메쉬로 변이 맵을 표현해서 최종적인 객체를 획득한다.

#### 3.1 위상 계산의 중복성 제거

일반적으로 푸리에 변환은 cosine kernel에 의한 실수 성분과 sine kernel에 의한 허수 성분으로 구분되고 각 성분에 tangent 함수의 역함수로서 위상이 계산되어진다. 푸리에 변환을 고속으로 계산하기 위해 개발된 고속푸리에 변환은 창의 크기를 2의 지수승으로 증가시켜 삼각함수 계산의 중복성을 제거한 알고리즘이다. 그러나 고속 푸리에 변환을 WFP에 적용하면 창의 크기가 2, 4, 8, 16, 32, ...의 형태로 증가되기 때문에 파형의 변화가 너무 불연속하게 분포하므로 계층기반에서 연속성이 파괴되어지고 정확한 정함을 수행할 수 없게 된다. 그러므로 본 논문에서는 창의 크기를 정수

### 고속 푸리에 변환을 이용한 계층적 위상기반 3차원 객체 추출 기법

단위로 증가시키면서 각 삼각함수의 반복성을 이용하여 sine과 cosine kernel의 중복을 제거한다. sine과 cosine함수는 한 주기( $2\pi$ )내에서  $\pi$ 를 기준으로 같은 값이 반복되기 때문에 고속푸리에 변환과 같이 삼각함수의 kernel 값을 메모리 공간에 저장하면 매번 반복되는 계산을 메모리 참조로 대체시켜 속도를 높일 수 있다.

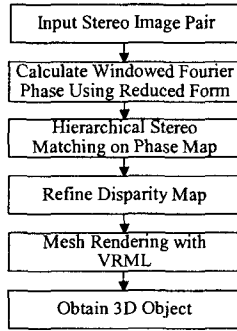


그림 3. 제안한 시스템의 전체 블럭도

기존의 WFP 계산량을 살펴보기 위해 창 크기의 5인 경우를 고려하면 표 1에서와 같이 10번의 kernel 계산이 요구된다. 이것은 창 크기가 변할 경우 새로운 kernel 값을 계산해야만 한다.

표 1. 창 크기가 5일 때 각 kernel의 값( $v=1/M$ )

| 공간변수<br>kernel (x) | 0 | 1     | 2      | 3      | 4      |
|--------------------|---|-------|--------|--------|--------|
| $\cos(2\pi x/M)$   | 1 | 0.309 | -0.809 | -0.809 | 0.309  |
| $\sin(2\pi x/M)$   | 0 | 0.951 | 0.587  | -0.587 | -0.951 |

신호의 크기가 512-sample이고 창이 5~32까지 변화된다고 가정할 때, 기존의 WFP의 계산량은 다음과 같다.  $O(WFP) = 512 \times 2 \times (5 + 6 + 7 + \dots + 32) = 265,216$  연산이 필요하다. 이것은 1차원 신호에 대한 것이고 2차원 영상신호라면 수직 크기만큼의 계산이 더 곱해져야 한다. 만약 kernel값이 메모리에 저장되어 있다면 표 1에서 보는 것과 같이 공간변수 3이상에서는 이전 값이 반복되는 것을 볼 수 있다. 즉

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

성질을 이용하고 모든 화소위치에서 kernel을 계산하지 않고 메모리 참조를 이용하면 계산의 비용은  $O(RWFP) = 1/2 \times (5 + 6 + 7 + \dots + 32) + (\text{메모리 참조 overhead}) = 259$ 연산 + 메모리 참조 비용이 된다. 512 크기의 신호인 경우 연산의 횟수는 1024배 정도 줄었지만 메모리 참조가 많아서 전체 속도는 실험에서 볼 수 있지만 8배 정도의 평균 계산 시간을 줄일 수 있었다.

### 3.2 계층적 위상기반 정합

그림 4는 원 신호와 이동된 신호의 WFP를 각각 보여준다. 그림 4의 각 파형은 국부적으로 위상이 이동된 변이만큼 약간씩 변화되어 있다는 것을 알 수 있다. 그러므로 정합에 사용될 위상의 특징은 위상에서

모든 값을 사용하지만 초기 정합의 성능을 높이기 위해 위상의 정점(peak)과 영교차점(zero crossing point)이 사용되었다. 왜냐하면 위상의 정점과 영교차점이 파형의 정점과 valley에 각각 대응되기 때문이다. 그림 5는 그림 4에서 각각 위상의 특징을 구한 것을 나타낸다. 특징 선에서 검은 부분이 정점들이고 회색선은 영교차점이다.

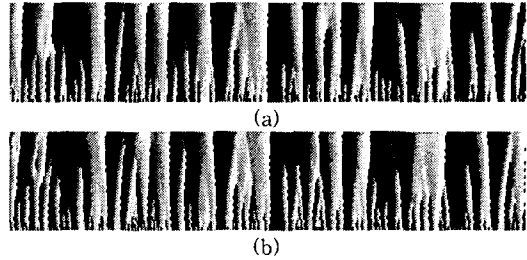


그림 4. 원 신호와 이동된 신호의 WFP. (a)원 신호 (b)이동된 신호

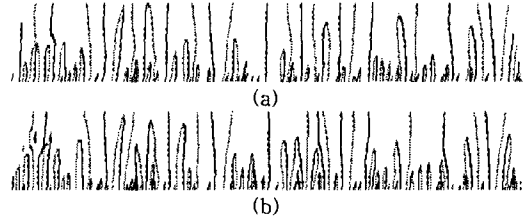


그림 5. 그림 4의 위상 맵에서 구한 특징 영상. (a)원 신호 (b)이동된 신호

정합은 저해상도(주기가 큰 곳)에서 두 위상의 특징들을 정합하고 정합 과정은 다음과 같이 수행한다.

- 1) 각 조각(segment)의 labeling(2-pass algorithm)
- 2) 탐색 영역 내부에서 후보(candidate) 조각 추출
- 3) 원 신호의 밝기 값과 위상 차가 가장 작은 일치(correspondence) 조각 선택
- 4) 최고 해상도(최저 크기의 창)의 변이 값 산출
- 5) 추출된 feature에서 계산된 변이 값을 기준으로 주변 값을 선형보간(linear interpolation)
- 6) 보간된 값을 기준으로 주변  $\pm 1$ 화소 값 사이에서 변이 값 refinement

정합에 사용된 거리척도(distance measure)는 위상 및 휘도 값을 모두 고려했으며 최소거리를 갖는 후보를 정합 점으로 선택하였다. 사용된 척도는

$$D(s_i, s_j) = w_p MAD_p(s_i, s_j) + w_i MAD_i(s_i, s_j) \quad (9)$$

와 같이 각 조각의 위상과 휘도의 MAD(mean of absolute difference)를 사용하였다.  $w_p$ 와  $w_i$ 는 위상과 휘도의 크기를 보상하는 scale factor이다. 휘도는 최대 차이가 255이지만 위상은  $-\pi \sim \pi$ 이므로 각각의 크기를 normalization 시켜서 사용하였다.

### 3.3 Mesh rendering

지금까지 계산된 변이로부터 3차원 모델을 생성시키기 위해서 단순한 삼각법을 이용하였다. 공간적으로

인접한 4개의 정점,  $V_{ij}(i, j, d_{ij})$ ,  $V_{i+1,j}(i+1, j, d_{i+1,j})$ ,  $V_{i,j+1}(i, j+1, d_{i,j+1})$ , and  $V_{i+1,j+1}(i+1, j+1, d_{i+1,j+1})$ 이 있다고 가정하자. 여기서  $d_{ij}$ 는 위치  $(i, j)$ 에서 변이를 나타낸다. 따라서 두 개의 삼각형  $T_1$ 과  $T_2$ 가 생성될 수 있다.

$$T_1 = F(V_{ij}(i, j, d_{ij}), V_{i+1,j}(i+1, j, d_{i+1,j}), V_{i,j+1}(i, j+1, d_{i,j+1}))$$

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

(10)

여기서  $F(\cdot)$ 는 삼각형을 생성시키는 함수이다.

#### IV. 실험

제안한 3차원 객체추출 기법의 성능을 분석하기 위해 그림 6, 7과 같이  $128 \times 128$  크기의 50% RDS (random dot stereogram)과  $512 \times 512$  크기의 "pentagon" 영상을 사용하였다. 표 2는 RDS에 대한 오정합율의 비교를 나타내었다. 오차값은 SSE(sum of squared error)와 MSE(mean of squared error)를 비교하였다. 표 3은 pentagon 영상에 대한 계산시간 비교를 나타낸다. 시뮬레이션 환경은 IBM PC(pentium-III, Clock 750MHz)에서 실험한 계산시간을 나타낸다.

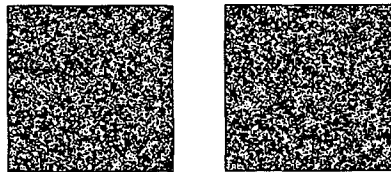


그림 6. 50% RDS(random noise 20%)



그림 7. "Pentagon" 영상

표 2와 3에서 보는 것과 같이 이완처리를 포함한 휘도 기반의 정합에 비해 오정합율 40%정도 줄었으며 계산량 또한 7.6배 정도 줄일 수 있었다. 그림 8과 9는 계산된 변이를 VRML로 표현한 3차원 객체를 각각 나타낸다. 그림 8(a)와 (b)를 비교해 볼 때, 제안한 방식으로 더 나은 결과를 얻을 수 있었다. 또한 그림 8(c)와 9(b)는 원 영상을 획득한 3차원 메시에 겹쳐서 표현한 것으로 인터넷 환경이나 컴퓨터 환경에서 사용할 수 있는 결과이다.

#### V. 결론

제안한 3차원 객체 추출 기법은 기존의 WFP의 위상 계산을 sine과 cosine kernel의 중복계산을 줄여  $512$  크기의 영상에서 비교할 때, 연산을 1024배로 줄였고 전체 계산시간은 7.6배 줄일 수 있었다. 또한, 이완처리를 포함한 보편화된 BMA기법에 비해 40% 정도의 오정합율을 줄일 수 있었다. 그러므로 본 논문에서 제안한 기법은 오정합율과 성능면에서 안정된 결과를 제시할 수 있었다 그러나 변이로부터 3차원 물체를

표현하기 위해 단순한 삼각법을 사용하였기 때문에 3차원 자료를 단순화시키지 않았다. 향후 보다 효율적인 3차원 객체를 다루기 위해 메쉬를 단순화하는 연구가 더 필요하다고 할 수 있다.

표 2. 50% RDS에 대한 정합 오차비교

| 정합 기법 | 휘도기반<br>(이완처리 포함) | 제안한 방법 |
|-------|-------------------|--------|
| SSE   | 1375              | 836    |
| MSE   | 0.084             | 0.051  |

표 3. "Pentagon"에서의 계산량 비교

| 정합 기법       | WFP       | 제안한 방법   |
|-------------|-----------|----------|
| 평균시간        |           |          |
| 1 line 정합시간 | 0.244sec  | 0.032sec |
| 전체 정합시간     | 125.17sec | 16.53sec |

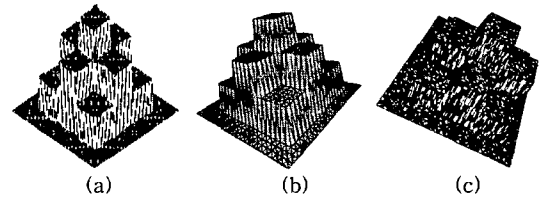


그림 8. RDS의 결과. (a)이완처리를 포함한 BMA (b)제안한 방식 (c)제안한 방법의 영상 포함 결과

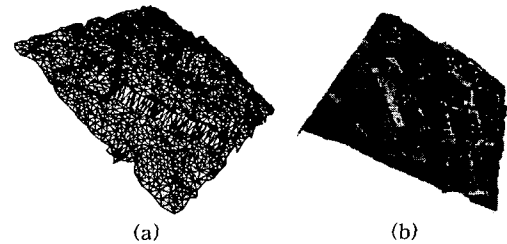


그림 9. 제안한 방식의 "pentagon" 결과. (a)3차원 mesh (b)영상이 포함된 결과

#### 참고문헌

- [1] K. Sugihara, "Three principle in stereo vision," Robotics, vol. 1, no. 4, pp. 391-400, 1986.
- [2] W. Eric L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm," *IEEE Trans. Patt. Anal. Machine Intell.* vol. 7, no. 1, pp. 17-34, 1985.
- [3] D. De Vleeschauer, "An Intensity-based, Coarse-to-fine Approach to Reliably Measure Binocular Disparity," *CVGIP: Image Understanding* vol. 57, no. 2, pp. 204-218, 1993.
- [4] Juyang Weng, "Windowed Fourier Phase: Completeness and Signal Reconstruction," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 657-666, 1993.