Chapter 9. 인식

1. 도전

컴퓨터 비전에서 도적적인 연구 주제는 인식으로, 여러 대회가 주기적으로 열린다.

2. 사례 인식

사례는 특정한 물체를 뜻하는데 과거와 달리 최근에는 물체의 모양에 제한을 두지 않고 여러 환경을 고려한다.

초창기 사레인식에는 기하학적 접근 방식이 유용했다. 물체 모델을 3차원공간으로 표현하고 영상에서 물체 분할을 하여 모델과 매칭하여 인식을 수행했다. 여러가지 가정과 함께 일반 실린더, 양상 그래프등 물체 표현 방법이 제안되었지만 문제를 잘 해결하지 못하였다. 기하 정렬을 사용한 알고리즘은 모델 구축 단계와 인식 단계로 구성된다. 모델 구축 단게는 관심점을 검출, 저장하고 전처리를 적용한다. 그리고 관심점 각각에 대해 특징 벡터를 추출하고 부류 정보와 저장한다. 인식 단께는 영상이 입력되면 물체의 정보인 기하 변환 행렬을 계산하여 출력하고 물체 각각에 대해 인식을 시도한다. 기하 정렬된 쌍의 집합의 품질이 임계값을 넘으면 물체를 인식한 것으로 한다.

기하 정렬은 계산이 너무 오래 걸리기 때문에 단어 가방 기법이 고안되었다. 정보 검색은 매우 빠르게 가능한데 이는 문서를 빈도 벡터로 변환하여 간략화했기 때문에 가능한 것이다. 문서간 유사성을 측정하기 위해 단어 빈도 벡터(tf벡터)를 가공하여 tf-idf벡터를 이용한다. 이 벡터들의 유사도는 코사인 유사도로 측정한다. 영상에서 이 정보 검색이론을 적용하면 특징 공간을 벡터양자화로 구간으로 나누어 특징 공간에 가까이 있는 특징 벡터들을 모아 그들의 대표를 단어로 삼게 된다. 이 단어를 시각 단어라 부른다. 정보 검색과 마찬가지로 유사도를 계산하게 된다.

3. 범주 인식

사례인식과 달리 범주인식은 부류 내 변화가 매우 심해서 힘든 문제이다. 범주 인식은 배경을 모두 지우고 해당 물체 영역만 주는 강한 지도가 있고, 어떤 범주의 물체가 존재하는지 주석만 다는 약한 지도가 있는데 최근엔 약한 지도를 추구하는 추세이다.

사례인식은 매칭 연산을 사용했지만 범주 인식은 분류 연산을 사용하게 된다. 단어 가방을 이용한 범주 인식의 처리 과정은 특징 검출, 벡터 양자화, tf-idf벡터 계산으로 특징 벡터를 추출하고 기계 학습을 적용하여 학습을 수행해 분류 과정을 거쳐 인식을 하게 된다. 단어 가방은 물체 영역과 배경 영역을 구분하지 않는다.

부품 모델 기법은 단어 가방과 달리 물체를 구성하는 부품과 그들의 위치 관계를 고려한다. 사람의 얼굴은 다양한 변화를 가질 수 있지만 부품의 외관과 부품들의 위치 관계인 모양은 일정한 범위에 있다. 이를 이용해 별자리 모델을 설계할 수 있다. 하지만 계산 시간이 너무 오래 걸리고 모양에 해당하는 특징을 무엇으로 할지가 문제이다.

최근에는 컨볼루션 신경망을 이용한 범주 인식 문제가 제일 활발하게 이루어 지고 있다.

4. 사람 인식

얼굴 인식은 사람 인식중 가장 많은 연구가 이루어진 분야이고 이를 통해 나이 인식, 성별 인식, 인종 인식등을 통해 고객 관리나 표적 광고에 활용할 수 있다.

5. 모바일 기기에 적용된 인식 기술

최근 모바일 기기의 발전으로 얼굴 인식을 통한 잠금해제, leafsnap, 증강 현실 등이 개발되었다.

Chapter 10. 모션

1. 움직이는 상황

동적 비전이 처리하는 영상은 연속 영상(동영상)이며 장면에 담긴 물체에 대한 정보를 알아내야 한다. 영상은 공간 일관성을 가지며 연속 영상은 시관 일관성도 갖는다. 이 두 일관성을 활용해 정보를 알아낸다.

차 영상은 이전 프레임과의 차를 구하는 것이다. 움직임이 큰 부분이 차이가 발생할 수밖에 없다는 원리를 이용한다. 하지만 배경과 물체의 색상이나 명암의 큰 변화가 없어야 한다.

모션 필드는 차 영상의 단점을 극복한 기법이다. 연속된 두장의 영상으로부터 움직임이 발생한 모든 점의 2차원 모션벡터를 구해 모션 필드를 얻는다. 모션 필드를 알 수 있으면 목표물 추적, 제스처 인식등의 문제를 해결할 수 있지만 벡터를 추정하는 일은 쉽지 않다.

2. 광류

광류는 두 영상의 명암 변화를 통해 움직임 정보를 얻는다. 조명의 변화가 없고, 각에 따라 명암의 차이가 있다는 사실을 무시한다고 가정하면 연속한 두 영상에 나타난 물체의 같은 점은 명암값이 같거나 비슷하다는 밝기 항상성을 이용한다. 인접한 두 영상은 광류 조건식(그레이디언트 조건식)을 만족한다. Lucas-Kanade 알고리즘과 Horn-Schunck 알고리즘은 각각 가정을 추가하여 지역적 방법, 전연적 방법으로 모션벡터를 확정 짓는다. LK는 지역적이기 때문에 명안 변화가 적은 곳에서 광류값이 정해지지 않는 영역이 발생할 수 있다는 단점과, HS는 물체 경게에 해당하는 불연속점이 불분명해지는 단점이 있기 때문에 두 알고리즘을 결합한 새로운 광류 추정 알고리즘이 쓰인다. 광류는 로봇 항해와 독화술같이 다양한 응용 문제를 푸는데 활용되고 있다.

3. 물체 추적

광류 맵은 물체의 궤적에 관한 정보이고 명시적이지 않고 암시적이다. 따라서 물체 추적을 위해선 광류맵에서 추출해낸 특징을 정보로 변환하는 과정이 필요하다. KLT 추적 알고리즘은 특징점을 찾기 위해서 H행렬의 고유값을 이용하여 안정적인 특징 점을 찾아낸다. 특징점을 검출한 뒤부턴 그것들을 추적하고, 영상 밖으로 가거나 다른 물체에 가려진 경우 H행렬 고유값을 이용하여 실종 처리한다.

광류 맵 추정 알고리즘은 큰 이동이 일어나는 영역에서 모션 벡터를 추정하지 못한다. 이 문제는 큰 이동을 찾아내는 대응점 찾기를 통해 희소한 모션 벡터를 생성한 후 밀집된 광류 맵을 알아내는 HS 알고리즘을 적용하여 해결될 수 있다.

Chapter 9. 3차원 비전

1. 본질 영상

인간은 외관과 본질을 구별하는 능력이 있지만 컴퓨터는 외관 정보만으로 영상을 분할한다. 본질 영상을 복원하기 위한 시도는 많이 있었다. 실제 영상이 아닌 단순화된 영상이며 화소의 컬러값을 조명값과 반사율의 곱이라고 가정하고 또 조명 영상은 서서히 변하고 급격한 변화는 없다고 가정을 할 경우 본질 영상 복원이 가능하다. 또 조명만 서서히 변화시키며 여러 장의 컬러 영상을 획득한 경우 그림자를 제거하는 영상을 얻을 수도 있다.

2. 스테레오

스테레오 비전은 사람의 눈과 같은 원리를 사용해 깊이를 알아내는 것이다. 사람의 눈처럼 두 대의 카메라를 사용하는 스테레오 비전의 원리를 사용하면 물체를 한 점으로 특정지어 거리를 알 수 있다. 하지만 대응점을 찾아야 하는데 스테레오는 잘못된 매칭인 거짓 긍정을 가려낼 수 없고, 매칭이 발생한 점에서만 깊이 정보를 알아낼 수 있어 희소한 깊이 영상을 얻는다는 문제점이 있다. 에피폴라 기하를 이용하면 대응점을 찾을 수 있다.

희소하지 않은 밀집된 깊이 영상을 얻기 위해서 모든 화소에 대해 에피폴라 선을 이용하여 대응점을 찾을 수 있지만 모든 화소가 개별적으로 대응점을 결정하기 때문에 틀리게 매칭될 수 있다. 공간 일관성을 적용하기 위해 변위 공간 영상을 만들어 전역 탐색 방법을 사용해 문제를 해결한다.

3. 능동 센서

카메라에 들어오는 정보만 수동적으로 받아들이는 것이 아닌 능동적으로 투사하여 되돌아오는 신호로부터 거리를 알아내는 것이 능동 센서이다.

구조 광 기법은 스테레오 기법에서 하나의 카메라를 프로젝터로 교체하면 된다. 구조 광 기법도 대응점을 찾아야 하는데 2차원의 격자 무늬를 사용해서 대응 점을 쉽게 얻을 수 있다. 그리고 카메라와 물체의 상의 각, 프로젝터와 물체의 상의 각은 캘리브레이션 작업을 통해 알 수 있다. 키넥트가 구조광을 이용한 제품이고 이것은 사람 자세 추정, 행위 인식, 손 제스처 인식에 사용된다,

4. 깊이 영상의 인식

컬러 영상에서 사람의 자세를 추정하는 것은 매우 힘들다. 사람의 모양과 외관은 심히 변하고 조명도 변하기 때문이다.

깊이 영상은 옷 과 조명에 불변한 영상을 제공하고 3차원 정보를 가지기 때문에 더 좋은 성능을 보인다. 차 영상을 이용하여 사람을 검출하고 컬러 영상을 이용하여 추적하는 컬러와 깊이 영상을 결합한 방법도 있다.

깊이 영상을 임의 숲 분류기를 이용하여 화소를 분류하고, 분류된 구성요소를 군집화하여 군집의 대표를 골격의 연결점으로 사용하여 연결하면 자세 정보를 얻을 수 있다.

Chapter 12. 장면 이해

1. 동물은 어떻게 보나?

컴퓨터 비전 연구는 사람의 시각을 크게 고려하지 않고 진행되어왔다. 사람의 두뇌에서 시각 정보가 처리되는 과정에 비해 컴퓨터 비전 알고리즘은 심하게 간단했기 때문에 계층 처리 구조, 정보 채널의 분리 처리, 처리 과정이 정보를 주고 받는 피드백, 고정된 처리와 학습의 균형 이렇게 네가지 원칙을 적용해야 한다고 한다.

2. 이해로 가는 길

장면 속 물체의 인식을 넘어선 장면 이해는 사람의 시각 기능을 모방하여 연구되고 있다. 사람은 장면을 볼 때 주어진 임무에 따라 선택적 주의 집중을 한다. 선택적 주의 집중은 단속성 운동과 고착 현상으로 나타나고 이는 현저성 맵을 통해 구현된다. 컴퓨터 비전에서도 사람이 문맥을 사용하는 것만큼 중요하지만 현재는 연구가 매우 느리게 진행되고 있다. 언어 처리에서 파싱은 주로 사용되고, 컴퓨터 비전에서도 사용된다. 영상 파싱은 영상을 구성요소로 나누어 각 요소에 정보를 부여하고, 요소 간의 위치 관계와 소속 관계를 계층적으로 표현한다.