

효과적인 3차원 객체 인식 및 자세 추정을 위한 외형 및 SIFT 특징 정보 결합 기법[#]

논 문
59-2-34

Combining Shape and SIFT Features for 3-D Object Detection and Pose Estimation

탁 윤 식* · 황 인 준[†]
(Yoon-Sik Tak · Eenjun Hwang)

Abstract - Three dimensional (3-D) object detection and pose estimation from a single view query image has been an important issue in various fields such as medical applications, robot vision, and manufacturing automation. However, most of the existing methods are not appropriate in a real time environment since object detection and pose estimation requires extensive information and computation. In this paper, we present a fast 3-D object detection and pose estimation scheme based on surrounding camera view-changed images of objects. Our scheme has two parts. First, we detect images similar to the query image from the database based on the shape feature, and calculate candidate poses. Second, we perform accurate pose estimation for the candidate poses using the scale invariant feature transform (SIFT) method. We carried out extensive experiments on our prototype system and achieved excellent performance, and we report some of the results.

Key Words : Pose estimation, 3-D object retrieval, Shape-based retrieval, Distance curve, SIFT

1. 서 론

단일 질의 영상 분석을 통한 3차원 객체 검색 및 자세 예측 기법은 로봇 비전, 의학 영상 분석, Unmanned Aerial Vehicles (UAVs) 및 산업 자동화 등의 다양한 영역에서 핵심 기술로써 연구되고 있다. 이렇게 단일 질의 영상을 통한 객체 검색 및 자세 예측을 가능하게 하기 위해서는 객체의 회전에 의해 발생할 수 있는 다양한 각도에서 추출한 영상을 분석하여야 한다. 하지만, 객체의 가능한 모든 영상을 분석하기 위해서는 엄청난 연산 시간을 필요로 하기 때문에 보다 효과적인 객체 검색 및 자세 추정을 위해서 다양한 기법들이 연구되었다. [2]에서 Vacchetti와 Lepetit는 단일 카메라로부터 얻은 영상으로부터 추출한 standard corner 정보를 활용하여 3차원 객체를 트래킹 하는 기법에 대한 연구를 수행하였다. [3]에서는 probabilistic boosting network 기법을 사용한 실시간 객체 검색 및 자세 추정 기법을 제안하였고, [4]에서는 최소 회기 분석 방법을 통한 객체 자세 추정 기법을 제안하였다. 이러한 기법들 외에 3차원 객체 검색 및 자세 추정을 위해서 널리 사용되는 기법은 SIFT[5]이다. [6]에서 Panin와 Knoll은 SIFT 알고리즘을 적용한 3차원 객체 트래킹 기법에 대한 연구를 수행하였으며, [7]에서는 Wagner에 의해 SIFT와 Fern을 사용한 객체 자세 추정

기법이 제안되었다.

이러한 SIFT 알고리즘은 정확한 3차원 객체 검색 및 자세 예측을 가능하게 하기 때문에 로봇 및 자동차 등의 자동 주행과 관련된 많은 응용 연구들에서 사용되고 있다. 하지만, SIFT는 엄청난 양의 연산을 필요로 하기 때문에 방대한 양의 데이터베이스를 사용한 실시간 환경에서는 적용하기가 어렵다. 따라서 SIFT 기반의 기법들은 데이터 셋이 제한된 실내 환경과 같이 제한된 환경에서 적용되거나 정확도의 손상을 최소화시키면서 연산 시간을 줄일 수 있도록 변형되어 이용되고 있다.

이러한 기법들의 문제를 해결하기 위해, 본 논문에서는 외형 정보를 사용한 새로운 3차원 객체 검색 및 자세 추정 기법을 제안한다. 제안된 기법은 영상간의 유사도 비교에 비교적 적은 연산 시간이 걸리는 객체의 외형 정보를 사용하고 외형 정보에 의해 판별될 수 없는 동일 외형을 가진 영상들에 한해 SIFT 알고리즘을 적용함으로써 검색 시간을 확연히 줄일 수 있다. 이러한 외형 기반 기법과 SIFT 기법을 결합함으로써 정확한 객체 검색 및 자세 추정을 실시간에 제공할 수 있게 하였다. 또한 객체의 대칭성에 따른 반복되는 외형 패턴을 인덱스 구조에 제외시킴으로써 보다 더 효과적인 검색을 가능하게 한다.

본 논문에서는 외형 특징 정보로써 영상에 포함된 객체의 외곽선과 중심점을 연결한 거리 곡선을 사용하며, 3차원 객체로부터 영상을 추출하고 외형 정보로써 사용하는 거리 곡선을 계산하는 과정은 그림 1에 나타나 있다. 정확한 거리 곡선을 추출하기 위해서는 [8]과 같이 영상으로부터 객체의 영역을 인식하기 위한 기법이 추가적으로 필요한데, 본 논문의 범위를 벗어나기 때문에 객체를 제외한 영상은 단일 배경색을 가진 것으로 한정하도록 한다.

* 비 회 원 : 고려대학교 전자컴퓨터공학과 석·박사 통합과정

[†] 교신저자, 정회원 : 고려대학교 전기전자전파공학과 부교수

E-mail : ehwang04@korea.ac.kr

접수일자 : 2009년 12월 1일

최종완료 : 2009년 12월 29일

본 논문은 기존 연구인 [1]을 확장 하였습니다.

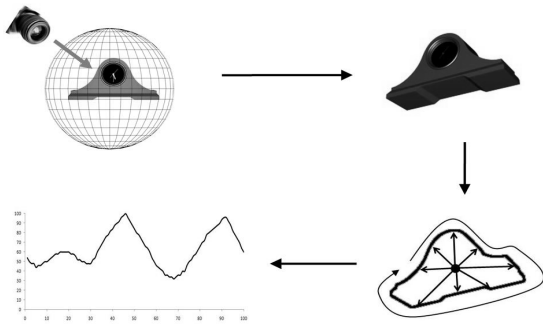


그림 1 3차원 객체로부터 영상 추출 및 외형 정보 (거리곡선) 계산 과정

Fig. 1 Shape feature extraction from 3D objects

본 논문의 구성은 다음과 같다. 2장에서는 본 논문의 배경이 되는 기술들에 대해 설명하며, 3장에서는 객체 검색 및 자세 예측을 위한 검색 기법들에 대해 소개하며, 4장에서는 제안된 기법들의 성능을 평가하기 위한 다양한 실험 결과를 보인다. 마지막으로 5장에서는 본 논문을 결론짓는다.

2. 배경 기술

2.1 GEMINI Framework

거리 곡선 (time series)을 특징 정보로써 사용하는 k-NN 검색에서는 우선순위 큐 기반의 GEMINI framework [9][10]가 순차 검색 방식보다 검색 시간을 확연히 줄일 수 있기 때문에 널리 사용되고 있다.

우선순위 큐 기반의 GEMINI Framework는 비교적 연산 시간이 짧고 하위의 유사도 비교 알고리즘보다 질의 영상과의 차이가 언제나 작거나 같은 값을 가지는 Low Bound 함수들을 계층적으로 설정함으로써 검색 시간을 최소화하는 기법이다. 본 논문에서 사용하는 Low Bound 함수들 및 매칭 알고리즘들은 다음과 같으며, 함수들에 대한 자세한 설명은 [11]에 나타나 있다.

- F_Dist : 퓨리에 계수들간의 유사도
- MINDIST : 인덱스 상의 퓨리에 포인트 (FP) 값들의 경계와 질의 영상의 퓨리에 계수간의 유사도
- LB_K : 두 곡선에 동적 시간 정합 기법을 적용했을 때 발생할 수 있는 점들 간의 매핑 간격을 고려한 동적 시간 정합 기법의 Low Bound 함수
- DTW : 두 곡선간의 유사도를 계산하기 위한 매핑 함수 (동적 시간 정합 기법)

2.2 카메라 영상 최적화 기법

정확한 3차원 객체 검색과 자세 추정을 위해서는 객체의 가능한 모든 각도에서의 영상을 고려하여야 한다. 하지만, 이러한 많은 양의 영상을 고려하기 위해서는 엄청난 양의 저장 공간과 검색 시간이 필요하다. 이러한 문제점을 해결하기 위해서, 본 논문에서는 효과적인 인덱스 구축을 위한 중복되는 외형 패턴을 제거하는 기법을 소개한다.

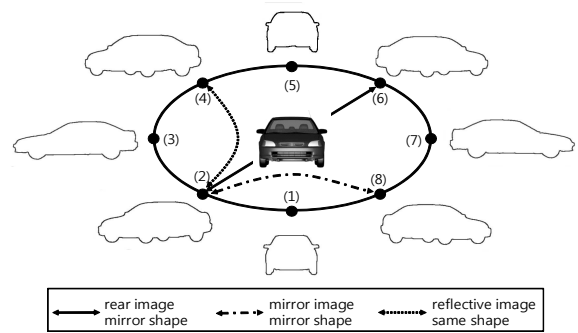


그림 2 객체의 회전에 따른 외형 패턴

Fig. 2 View points and their shape patterns

실생활에 존재하는 대부분의 인위적으로 생성된 객체들은 좌우 혹은 위아래로 대칭인 형태를 가지고 있다. 이러한 대칭 형태에 따라서 외형 패턴이 반복되게 되는데, 그림 2에서는 이러한 반복되는 외형 패턴에 따른 다음과 같은 영상들의 특징을 정할 수 있다. (i) 대칭을 가지는 물체는 거울상 (mirror image)을 가지는 영상은 외형이 좌우 대칭 (mirror shape)을 가진다. (ii) 대칭을 가지는 물체는 외형의 패턴이 반복됨에 따라 동일 외형을 가지는 영상 (reflective image)을 가진다. (iii) 어떠한 객체라도 반대편 영상 (rear image)간에는 좌우 대칭을 가진다.

이러한 특징들에 따라, 본 논문에서는 외형이 중복되는 영상들을 정확도의 손상 없이 인덱스 구조상에서 줄임으로써 검색 시간을 효과적으로 줄일 수 있는 카메라 영상 최적화 기법을 소개한다. 이 기법은 거울상을 가지는 영상들을 하나의 특징 정보로써 인덱스에 포함시키는 거울상 영상 결합 기법과 중복되는 외형 패턴을 가진 영상들을 인덱스에서 제거 시키는 중복 외형 패턴 제거 기법으로 분리된다.

가. 거울상 영상 결합 기법

모든 객체들은 객체를 중심으로 동일한 거리에서 마주볼 때 동일한 외형을 가지지만 좌우가 뒤바뀐 거울상 외형을 가지게 된다. 이러한 거울상 외형을 가진 영상들로부터 특징 정보로써 거리 곡선을 가질 경우, 거리 곡선 또한 모양은 같지만 좌우가 바뀐 형태를 가지게 된다. 이러한 형태의 거리 곡선에 대해서 discrete Fourier transform (DFT)을 적용할 경우, 두 곡선에 대해 동일한 계수 값을 얻을 수 있다. 따라서 이러한 거울상 외형을 가진 영상들로부터 추출한 동일 계수 값을 사용해 인덱스를 구축할 경우, 하나의 계수 값을 사용하더라도 두 개의 영상들을 동시에 접근할 수 있다. 이러한 거울상 영상 결합을 통해 인덱스를 구축할 경우, 검색 정확도에 지장을 주지 않으면서도 인덱스의 크기를 반으로 줄일 수 있다.

나. 중복 외형 패턴 제거 기법

3차원 객체로부터 2차원 영상을 추출하기 위한 촬영 지점들은 중심점으로부터 카메라의 수평 및 수직 이동을 통해 접근할 수 있으므로 객체의 대칭을 수평 및 수직 평면 모두를 고려한다. 먼저, 수평 평면에 대해서 객체의 대칭에 따라 아래와 같이 H1에서 H4까지의 4가지 객체 타입을 정할 수 있다.

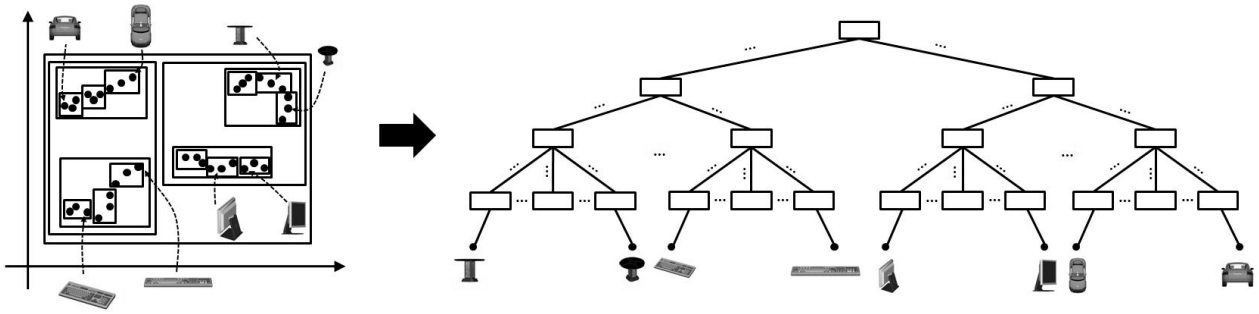


그림 3 3차원 객체로부터 추출한 영상들의 인덱스 구조
 Fig. 3 Index structure of extracted views from 3D objects

- H1: 이 타입은 객체를 수평으로 회전시키며 어느 곳에서 보더라도 동일한 외형을 가지는 객체들을 포함한다. (예: 구)
- H2: 이 타입은 90도 간격으로 외형의 패턴이 반복되는 객체들을 포함한다. (예: 자동차)
- H3: 이 타입은 90도보다 더 짧은 주기로 외형의 패턴이 반복되는 객체들을 포함한다. (예: 주사위)
- H4: 이 타입은 수평으로 객체를 회전시킬 경우, 반복되는 외형의 패턴을 찾을 수 없는 객체들을 포함한다.

수직 평면에 대해서도 수평 평면과 동일한 방식으로 4개의 타입을 정할 수 있으며, 이러한 객체의 타입을 분류하는 방법에 대해서는 [11]에 자세히 설명되어 있다.

수평 및 수직 평면에 대해 정의된 객체의 타입에 따라 중복되는 외형 패턴들 중 하나의 외형 패턴을 제외한 나머지 패턴들은 인덱스를 구축하는 과정에서 고려하지 않음으로써 인덱스의 크기를 확연히 줄일 수 있다. 중복되는 외형 패턴들은 추후에 동일 외형을 가지는 영상들을 고려하여 객체의 자세를 추정함으로써 검색의 정확도에 영향을 주지 않으면서도 검색 시간 및 인덱스 크기를 효과적으로 줄일 수 있다.

3. 검색 프레임워크

이번 장에서는 영상들로부터 추출한 외형 정보 기반의 객체 검색 및 자세 추정을 위한 검색 방법에 대해 설명한다. 이를 위해, 우선 효과적인 검색을 위한 인덱스 구축 기법에 대해 설명한다. 다음으로 구축된 인덱스 구조를 활용한 외형과 SIFT 정보를 혼합한 K-NN 검색과 범위검색 (Range Search) 기법에 대해 설명한다.

3.1 인덱스 구축

임의의 각도에서 촬영한 객체의 영상을 바탕으로 해당 객체 인식 및 자세를 추정하기 위해서는 엄청난 양의 다각도 영상을 검색에서 고려하여야 한다. 따라서 효과적인 검색을 위해서는 적합한 인덱스 구조가 필수적이다. 하지만 본 논문에서 외형 정보로써 사용하는 거리 곡선은 그 값들을 직접 사용할 경우 지나친 인덱스 차원의 발생에 따라 오히려 효율이 떨어지는 “Curse of Dimensionality [12]”가 발생할 수 있다. 따라서, 본 논문에서는 이러한 문제를 위해 거리 곡선의 값들을 Discrete Fourier Transform (DFT)을 통해

변환된 계수 값들 중 상위 값 일부를 사용하여 인덱스를 구축하도록 한다. 본 논문에서 사용하는 인덱스 구조는 그림 3에 나타나 있으며, 구축 과정은 다음과 같다:

1. 각각의 다각도 영상으로부터 거리곡선을 계산한 후, Fourier 계수를 추출한다. 각각의 계수들은 Fourier 공간 속에서 하나의 점 (Fourier Point: FP)을 형성하며, 각각의 FP는 해당 영상을 포함하는 객체의 id (Object id)와 수평 및 수직 타입 (H_Type, V_Type), 해당 영상을 촬영한 좌표 값, 계산된 거리곡선을 포함한다.
2. 형성된 FP들을 객체별로 R-Tree기법을 사용해 트리구조로 구성한다.
 - 객체별로 근접한 FP들을 Minimum Bounding Rectangle (MBR)을 사용해 그룹 짓는다. 이렇게 MBR을 형성할 때 포함된 FP들의 Fourier 계수값들의 최저 값과 최대 값을 정의해 값의 경계를 형성한다.
 - 이렇게 형성된 MBR들을 객체별로 하나의 MBR에 해당 객체의 모든 영상이 포함되도록 반복적으로 상위 MBR로 그룹 짓는다.
3. 각각의 객체별로 구성된 MBR의 트리구조를 하나의 루트 MBR이 모든 객체의 영상을 포함하도록 반복적으로 상위 MBR로 그룹 짓는다.

이러한 방법으로 인덱스를 구축한 후 질의 영상 Q가 입력되었을 경우, 인덱스 구조에 포함된 MBR들의 값의 경계들과 질의 영상으로부터 추출된 Fourier 계수 값들 간의 차이를 계산하며 점차적으로 영상에 접근한다.

3.2 K-NN 검색

가. 외형 기반의 객체 및 후보 자세 검색

기존에 제안되고 개량되었던 GEMINI Framework 기반의 기법들은 단순 이미지 검색에 사용되기 위해 제안되었기 때문에, 본 논문에서와 같이 3차원 객체별로 그룹 지어져 있는 인덱스 구조를 가진 환경에는 적합하지 못하다. 따라서 본 논문에서는 소개한 인덱스 구조를 최대한 활용하기 위해 기존의 기법을 변형한 k-NN 검색 기법을 소개한다. 이러한 Low Bound 함수들을 적용한 변형된 k-NN 검색 기법에 대한 자세한 내용은 Algorithm 1에 기술되어 있다.

Algorithm 1: k-NN Search (Q, k)

```

Variable queue : MinPriorityQueue;
1. queue.push(root); result = {};
2. while not queue.IsEmpty() do
3.   top = queue.Pop();
4.   if top.id is in the result
5.     continue;
6.   else
7.     else if top is a leaf node
8.       for each FP P in top do
9.         queue.push(P, F_Dist(Q,P));
10.    else if top is a Fourier Point P
11.      retrieve its full sequence S;
12.      queue.push(S, LB_K(Q,S));
13.      calculate reverse sequence S' of S
14.      queue.push(S', LB_K(Q,S'));
15.    else if top is a sequence S with LB_K dist
16.      queue.push(S, DTW(Q, S));
17.    if top is a sequence with DTW Dist.
18.      add top to result;
19.      if | result | = k
20.        return result;
21.    else // Non-leaf Node
22.      for each child node C in top
23.        queue.push(C, MINDIST(Q,C));
24. Pose Estimation (result);

```

나. SIFT 알고리즘을 이용한 객체 자세 추정

외형 기반의 객체 검색 기법에서는 추출한 외형 특징 정보가 가장 유사한 영상을 포함하는 객체가 질의 영상의 객체와 가장 유사하다고 판단된다. 하지만, 본 논문에서와 같이 객체의 자세는 다른 각도에서 촬영한 영상이 동일한 외형을 가질 수 있으므로, 외형 특징 정보만으로는 해당 객체의 자세를 판단하기 어렵다. 따라서 본 논문에서는 외형 기반의 가지는 이러한 문제를 해결하기 위해, 동일한 외형을 가지는 영상들에 대해서 정확한 인식이 가능한 SIFT 알고리즘을 추가적으로 적용하도록 한다. SIFT 알고리즘을 사용한 최종적인 객체의 자세 인식 방법은 Algorithm 2에 기술되어 있으며, 전체적인 흐름은 다음과 같다.

1. Algorithm 1에 기술된 k-NN 알고리즘의 결과에 포함된 각각의 영상에 대해 다음의 과정을 수행한다. (Line 2 - 25)
 - 1) 해당 영상과 동일한 외형을 가지는 영상들을 추출한다. 이러한 영상들은 카메라 영상 최적화 기법에 의해 인덱스에서 제거가 되었기 때문에 인덱스를 통한 접근은 불가능하며, 이러한 영상들의 카메라 좌표의 값을 계산한다. (Line 4 - 22)

Algorithm 2: Pose Estimation (result R, Q)

```

1. final_result = {};
2. while not R.IsEmpty() do
3.   extract an element E in R;
4.   set the shape repeating period (HRperiod, VRperiod)
      of E using E.HType and E.VType;
5.   if H_Type is 1
6.     NumofHR = 360 / HRperiod;
7.   else // H_Type 2, 3, 4
8.     NumofHR = 180 / HRperiod;
9.   if V_Type is 1
10.    NumofVR = 360 / VRperiod;
11.  else // V_Type 2, 3, 4
12.    NumofVR = 180 / VRperiod;
13.  for i is 1 to NumofVR
14.    if i is an odd number
15.      Vi = VRperiod * (i+1) - E.v_coordinate;
16.    else // i is an even number
17.      Vi = VRperiod * i + E.v_coordinate;
18.  for j is 1 to NumofHR
19.    if j is an odd number
20.      Hj = HRperiod * (j+1) - E.h_coordinate;
21.    else // j is an even number
22.      Hj = HRperiod * j + E.h_coordinate;
23.  extract views V(Hi,Vj) and insert into VS
      where 1 ≤ i ≤ NumofHR, 1 ≤ j ≤ NumofVR
24.  match V in VS with Q using SIFT;
25.  insert best matched V in VS to final_result;
26. return final_result;

```

- 해당 영상에 포함된 객체의 타입에 따라 수평 및 수직 평면으로 반복되는 외형의 패턴 주기를 계산한다. (Ex. H1 : 1°, H2 : 90°, etc.)
 - 반대편에서 바라보는 거울 상 영상들의 외형이 동일한 H1 및 V1 타입은 360°, 나머지는 180°를 기준으로 수평 및 수직으로 반복되는 외형 패턴의 수를 계산한다. (NumofHR, NumofVR)
 - 각각의 반복되는 외형 패턴별로 하나의 동일 외형 영상이 존재하므로, 각각의 외형 패턴에 위치하는 동일 외형 영상의 카메라 위치를 계산한다.
- 2) 1)에서 계산된 수평 및 수직 카메라 지점에 위치하는 영상들을 추출한다. (Line 23)
 - 3) 추출된 영상들과 질의 영상간의 유사도를 SIFT 기법을 적용하여 계산하고, 가장 유사한 영상을 해당 영상에 포함된 객체의 자세로 추정한다. (Line 24 - 25)
2. 결과에 포함된 모든 영상들에 대해서 자세 추정이 끝난 후, 계산된 자세들을 최종 결과로써 제공한다. (Line 26)

Algorithm 3: Range Search (N, Q, e)

1. **if** N.id is in the result
2. **skip** node N
3. **else**
4. **if** N is a non-leaf node
5. **for** each child node C of N
6. **if** MINDIST(Q, C) \leq e
7. Range Search(C, Q, e);
8. **else** // N is a leaf node
9. **for** each Fourier points C in N
10. **if** F_Dist(Q, C) \leq e
11. **retrieve** full sequence C from DB;
12. **if** LB_K(Q, C) \leq e && DTW(Q, C) \leq e
13. **add** C to the result;
14. **else**
15. **reverse** a full sequence C to C'
16. **if** LB_K(Q, C') \leq e && DTW(Q, C') \leq e
17. **add** C' to the result;

3.3 범위 검색 (Range Search)

범위 검색은 k-NN 검색과 더불어 영상 검색 시스템에서 널리 쓰이는 검색 기법이며, k개의 유사한 영상을 추출하는 k-NN 알고리즘과는 달리, 일정 범위 (e)에 포함되는 모든 영상들을 검색하는 기법이다. 이러한 일반적인 영상 검색에서의 범위 검색 기법과는 달리, 본 논문에서 소개하는 범위 검색 기법은 일정 범위(e)에 포함되는 모든 3차원 객체들을 해당 객체의 자세로 추정되는 영상과 함께 검색한다. 소개된 범위 검색 기법은 k-NN 알고리즘과 마찬가지로 사용된 인덱스 구조를 최대한 활용할 수 있도록 기존의 GEMINI framework 기반의 범위 검색 기법을 변경하여 사용하도록 한다. 자세한 과정은 Algorithm 3에 나타나 있으며, 사용된 Low bound 및 매칭 알고리즘들은 k-NN 알고리즘과 동일하다. 범위 검색 기법은 초기 값으로써 인덱스의 루트노드 R, 질의 영상 Q, 주어진 범위 e를 입력받아 Range Search (R, Q, e)를 실행함으로써 초기화 된다. 전체적인 범위 검색의 동작 과정은 다음과 같다:

1. k-NN 검색과 마찬가지로 검색의 결과로써 하나의 객체당 하나의 자세를 포함하는 영상만이 필요하다. 따라서 만약 현재 노드의 id가 결과에 포함되어 있을 경우, 현재의 노드를 더 이상 확장하지 않고 해당 노드에 대해서 검색을 중단한다. (Line 1-2)
2. 만약 현재 노드가 중간노드일 경우, 해당 노드에 포함된 자식 노드들 (MBRs)과 질의 영상간의 유사도를 MINDIST를 사용해서 계산한 후, 계산된 값이 정해진 범위 (e) 이내인 자식 노드들을 계속해서 확장하며 검색한다. (Line 4-7)

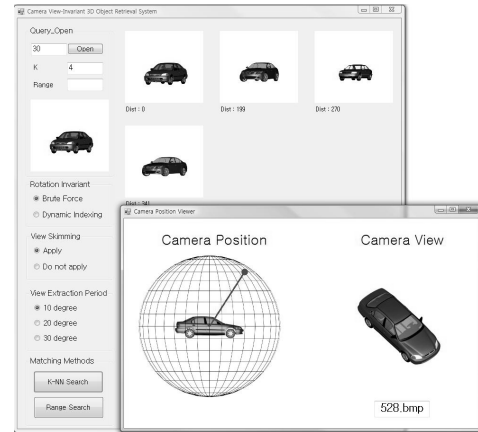


그림 4 프로토타입 시스템

Fig. 4 Prototype System

3. 만약 현재 노드가 단말 노드일 경우, 해당 노드에 포함된 FP들에 대해서 다음과 같은 과정을 수행한다. (Line 8-17)
 - 1) 만약 FP의 계수 값과 질의 영상의 Fourier 계수간의 차이가 정해진 범위 (e) 이내인 경우, 해당 FP의 거리 곡선을 추출하고 2)를 실행한다.
 - 2) 만약 추출된 거리 곡선과 질의 영상으로부터 계산된 거리 곡선간의 LB_K의 값이 정해진 범위 (e) 이내이고 두 곡선간의 DTW 값 또한 (e) 이내인 경우 해당 영상의 정보를 result에 넣는다. 만약 두 값 모두 (e) 이내인 경우가 아니면 3)을 실행한다.
 - 3) 해당 영상의 맞은편에 위치한 거울 상 영상을 고려하기 위해 거리 곡선의 순서를 거꾸로 하여 계산된 거울 상 거리 곡선을 생성한다. 만약 거울상 거리 곡선과 질의 영상으로부터 계산된 거리 곡선간의 LB_K의 값이 정해진 범위 (e)이내이고 두 곡선간의 DTW 값 또한 (e)이내인 경우 해당 영상의 정보를 result에 넣는다.

4. 실험

제안된 기법들의 성능을 평가하기 위해, 본 논문에서는 3차원 객체 검색과 자세 예측을 위한 프로토타입 시스템을 구축하고 다양한 실험을 수행하였다. 3차원 객체들로부터 추출한 14,000 여개의 다각도 영상을 데이터셋으로 활용하였으며, 4GB 메모리를 가진 Intel Core2Duo CPU를 탑재한 시스템을 사용하였다. 본 논문에서 제안된 기법들은 모두 C# 언어를 사용하여 구현되었다.

그림 4에서는 구현된 프로토타입 시스템의 인터페이스를 보여주고 있다. 질의가 주어졌을 때, 사용자는 K-NN 알고리즘과 범위 검색 알고리즘을 선택할 수 있으며, 영상의 회전을 고려하기 위한 기법으로 회전이 가능한 모든 형태를 고려하는 "Brute Force" 알고리즘과 [13]에서 제안한 "Dynamic Indexing" 기법을 선택할 수 있다. 주어진 질의 영상에 대해서 시스템은 가장 유사한 객체들의 자세가 포함된 영상들의 리스트를 반환하며 검색창에 디스플레이 된다. 또한, 사용자는 해당 영상에 포함된 객체의 다른 다각도 영상들과 함께 해당 영상의 시점을 확인해 볼 수 있다.

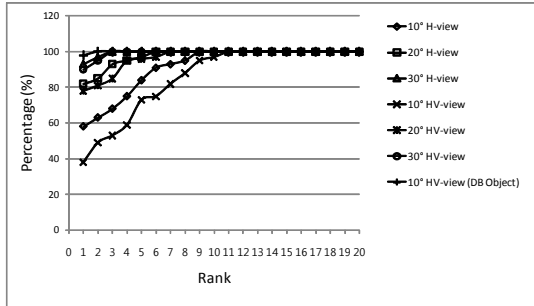


그림 5 정확도 계산을 위한 CMC
Fig. 5 CMC curve for camera angle invariance

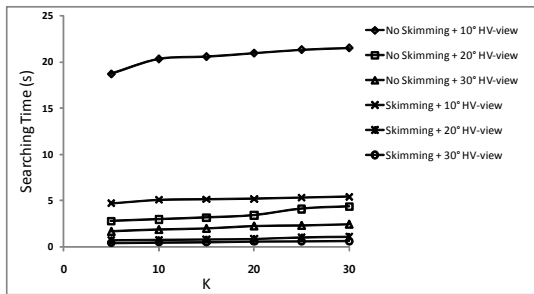


그림 6 K-NN 검색의 실행 시간
Fig. 6 Search time for K-NN

Experiment 1: 데이터베이스에 포함된 3차원 객체와 동일한 실제 객체로부터 (DB Object) 인덱스 구조와 동일한 영상 추출 간격을 가진 임의의 각도에서 추출한 영상을 질의로 하였을 때 제안된 기법의 정확도를 계산하였다. 그림 5에서는 이러한 정확도를 Cumulative Match Curves (CMC)를 사용하여 나타내었다. 실험을 통해, 모델링된 3차원 객체와 동일한 실제 객체로부터 인덱스 구조와 동일한 영상 추출 간격을 가진 임의의 각도에서 질의 영상을 추출하더라도 동일한 3차원 모델 객체로부터 10도 간격으로 수평, 수직 이동을 고려한 영상들을 추출하여 데이터베이스를 구축할 경우, 대부분의 경우에서 완벽하게 해당 객체를 자세와 함께 찾을 수 있음을 알 수 있다.

Experiment 2: 1번 실험에서보다 복잡한 상황에서의 시스템의 성능을 평가하였다. 앞의 경우에서와 달리, 이번 실험에서는 데이터베이스에 모델링되어 저장된 3차원 객체들과 유사하며 개념적으로는 같지만 (예: SUV 자동차), 정확하게 일치하지 않은 실제 객체들을 인덱스 구조와 동일한 영상 추출 간격을 가진 임의로 선택된 각도에서 촬영하여 질의 영상으로 사용하였다. 객체로부터 다각도 영상을 추출하는 간격을 10에서 30도로 변화를 주며 검색 정확도의 변화를 측정하였다. 이 경우, 객체의 수평 회전만을 고려할 경우 (H-View) 10에서 30도의 영상 추출 간격을 가지더라도 비교적 정확한 검색이 가능하였다. 하지만 수평과 수직 회전 모두를 고려할 경우 (HV-view), 10도의 영상 추출 간격을 줄 경우 심각한 정확도의 저하가 발생하였고, 30도까지 영상 추출 간격을 늘릴수록 정확도가 올라가는 것을 볼 수 있다. 이는 데이터베이스의 객체와 질의 영상속의 객체가 완전히 같지 않을 경우, 해당 객체의 자세를 추정하는 것은

가능하지만 해당 객체의 자세한 자세를 추정하기에는 부적합함을 알 수 있다.

Experiment 3: 본 논문에서 소개한 K-NN 검색 알고리즘을 중복 패턴 제거 기법의 적용 유무와 영상 추출 간격의 변화에 따른 검색 시간을 측정하였다. 그림 6에서는 이러한 다양한 경우에서의 K개의 가장 가까운 영상들을 추출하는데 걸리는 시간을 그래프로써 나타내고 있다. 그래프에 나타나 있듯이, 객체의 자세한 자세를 측정하기 위해 영상 추출 간격을 늘려감에 따라 폭발적으로 검색시간이 증가함을 알 수 있다. 하지만, 소개된 중복 패턴 제거 기법이 모든 영상 추출 간격에서 효과적으로 검색시간을 줄일 수 있기 때문에 좁은 영상 추출 간격을 사용하더라도 효과적으로 객체 및 자세 검색이 가능함을 알 수 있다. 전체적인 실험에서, 통일되게 이미지 회전을 보정하기 위해, Dynamic Indexing 기법[13]을 적용하였다.

5. 결 론

단일 영상 질의를 사용한 보다 빠르고 정확한 3차원 객체 검색 및 자세 예측을 위해, 본 논문에서는 외형 정보 및 SIFT 알고리즘을 사용한 새로운 검색 기법을 제안하였다. 정확하면서도 실시간 성능을 보장하기 위해, 제안된 기법은 크게 두 개의 부분으로 나눌 수 있는데, 1) 외형기반의 유사 객체 검색 및 후보 자세 검색과 2) 느리지만 정확한 SIFT 기법을 활용하여 동일 외형 영상들 중 최종 객체의 자세를 추정하는 부분이다. 또한 보다 빠른 검색을 지원하면서도 정확도에 손상을 주지 않기 위해, 본 논문에서는 거울상 외형을 가지는 영상들을 하나의 계수로써 결합시키고 반복되는 외형 패턴을 제거하는 기법인 카메라 영상 최적화 기법을 적용하였다. 실험을 통해, 제안된 기법은 저장된 객체의 질의에 대해서 대부분의 경우 정확한 결과를 제공함을 보였다. 또한 데이터베이스에 모델링된 3차원 객체와 SUV 자동차와 같이 객체 분류만 같은 실제 객체에 대해서도 대략적인 객체의 자세를 추정하기에는 적합한 정확도를 가질 수 있음을 보였다. 마지막으로 속도가 느린 SIFT 알고리즘을 동일 외형 영상에만 적용하고 카메라 영상 최적화 기법을 통한 인덱스 구축으로 빠른 시간 내에 결과를 얻을 수 있음을 보였다.

감사의 글

본 연구는 2008년도 정부(교육인적자원부)의 재원으로 한국 학술진흥재단의 지원을 받아 수행된 연구 (KRF-2008-313-D00858)로서, 관계부처에 감사 드립니다.

참 고 문 헌

[1] Y. Tak and E. Hwang, "Indexing and Matching Scheme for Recognizing 3D Objects from Single 2D Image," International Conference on Internet and Multimedia Systems and Applications (IMSA'09), pp.

60-67, 2009.

[2] L. Vacchetti, V. Lepetit and P. Fua, "Stable real-time 3D tracking using online and offline information," IEEE Transactions on PAMI, vol. 26, no. 10, pp. 1385-1391, 2004

[3] J. Zhang, S.K. Zhou, L. McMilan and D. Comaniciu, "Joint Real-time Object Detection and Pose Estimation Using Probabilistic Boosting Network," CVPR'07, pp.1-8, 2007

[4] M. Saito and K. Kitaguchi, "Appearance based object pose estimation using regression models," SICE Annual Conference, pp.1926-1929, 2008

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," IJCV, vol. 60, no. 2, pp. 91-110, Nov. 2004

[6] G. Panin and A. Knoll, "Fully Automatic Real-Time 3D Object Tracking using Active Contour and Appearance Models," Journal of Multimedia, vol. 1, no. 7, pp. 62-70, 2006.

[7] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond and D. Schmalstieg, "Pose Tracking from Natural Features on Mobile Phones," IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 125-134, 2008.

[8] Y. Nam, E. Hwang and D. Kim, "A similarity-based leaf image retrieval scheme: Joining shape and venation features," Computer Vision and Image Understanding, vol. 110, no. 2, pp.245-259, 2008.

[9] E. Keogh and C. Ratanamahatana, "Exact indexing of dynamic time warping," Knowledge and Information Systems, Vol. 7, pp. 358-386, 2005.

[10] C. Faloutsos and K. Lin, "FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets," ACM SIGMOD, pp.163-174. 1995.

[11] H. Kim, Y. Tak and E. Hwang, "A Shape-Based Indexing Scheme for Camera View-Invariant 3-D Object Retrieval," Multimedia Tools and Applications, Accepted.

[12] C. Faloutsos, M. Ranganathan and Y. Manolopoulos, "Fast subsequence matching in time-series databases," ACM SIGMOD, pp.419-429, 1993

[13] Y. Tak and E. Hwang, "An indexing scheme for efficient camera angle invariant image retrieval," International Conference on Computer and Information Technology, pp.143-148, 2008

저 자 소 개



탁 윤 식 (卓 允 植)

2005 : 동국대학교 컴퓨터공학과 졸업 (공학사)

2005~현재: 고려대학교 대학원 전자컴퓨터공학과 (석·박사통합과정)
주관심분야

데이터베이스, 멀티미디어 시스템, 타임 시리즈, 영상 인덱스 및 검색 프레임워크 개발
E-mail : life993@korea.ac.kr



황 인 준 (黃 仁 俊)

1988 : 서울대학교 컴퓨터공학과 졸업 (공학사)

1990 : 서울대학교 컴퓨터공학과 (공학석사)

1998 : Univ. of Maryland at College Park 전산학과 (공학박사)

1998~1999: Bowie State Univ., 조교수
1999~1999: Hughes Research Lab. 연구교수
1999~2003: 아주대학교 정보통신전문대학원 조교수
2003~2004: 아주대학교 정보통신전문대학원 부교수
2004~2007: 고려대학교 전기전자전파공학과 조교수
2008~현재: 고려대학교 전기전자전파공학과 부교수
주관심분야
데이터베이스, 멀티미디어 검색, 정보통합, 전자상거래, 영상처리, 유비쿼터스 컴퓨팅
E-mail : ehwang04@korea.ac.kr