

논문 2009-46SP-4-15

저니키 모멘트 기반 지역 서술자를 이용한 실시간 특징점 정합

(Real-Time Feature Point Matching Using Local Descriptor Derived by Zernike Moments)

황 선 규*, 김 회 율**

(Sun-Kyoo Hwang and Whoi-Yul Kim)

요 약

서로 다른 시점의 두 영상에서 동일한 점들을 정합하는 특징점 정합은 다양한 영상 처리 분야에서 널리 사용되고 있으며, 최근에는 실시간으로 동작하는 특징점 정합에 대한 요구가 높아지고 있다. 본 논문은 저니키 모멘트 기반의 지역 서술자를 이용하여 특징점을 실시간으로 정합하는 방법을 제안한다. 빠른 모서리 점 검출 방법을 이용하여 입력 영상으로부터 특징점을 추출하고, 각 특징점에서 저니키 모멘트를 이용한 지역 서술자를 생성한다. 저니키 모멘트 기반의 지역 서술자는 특징점 주변의 부분 영상을 적은 차수의 특징 벡터로써 효율적으로 표현하며, 영상의 회전과 밝기 변화에 강인하다. 본 논문에서는 저니키 모멘트 계산을 실시간으로 수행하기 위하여 고정된 크기의 저니키 기저 함수를 미리 계산하여 이를 룩업 테이블에 저장하여 사용한다. 특징점 정합 단계에서는 근사 최근방 이웃(ANN) 방법을 사용하여 초기 정합 결과를 얻고, 이 중 잘못된 정합은 RANSAC 알고리즘을 이용하여 제거함으로써 최종 정합 결과를 얻는다. 실험 결과 제안하는 방법은 다양한 변환이 존재하는 영상에 대하여 실시간으로 특징점 정합을 수행함을 확인하였다.

Abstract

Feature point matching, which is finding the corresponding points from two images with different viewpoint, has been used in various vision-based applications and the demand for the real-time operation of the matching is increasing these days. This paper presents a real-time feature point matching method by using a local descriptor derived by Zernike moments. From an input image, we find a set of feature points by using an existing fast corner detection algorithm and compute a local descriptor derived by Zernike moments at each feature point. The local descriptor based on Zernike moments represents the properties of the image patch around the feature points efficiently and is robust to rotation and illumination changes. In order to speed up the computation of Zernike moments, we compute the Zernike basis functions with fixed size in advance and store them in lookup tables. The initial matching results are acquired by an Approximate Nearest Neighbor (ANN) method and false matchings are eliminated by a RANSAC algorithm. In the experiments we confirmed that the proposed method matches the feature points in images with various transformations in real-time and outperforms existing methods.

Keywords: Feature point, matching, Zernike moments, local descriptor, real-time

* 정회원, 한양대학교 산업과학연구소
(Research Institute of Industrial Science,
Hanyang University)

** 정회원, 한양대학교 전자컴퓨터통신공학파
(Dept. of Electronics and Computer Engineering,
Hanyang University)

※ 이 논문은 2006년 정부(교육인적자원부)의 재원으로
한국학술진흥재단의 지원을 받아 수행된 연구임.
(KRF-2006-352-D00150)

접수일자: 2008년12월16일, 수정완료일: 2009년6월10일

I. 서 론

특징점 정합은 시점이 다른 두 영상에서 동일한 위치의 점들을 찾아 서로 연결하는 작업을 의미한다. 3차원 공간을 다른 시점에서 촬영한 영상들 사이에는 회전, 크기, 이동 등의 변환이 발생하게 된다. 이러한 변환이 존재하는 두 영상에서 동일한 특징점을 찾아 정합하는

작업은 객체 인식^[10, 16], 3차원 복원^[7], 스테레오 비전^[14], 증강 현실^[9] 등의 다양한 영상 처리 분야에서 활용되고 있다. 최근에는 이동 로봇의 자기 위치 추정 및 지도 작성(SLAM)에서도 특징점 정합 기술이 사용되고 있다^[17]. 이 중 SLAM, 스테레오 비전, 증강 현실 등의 영상 처리 분야에서는 특징점 정합의 높은 정확도뿐만 아니라 실시간 동작을 필요로 하고 있다.

특징점 정합 과정은 일반적으로 특징점 검출, 특징점 기술, 그리고 실제 특징점 정합의 세 단계로 구성된다. 첫 번째 특징점 검출이란 영상에 다양한 변환이 가해져도 동일한 위치의 점을 반복적으로 찾는 과정을 의미한다. 특징점 기술은 특징점 주변의 국지적 픽셀 분포 정보를 이용하여 각 특징점을 표현하는 특징 벡터를 계산하는 단계를 의미한다. 이 특징 벡터는 특징점 근방의 영상 특성을 표현하기 때문에 지역 서술자라고 부른다. 마지막으로 특징점 정합은 추출된 특징 벡터를 이용하여 실제 정합을 수행하는 단계를 의미한다. 이 세 가지 단계가 모두 적은 연산량으로 동작해야 실시간 특징점 정합이 수행될 수 있다.

특징점 검출과 기술을 위해 널리 사용되는 방법 중 하나는 Lowe가 제안한 SIFT(Scale-Invariant Feature Transform)이다^[10]. SIFT는 영상의 크기, 회전 등에 불변한 특징점을 찾아 이를 방향 히스토그램으로 표현하는 방식이다. 이 방법은 스케일의 변화가 큰 영상에서도 특징점을 효과적으로 검출하고 기술하는 장점이 있다. 그러나 SIFT 알고리즘은 다양한 스케일의 영상을 생성하여 특징점을 찾고 특징 벡터를 추출하기 때문에 연산 속도가 느리다는 단점이 있다. 그러므로 SIFT는 실시간 연산이 필요한 로봇 비전, 증강 현실 등의 응용 프로그램에서 사용되기에는 적합하지 않다. Lepetit와 Fua는 특징점 정합을 패턴 분류 문제로 인식하여 카메라 포즈에 강인한 정합 알고리즘을 제안하였다^[9]. 기존 영상으로부터 다양한 시점의 영상을 생성하여 각 영상에서 특징점을 추출하고, 임의 트리(randomized tree) 자료 구조를 사용하여 특징점에 대한 정보를 저장한다. 그리고 입력 영상이 들어오면 트리 구조로부터 유사한 특징점을 빠르게 검색한다. 이 방법은 특징점 정합 자체는 실시간으로 동작하지만 참조 영상으로부터 임의 트리를 구성하는 과정에서 수 십 초 이상의 시간이 소요된다는 단점이 있다. Tran과 Marchand는 비주얼 서보잉(visual servoing)을 위한 빠른 특징점 정합 방법을 제안하였다^[15]. 이 방법은 영상 내의 모서리 점을 찾아

이를 특징점으로 삼고, 각 특징점에서 주 방향 성분을 찾아 회전한 후, 주축 성분 분석(PCA)에 기반한 지역 서술자를 생성하여 정합에 사용하였다. 그러나 이 방법은 PCA를 통한 고유 벡터 추출 과정이 필요하고, 특징 벡터 추출을 위해 각 특징점 근방의 부분 영상을 회전하는 과정이 필요하다는 단점이 있다. 또한 모서리 점에서 그래디언트 방향 성분을 이용하여 검출된 주축 방향의 정확도가 낮기 때문에 회전 변환에 대하여 다소 부정확한 정합을 수행한다는 단점도 존재한다.

본 논문은 저니키 모멘트로부터 유도한 지역 서술자를 이용하여 실시간으로 특징점을 정합하는 방법을 제안한다. 먼저 빠른 모서리 점 검출 방법을 이용하여 입력 영상으로부터 특징점을 검출하고, 각 특징점 근방의 부분 영상으로부터 저니키 모멘트 기반의 특징 벡터를 추출한다. 저니키 모멘트는 적은 수의 모멘트 개수를 이용하여 영상의 특징을 효과적으로 표현할 수 있기 때문에 다양한 패턴 인식 시스템에서 사용되어왔다. 특히 저니키 모멘트의 절대값 성분은 영상의 회전에 불변한 특징을 가지고 있기 때문에 특징 벡터 추출을 위하여 영상을 회전하는 과정이 필요하지 않다. 또한 조명 변화에도 강인한 특성을 부여하기 위하여 저니키 모멘트 절대값의 정규화 과정을 수행하였다. 제안하는 방법은 고정된 크기의 부분 영상으로부터 저니키 모멘트를 계산하기 때문에 저니키 기저 함수를 미리 계산하여 룩업 테이블(lookup table)에 저장하고, 이를 이용하여 저니키 모멘트를 계산한다. 이 경우 특징점 근방의 부분 영상을 룩업테이블에 저장된 저니키 기저 함수에 투영하고 그 합을 구함으로써 매우 빠르게 저니키 모멘트를 계산할 수 있다. 실제 특징점 정합 과정에서는 근사 최 근방 이웃(ANN) 방법을 이용하여 초기 정합을 수행하고, RANSAC 알고리즘을 사용하여 잘못된 정합을 제거한다. 실험에서 제안하는 방법은 크기, 회전, 이동, 밝기 변화, 그리고 가려짐이 있는 영상에 대하여 실시간 특징점 정합을 성공적으로 수행함을 확인하였다.

본 논문의 구성은 다음과 같다. II장에서는 빠른 특징점 추출 방법에 대하여 설명한다. III장에서는 저니키 모멘트와 저니키 모멘트 기반의 지역 서술자에 대하여 자세히 서술한다. IV장에서는 근사 최 근방 이웃 방법과 RANSAC 알고리즘에 의한 특징점 정합 방법에 대하여 설명한다. 다양한 영상에 대한 실험 결과는 V장에서 제시하고, VI장에서 결론을 내린다.

II. 특징점 검출

특징점은 이동, 크기, 회전 등의 다양한 기하학적 변형이 일어난 영상에서도 반복적으로 검출되는 점을 의미한다. 주로 국지적 변화가 큰 점들이 특징점으로 사용되기 적합하며, 주로 모서리 점들이 특징점으로 사용되었다. 모서리 점을 찾는 방법으로는 Harris 모서리 검출 방법^[5], 가우시안-라플라시안 방법^[11], SUSAN^[13] 방법 등이 주로 사용되었다. 크기 변화가 있는 영상에서도 동일한 특징점을 찾기 위해서 이들 방법에 스케일 스페이스(scale space) 개념이 함께 사용되기도 하였다. 그러나 이러한 방법들은 대부분 많은 연산량을 필요로 하기 때문에 실시간 응용프로그램에서 사용하기에는 무리가 있다.

최근에 빠르고 간단한 FAST(Feature from Accelerated Segment Test) 모서리 점 검출 방법이 제안되었다^[12]. 이 방법은 영상의 모든 픽셀에 대하여 주변 16개 픽셀을 검사하여 모서리 점인지 아닌지를 판별한다. 그림 1은 FAST 모서리 점 검출을 위해 사용되는 픽셀들의 분포를 보여준다. 점 p 주변의 검정색 테두리로 표시된 16개의 픽셀 값을 검사하여 p의 픽셀 값보다 밝거나 또는 어두운 픽셀이 연속으로 n개 이상 나타나면 이 점은 모서리 점으로 분류한다. 이 방법은 그림 1에 표시된 원형의 주변 픽셀들 중에서 1, 9, 5, 13번째 픽셀만을 검사하여 모서리 점이 아닌 픽셀을 미리 걸러낼 수 있다. 실험에서 FAST 모서리 점 검출 방법은 320×240 크기의 영상에 대하여 모서리 점들을 검출하는데 1~2ms 정도의 시간이 소요됨을 확인하였다. 본 논문에서는 FAST 모서리 점 검출 방법으로 추출된 픽셀들을 특징점으로 사용한다.

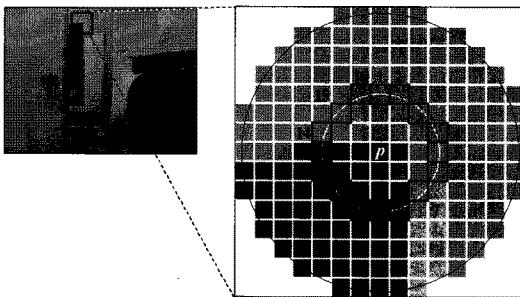


그림 1. FAST 모서리 점 검출 방법
Fig. 1. FAST corner point detector.

III. 저니키 모멘트를 이용한 지역 서술자

1. 저니키 모멘트

저니키 모멘트는 적은 개수의 모멘트를 이용하여 영상의 특징을 효과적으로 표현하기 때문에 많은 패턴 인식 시스템에서 사용되어 왔다^[8]. 저니키 모멘트는 입력 영상을 저니키 기저 함수 위에 투영하여 그 합을 계산함으로써 구해진다. 차수(order)가 n이고 반복수(repetition)가 m인 저니키 모멘트 Z_{nm} 은 다음 수식으로 정의된다.

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x,y) V_{nm}^*(x,y) dx dy \quad (1)$$

여기서 $f(x,y)$ 는 입력 영상을 의미하고, $V_{nm}(x,y)$ 는 저니키 기저 함수를 나타낸다. 기호 *는 복소 공액을 나타낸다. 저니키 기저 함수는 반지름이 1인 원형 안에서 정의되기 때문에 입력 영상은 $x^2+y^2 \leq 1$ 인 단위 원 크기로 정규화되어 계산에 사용된다. 위 식에서 차수 n은 0 또는 양의 정수이고, 반복수 m은 $n-|m|$ =(짝수)와 $|m| \leq n$ 을 만족하는 정수이다.

저니키 기저 함수 $V_{nm}(x,y)$ 는 다음 수식으로 정의된다.

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho) \exp(jm\theta) \quad (2)$$

여기서 $j = \sqrt{-1}$ 이고, $R_{nm}(\rho)$ 은 저니키 실수 방사 다항식이다. 서로 다른 차수 또는 반복수를 갖는 저니키 기저 함수 $V_{nm}(x,y)$ 는 단위 원 안에서 직교하기 때문에 각각의 모멘트 간에 정보의 중복이 없다. 이러한 특징은 서로 다른 차수 또는 반복수의 저니키 모멘트는 영상의 고유한 특성을 반영함을 나타내고, 적은 개수의 저니키 모멘트로 영상의 특징을 효과적으로 표현할 수 있음을 의미한다.

저니키 실수 방사 다항식은 다음과 같이 정의 된다.

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! (\frac{n+|m|}{2}-s)! (\frac{n-|m|}{2}-s)!} \quad (3)$$

이산 영상에 대하여 저니키 모멘트를 구하기 위해서, 수식 (1)을 다음과 같은 식으로 근사화할 수 있다.

$$Z_{nm} = \frac{n+1}{\lambda_n} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) V_{nm}^*(x,y) \quad (4)$$

위 식에서 λ_n 은 실제 계산에 사용되는 픽셀의 개수를 나타낸다.

2. 저니키 모멘트 기반의 지역 서술자

저니키 모멘트는 모델 기반 객체 인식을 위한 지역 서술자로 사용되어 이미 그 성능을 입증하였다^[16]. 이 지역 서술자는 회전에 대하여 불변하고 잡음과 조명 변화에 강인한 특징을 갖는다.

영상의 중심을 기준으로 α 만큼 회전된 영상을 $f^r(\rho, \theta) = f(\rho, \theta + \alpha)$ 라고 하자. 이 영상에 대하여 저니키 모멘트를 구하면 다음과 같은 수식이 성립한다.

$$Z_{nm}^r = Z_{nm} \exp(jm\alpha) \quad (5)$$

이는 회전된 입력 영상은 복소 저니키 모멘트의 위상에만 영향을 주고 그 절대값은 변하지 않음을 의미한다 ($|Z_{nm}^r| = |Z_{nm}|$). 그러므로 저니키 모멘트의 절대값을 특징값으로 사용할 경우 회전에 불변한 특징을 가지게 된다.

조명 변화에 대하여 불변한 특성을 주기 위하여 저니키 모멘트의 절대값을 Z_{00} 으로 나누는 정규화 과정을 사용한다. 조명의 변화가 생긴 영상을 $f^i(x, y) = \beta f(x, y)$ 으로 설정할 경우, 이 영상에 대하여 저니키 모멘트의 절대값을 구하면 다음과 같다.

$$|Z_{nm}^i| = \beta |Z_{nm}| \quad (6)$$

그러므로 이를 $|Z_{00}|$ 으로 나누면 조명 변화에 불변한 특징값을 얻을 수 있다. 이 특징값은 다음 수식으로 정리할 수 있다.

$$f_{nm} = \frac{|Z_{nm}|}{|Z_{00}|} \quad (7)$$

그러므로 i 번째 특징점에 대한 특징 벡터는 다음과 같이 정의된다.

$$f^i = \{f_{nm}^i | 1 \leq n \leq N, n-m = (\text{짝수}), 0 \leq m \leq n\} \quad (8)$$

여기서 N 은 계산하고자 하는 최대 차수를 의미한다. 차수와 반복수가 모두 0인 저니키 모멘트 Z_{00} 는 조명 변화에 불변한 특징값 계산을 위해 사용되었기 때문에, 특징 벡터에서 저니키 모멘트의 차수는 1부터 증가함을 볼 수 있다. 본 논문에서는 특징 벡터 추출을 위하여 $N = 7$ 로 설정하였고, 그 결과 특징 벡터의 길이는 19이다.

3. 룩업테이블을 이용한 실시간 연산

본 논문에서 제안하는 특징점 정합을 실시간으로 수행하기 위해서는 저니키 모멘트의 빠른 계산이 필수적이다. 저니키 모멘트는 그 기저함수 정의식이 매우 복잡하기 때문에 저니키 기저 함수 계산에서 많은 연산량을 필요로 한다. 저니키 모멘트의 계산을 빠르게 하는 다양한 방법들이 연구되었지만^[2], 이들 방법들은 여전히 실시간 동작에는 적합하지 않다.

본 논문에서는 앞서 구한 특징점 주변에서 고정된 크기의 부분 영상을 추출하고 이 부분 영상으로부터 저니키 모멘트 기반의 지역 서술자를 계산한다. 실제 구현에서는 부분 영상의 크기를 15×15 픽셀로 설정하였고, 이 사각형 영역에 내접하는 원형의 픽셀들을 저니키 모멘트 계산에서 사용하였다. 이 경우 부분 영상의 크기가 고정되어있기 때문에 저니키 기저 함수의 크기 또한 고정되어 사용된다. 그러므로 본 논문에서는 15×15 크기의 저니키 기저 함수를 미리 계산하여 이를 메모리에 저장한 후 사용하는 룩업테이블(lookup table) 방법을 사용하였다.

룩업테이블에 저장되어 있는 저니키 기저 함수의 일부($n \leq 4$)를 그림 2에 그레이스케일 영상의 형태로 나타내었다. 저니키 기저 함수의 값이 [-1, +1] 사이의 값을 가지므로, 이를 [0, 255] 사이로 선형 변환하여 그레이스케일 값의 범위를 갖도록 하였다. 반복수 m 이 0인 경우, 저니키 기저 함수의 허수부는 항상 0이기 때문에 그림 2에서 단순히 회색으로 표시되었다. 각 특징점에서 15×15 부분 영상을 추출하여 이를 룩업테이블에 저장된 저니키 기저 함수에 투영하여 모든 픽셀의 값을 합하면 저니키 모멘트가 계산된다. 이 경우, 일정수의 곱셈과 덧셈 연산만을 필요로 하기 때문에 매우 빠르게 저니키 모멘트를 계산할 수 있다.

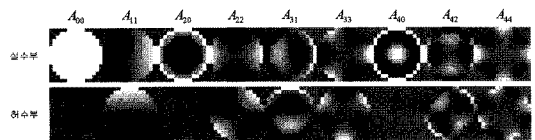


그림 2. 룩업테이블에 저장된 저니키 기저 함수의 일부
Fig. 2. Zernike basis functions stored in lookup table.

IV. 특징점 정합

기준 영상과 시점이 다른 입력 영상으로부터 특징점 추출과 지역 서술자 계산이 완료되면 특징점들의 실제

정합 과정이 필요하다. 본 논문에서는 두 단계로 구성된 특징점 정합을 사용한다. 첫 번째 단계에서는 빠르고 간단한 방법으로 정합을 수행하고, 두 번째 단계에서는 초기 정합 중에서 잘못된 정합을 제거하여 정확한 정합 결과만을 선택한다.

초기 정합을 수행하기 위하여 본 논문에서는 근사 최근방 이웃(Approximate Nearest Neighbor: ANN) 알고리즘을 사용하였다^[1, 18]. ANN 알고리즘은 기준 영상에서 추출된 특징 벡터들을 K-d 트리(K-d tree) 구조에 저장한 후, 입력 벡터와 가장 가까이에 위치한 벡터를 근사적으로 찾아주는 알고리즘이다. 본 논문에서 사용하는 저니키 모멘트 기반의 지역 서술자는 19차원의 특징 벡터를 갖기 때문에 K-d 트리에서 K 값은 19가 된다. ANN 알고리즘에서 두 특징 벡터 p와 q 사이의 거리 계산은 아래 수식과 같은 유클리디언 거리(Euclidean distance)를 이용한다.

$$dist(p, q) = \left(\sum_{i=0}^k (p_i - q_i)^2 \right) \quad (9)$$

ANN 알고리즘의 장점은 입력 벡터와 유사한 벡터를 매우 빠르게 검색한다는 점이다. 기준 영상에서 추출한 특징점들의 특징 벡터로 K-d 트리를 구성하는 복잡도 또한 $O(K \log K)$ 밖에 되지 않기 때문에 매우 빠르게 기준 영상을 등록할 수 있다.

그러나 단순한 유클리디언 거리로써 유사 특징 벡터를 검색하기 때문에 ANN 알고리즘에 의해 구한 초기 연결 관계는 잘못 연결된 정합이 다수 포함되어 있다. 이처럼 틀리게 정합된 결과를 걸러내기 위하여 RANSAC(RANdom SAmple Consensus) 알고리즘을 사용하였다^[3-4, 6]. 이 때 RANSAC 알고리즘에서 사용할 모델로는 두 영상 사이의 호모그래피(homography)를 사용하였다. RANSAC 알고리즘을 이용하여 잘못된 정합을 제거하는 방법은 다음과 같다.

1. 초기 정합 관계로부터 네 개의 연결 쌍을 임의로 선택한다.
2. 네 개의 연결 정보를 이용하여 호모그래피 H_{tmp} 를 계산한다.
3. 계산된 호모그래피를 이용하여 나머지 정합 중에서 정상인 정합의 개수를 확인한다.
4. 만약 정상인 정합 개수가 최대인 경우, 최종 호모그래피를 갱신한다. ($H=H_{tmp}$)
5. 1번부터 4번까지의 과정을 일정 횟수 동안 반복한다.

6. 최종으로 남는 호모그래피 H에 의해 정상으로 판별된 정합만을 최종 결과로 선택한다.

V. 실험

제안한 특징점 정합 방법의 성능을 평가하기 위하여 다양한 영상에 대한 실험을 수행하였다. 모든 실험은 윈도우 XP 환경에서 Visual C++ 2005로 구현하였고, 2.8GHz 데스크탑 컴퓨터에서 실행하였다. 다양한 영상에 대한 특징점 정합 결과를 그림으로 나타내고, 특징점 정합의 각 단계에서 소요된 연산 시간을 조사하여 정리하였다.

첫 번째 실험에서는 기준 영상에 회전, 크기, 이동 변환을 수행하고 밝기를 변화시킨 영상을 컴퓨터로 생성하여 이를 입력 영상으로 사용하였다. ANN 알고리즘에서 사용되는 K-d 트리는 기준 영상에서 추출한 특징 벡터 집합으로 처음 한 번만 생성하였다. 그림 3은 다양한 변환이 수행된 입력 영상에 대한 특징점 정합 결과를 보여준다. 그림 3(a)는 기준 영상을 90% 크기로 축소된 입력 영상에 대하여 ANN 알고리즘에 의한 초

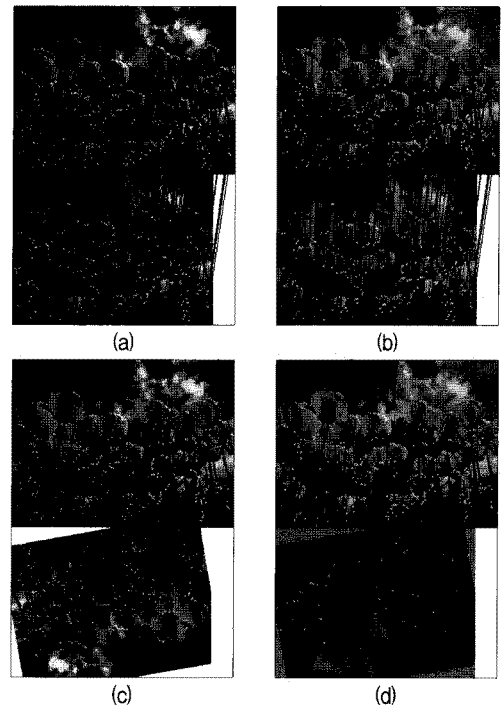


그림 3. 컴퓨터로 생성한 변환 영상에서의 특징점 정합
Fig. 3. Feature point matching with synthesized images.

표 1. 그림 3의 특징점 정합에 대한 연산 시간
Table 1. Computation time for each process for Fig. 3.

| 과정 | 시간 (ms) | | | |
|-----------|---------|--------|--------|--------|
| | 기준영상 | 그림3(b) | 그림3(c) | 그림3(d) |
| 특징점 검출 | 1 | | | |
| 지역 서술자 | 10(295) | | | |
| K-d 트리 | 1 | | | |
| 특징점 검출 | | 1 | 1 | 1 |
| 지역 서술자 | | 7(226) | 6(167) | 2(58) |
| ANN 초기 정합 | | 1 | 1 | 1 |
| RANSAC | | 8 | 4 | 2 |
| 합계 | 12 | 17 | 12 | 6 |

기 정합 결과를 보여준다. 이 결과 그림에서 다수의 잘못된 정합된 결과가 포함되어 있는 것을 확인할 수 있다. 초기 정합 결과에 RANSAC 알고리즘을 적용하여 올바른 정합만을 추출한 결과를 그림 3(b)에 나타내었다. 제안하는 특징점 정합 방법은 스케일 스페이스 개념을 사용하지 않았지만 약간의 크기 변환에는 강인하게 특징점을 정합하는 것을 확인하였다. 그림 3(c)는 그림 3(a)에서 사용한 축소 영상을 170° 회전하여 만든 영상에 대한 정합 결과를 보여준다. 저니키 모멘트 기반의 지역 서술자가 회전에 불변하기 때문에 단순한 회전 변환이 가해진 영상에 대하여 다수의 정합 결과를 얻을 수 있음을 확인하였다. 그림 3(d)는 그림 3(c)에서 사용된 영상을 70% 어둡게 변화시킨 후 정합을 수행한 결과이다. 이 경우, FAST 모서리 점 검출 방법에 의해 추출된 특징점의 개수가 다소 줄어들었지만 다수의 특징점 정합이 올바르게 수행된 것을 확인할 수 있다.

그림 3에 나타난 특징점 정합의 각 과정에서 소요된 시간을 표 1에 정리하였다. 기준 영상으로부터 특징점을 찾아 지역 서술자를 계산하고, K-d 트리를 생성하는 기준 영상 등록을 위해 모두 12ms 가 소요되었다. 이는 제안하는 특징점 정합 방법은 기준 영상을 등록하는 과정이 실시간으로 동작함을 의미한다. 크기, 회전, 밝기 등이 변환된 입력 영상으로부터 실제로 특징점을 찾고 정합하는데 걸린 시간은 20ms를 넘지 않았다. 지역 서술자 계산 시간 뒤 괄호 안에 써진 숫자는 검출된 특징점의 개수를 의미한다. 전체 연산 시간 중 대부분은 지역 서술자를 계산하는데 소요되며, 특징점의 개수가 적을수록 연산 시간이 단축되는 것을 볼 수 있다.

두 번째 실험에서는 스테레오 비전에서 사용되는 영상에 대하여 기존의 방법과 제안하는 방법으로 각각 특징점 정합을 수행하고, 그 성능을 비교하였다. 3차원 공

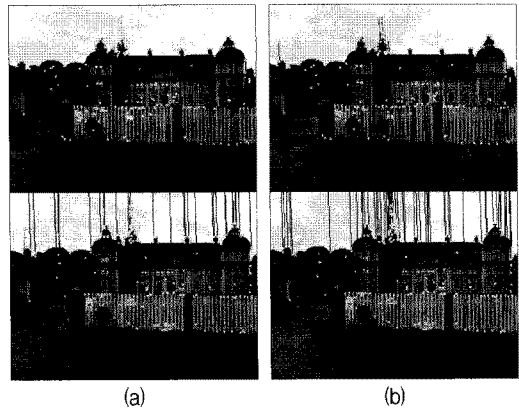


그림 4. 스테레오 영상에서 특징점 정합
Fig. 4. Feature point matching with stereo images.

표 2. 그림 4의 특징점 정합에 대한 연산 시간
Table 2. Computation time for each process for Fig. 4.

| 과정 | 시간 (ms) | |
|-------|-----------|---------|
| | 그림 4(a) | 그림 4(b) |
| 기준 영상 | 특징점 검출 | 1 |
| | 지역 서술자 | 68(194) |
| | K-d 트리 | 1 |
| 입력 영상 | 특징점 검출 | 1 |
| | 지역 서술자 | 66(189) |
| | ANN 초기 정합 | 2 |
| | RANSAC | 6 |
| 합계 | 70/75 | 9/14 |

간상의 건물을 두 대의 인접한 카메라로 촬영한 두 장의 스테레오 영상은 크기, 회전, 밝기 변화가 거의 없고, 약간의 카메라 시점의 차이만 존재한다. 이 영상에서 추출한 특징점을 Tran과 Marchand^[15]가 제안한 PCA 기반 지역 서술자와 본 논문에서 제안한 저니키 모멘트 기반의 서술자를 이용하여 표현하였다. 실제 특징점 정합은 두 방법 모두 동일하게 ANN 방법과 RANSAC 방법을 사용하였다. 이 두 영상에 대하여 PCA 기반의 지역 서술자를 이용하여 특징점 정합을 수행한 결과를 그림 4(a)에 나타내었고, 저니키 모멘트 기반의 지역 서술자를 이용한 특징점 정합 결과를 그림 4(b)에 나타내었다. 두 방법 모두 만족할만한 특징점 정합 결과를 나타내었다. 그러나 두 방법의 연산 시간 비교에서는 큰 차이가 존재하였다.

표 2는 그림 4에 나타난 두 가지 방법의 특징점 정합의 연산 시간을 보여준다. 실험 결과 Tran과 Marchand의 방법은 입력 영상이 들어왔을 때 특징점 정합에 전체 75ms가 걸리는 반면 제안하는 방법은 14ms 밖에 걸



그림 5. 가려짐과 회전이 있는 영상에서 특징점 정합
Fig. 5. Feature point matching with occlusion and rotation.

리지 않는 것을 확인할 수 있다. 이는 Tran과 Marchand의 방법이 각 특징점에서 그래디언트 성분의 주방향 성분을 구하고 회전 연산을 수행하기 때문에 발생하는 차이이다. 이 결과에서 Tran과 Marchand의 방법은 실시간 연산에 사용되기에는 다소 적합하지 않은 반면에 제안하는 방법은 실시간 응용프로그램에 사용하기에 충분함을 확인할 수 있다.

마지막으로 다양한 변환이 가해진 입력 영상에 대하여 제안하는 방법으로 특징점 검출을 수행하였다. 그림 5는 카메라를 고정된 상태에서 객체의 위치, 자세 등을 변환한 영상에 대하여 제안하는 방법으로 특징점 정합을 수행한 결과를 보여준다. 특히 이 실험에서는 영상의 일부가 가려지는 경우에 대해서도 제안하는 특징점 정합 방법이 만족할만한 결과를 보여줌을 확인할 수 있다. 두 가지 경우 모두 2ms 이하의 연산 시간에 특징점 정합이 수행되었고, 이는 제안하는 방법이 실시간으로 동작함을 의미한다.

VI. 결 론

본 논문에서는 저니키 모멘트 기반의 지역 서술자를 이용하여 두 영상의 특징점을 실시간으로 정합하는 방법을 제안하였다. 입력 영상으로부터 FAST 모서리 점 검출 방법으로 특징점을 추출하고, 각 특징점에서 저니키 모멘트 기반의 지역 서술자를 생성하였다. 이 지역 서술자는 영상의 회전과 밝기 변화에 강한 특징을 갖는다. 저니키 모멘트 계산을 실시간으로 수행하기 위하여 고정된 크기의 저니키 기저 함수를 미리 계산하여

록업 테이블에 저장하여 사용하였다. 실제 정합 단계에서는 ANN 알고리즘을 이용하여 초기 정합 결과를 얻고, RANSAC 알고리즘을 이용하여 잘못된 정합을 제거하였다. 다양한 영상에 대하여 실험한 결과 제안하는 방법은 시점이 다른 영상에 대하여 실시간으로 특징점 정합을 수행하는 것을 확인하였다. 기존의 특징점 정합 방법과의 비교 결과, 제안하는 방법이 보다 정확하고 빠르게 특징점 정합을 수행하는 것을 확인하였다. 제안하는 방법은 기존 영상을 등록하는 과정 또한 실시간으로 동작하였다.

참 고 문 헌

- [1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions", *Journal of ACM*, vol. 45, no. 6, pp. 891-923, Nov. 1998.
- [2] C.W. Chong, P. Raveendran, R. Mukundan, "A comparative analysis of algorithms for fast computation of Zernike moments," *Pattern Recognition*, vol. 36, no. 3, pp. 731 - 742, 2003.
- [3] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," *Proceedings of the 25th DAGM Symposium*, pp. 236-243, Sep. 2003.
- [4] M. Fischler and R. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, June 1981.
- [5] C.G. Harris and M.J. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, Manchester, 1988.
- [6] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2000.
- [7] S. Hu, J. Qiao, A. Zhang and Q. Huang, "3D Reconstruction from Image Sequence Taken with a Handheld Camera," *ISPRS Congress Beijing 2008*, pp. 559-562, Beijing China, July 2008
- [8] A. Khotanzad, Y.H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. on Pattern Anal. Mach. Intell.* vol. 12, no. 5, pp. 489-497, 1990.
- [9] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees", *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1465-1479, Sept. 2006.
- [10] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal*

of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[11] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest point," *IEEE International Conference on Computer Vision*, vol. 1, pp. 525-531, 2001.

[12] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," *In IEEE International Conference on Computer Vision*, vol. 2, pp. 1508-1515, Beijing, China, 2005.

[13] S. M. Smith and J. M. Brady, "SUSAN - a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45-78, May 1997.

[14] K. Takaya, "Feature point correspondence of stereo images by monogenic phase," *2007 IEEE*

Pacific Rim Conference on Communications, Computers and Signal Processing PACRIM 07, pp. 272-275, Victoria, B.C. Canada, Aug. 2007.

[15] T. T. H. Tran and E. Marchand, "Real-time key point matching: application to visual servoing", *IEEE International Conference on Robotics and Automation*, pp. 3787-3792, Roma, Italy, April 2007.

[16] 김성호, 김인철, 권인소, "지역 Zernike 모멘트를 이용한 모델 기반 확률적 물체 인식," *제 14회 영상처리 및 이해에 관한 워크샵*, pp. 87-92, 2002.

[17] 김정호, 김성호, 장기정, 권인소, "단일 카메라 전 방향 스테레오 시스템을 이용한 위치 추정 및 환경 지도 작성," *제1회 한국지능로봇 학회 종합 학술대회*, pp. 529-534, 2006.

[18] <http://www.cs.umd.edu/~mount/ANN/>

저 자 소 개



황 선 규(정회원)
 1997년 한양대학교 전자공학과 학사 졸업.
 1999년 한양대학교 전자공학과 석사 졸업.
 2006년 한양대학교 전자통신전파 공학과 박사 졸업.
 2007년~2009년 한양대학교 BK21 연구교수
 2009년~현재 한양대학교 산업과학연구소 연구원
 <주관심분야 : 영상처리, 컴퓨터비전, 패턴인식>



김 회 울(정회원)
 1980년 한양대학교 전자공학과 학사 졸업.
 1983년 Pennsylvania State University 전자공학과 석사 졸업.
 1989년 Purdue University 전자공학과 박사 졸업.
 1989년~1994년 University of Texas 조교수
 1994년~현재 한양대학교 전자통신컴퓨터공학부 정교수
 <주관심분야 : 영상처리, 컴퓨터비전, 패턴인식, 머신비전, MPEG-7>