

개선된 k-Nearest Neighbor를 이용한 물체 인식

정병수^o, 위승정^o, 김종혁^o, 김병기^{*}

^o 남부대학교 디지털정보학과, ^{*} 전남대학교 전산학과

bsjung@nabu.ac.kr

Object Recognition using Improved k-Nearest Neighbor

Byeongsoo Jung^o, Seungjung Wi^o, Jonghyeuk Kim^o

^oDept. of Digital Information Nambu Univ.,

^{*}Dept. of Computer Science Chonnam National Univ.

요약

모델 영상내의 각각의 물체의 대표 값을 만든 후에 실험 영상을 물체 공간에 투영 시켜서 나온 성분과 대표 값의 거리를 비교하여 인식하게 된다. 그러나 단순히 기존의 방법인 Point to Point 방식인 단순 거리 계산은 오차가 많기 때문에 본 논문에서는 개선된 Class to Class 방식인 k-Nearest Neighbor를 이용하여 몇 개의 연속적인 입력영상에 대해 각각의 모델영상들을 인식의 단위로 이용하였다.[1], [2]

1. 서론

모델 기반형 물체 인식을 하기위해 본 논문에서는 물체에 대해 5°씩 회전을 시켜서 물체 영상을 찍은 후에 그 사이에 들어온 입력 영상에 대해서는 그 주변의 학습 영상들을 이용해서 그룹단위로 인식을 수행한다.[2] 주성분 분석을 이용한 인식은 픽셀 값들의 분포를 이용하기 때문에 물체의 회전(Rotation), 평행 이동(Translation), 크기 변화(Scale), 조명 변화에 대해서 매우 민감하다. 그러므로 많은 학습 이미지가 필요하게 되고 영상의 크기를 모두 일정하게 맞추는 정규화 과정이 필요하게 된다. 주성분 분석은 조명 변화에 아주 민감하기 때문에 본 논문에서는 히스토그램 평활화를 이용하여 입력 영상을 정규화 해줌으로서 조명 변화에서도 일정한 인식률을 유지하게 되었다.[3]

2. 물체 영상 공간에서 상관관계와 거리

한 물체 영상 당 5°씩 회전을 해서 물체 영상을 습득하였다. 즉 한 물체는 72개의 이미지를 갖게 된다

$$C = \left(\frac{1}{N} \right) \sum_{i=1}^N x_i \quad \text{식(1)}$$

평균영상은 식(1)로 구하고 특이치 분해를 이용하여 고유 공간을 구하였다. 특이치 분해를 이용하여 식(2)에서처럼 행렬이 분해되는데, 식(3)처럼 영상 집합 X의 공분산 행렬에 대한 고유 벡터를 쉽게 얻을 수 있다.[4]

$$\begin{bmatrix} X \end{bmatrix} = \begin{bmatrix} V \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & & \\ & \dots & \\ & & \sigma_i \end{bmatrix} \cdot \begin{bmatrix} U^T \end{bmatrix} \quad \text{식(2)}$$

$$X = V \Sigma U^T \quad \text{식(3)}$$

식(4)에 의해 모든 고유벡터를 고유 공간 구성에 이용하지 않고 많은 영상을 대표할 수 있는 주성분의 벡터만을 이용할 수 있다.

$$\left(\sum_{i=1}^k \lambda_i\right) / \left(\sum_{i=1}^N \lambda_i\right) \geq T_1 \quad \text{식(4)}$$

여기서 T_1 은 고유벡터의 개수를 조정하는 문턱치이며, 인식과 물체 영상 평가 시 이용되는 고유벡터 $\{e_i | i = 1, 2, 3, \dots, k\}$ 는 저차원 공간을 구성하기 위해 ($k \ll N$)을 만족하며 실험에서는 $k=8$ 을 이용하였다. 이렇게 구성된 공간을 앞으론 물체 영상 공간으로 표현한다.

앞에서 얻어진 물체 영상 공간에서 평균 영상 C 에서 n 영상 집합 X 를 모두 식(5)을 이용하여 투영시킨다. 식(6)을 통해 영상을 복원 될 수 있다.

$$f_j = [e_1, e_2, e_3, \dots, e_k]^T (x_n - C) \quad \text{식(5)}$$

$$x_1 = \sum_{i=1}^N m_{1i} e_i + C \quad \text{식(6)}$$

$$x_1 \approx \sum_{i=1}^N m_{1i} e_i + C \quad \text{식(7)}$$

여기서 식(7)은 k 개의 물체영상공간에 투영한 물체영상을 표현하고, 두 영상간의 유사도는 식(8)로 측정한다.

$$\begin{aligned} \|x_1 - x_2\|^2 &= (x_1 - x_2)^T (x_1 - x_2) \\ &= 2 - 2x_1^T x_2 \end{aligned} \quad \text{식(8)}$$

$$\|x_1 - x_2\|^2 \approx \|m_1 - m_2\|^2 \quad \text{식(9)}$$

따라서 두 점간의 관계는 두 물체 영상간의 관계와 매우 유사하다는 것을 알 수 있다. 고유 공간에 투영된 점이 가까울수록 영상들의 높은 상관관계를 가진다는 것이 증명 되었다.

3. 개선된 k-Nearest Neighbor를 이용한 물체 인식

투영된 입력 영상이 모델 영상과의 매칭의 경우 (Point to Point) 여러 포즈들이 투영되어 있는 물체 영상 공간에서 실제 물체 영상은 매칭에 성공했는지라도 다른 물체 영상으로 인식하는 잘못된 매칭이 발생한다.[5], [6] 이러한 문제를 해결하기 위해 단일 물체 영상 단위의 매칭이 아닌 몇 개의 연속적인 입력 영상에 대해 각각 몇 개의 모델 영상들을 인식의 단위로 이용하였다(Class to Class).

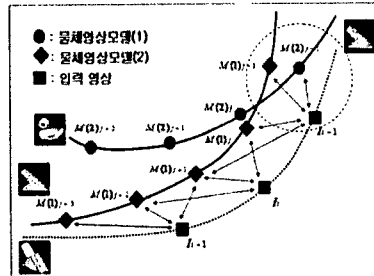
$$w = \frac{\text{Max}(\arg S(M_j)) - \text{Min}(\arg S(M_j))}{d(k-1)} \quad \text{식(10)}$$

매칭 알고리즘은 식(10), 식(11)에서 나타낸 것처럼

k -Nearest Neighbor 방법을 이용하였다. 여기서 $\arg S(M_j) = j$ 는 모델의 번호를 구하는 연산자이다.

$$\frac{\sum \sum w(I_i - M_j)}{k} \quad \text{식(11)}$$

위의 식(11)에 의해 구해진 값을 이용하여 모델 영상과 입력 영상의 인식을 결정한다. 여기서는 $k=3$ 을 이용하였다. [그림 1]은 공간상에 투영된 연속적인 입력 영상들과 모델 영상들은 이 방법을 이용한 매칭 방법을 나타낸 것이다. [그림 1]에서 나타나듯이 입력 영상과 모델 영상 간 가장 가까운 거리를 가지는 영상일지라도 실제 다른 물체 영상일 경우가 있다는 것을 보여준다. 이러한 문제를 해결하기 위해 본 논문에서 두 점간의 거리로만 판단하지 않고 연속적인 몇 개의 영상 단위로 매칭을 수행함으로써 물체 영상 전체인 케적간의 일치성까지 알 수 있으며, 이를 이용하여 얻어진 결과는 물체 영상 전체를 평가하는데도 이용할 수 있었다.



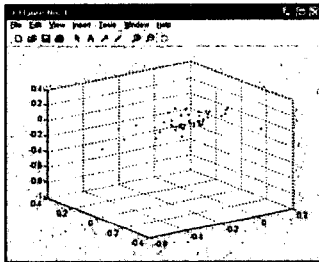
[그림 1] 개선된 방법을 적용한 물체 영상 인식

4. 실험 결과

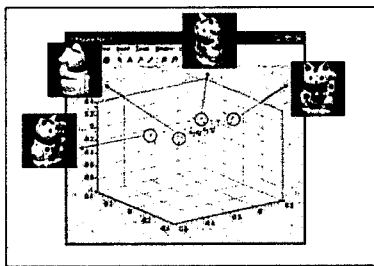
[그림 3]는 [그림 2]를 3차원 물체 영상 공간에 투영시킨 결과이다. [그림 4]은 물체 영상 공간에서 각 영상들이 매핑 되는 결과를 나타내었다.



[그림 2] 고양이 조각 물체의 방향이 틀어지는 경우의 모델 영상



[그림 3] [그림 2]를 3차원 물체 영상물체 영상들의 매핑



[그림 4] 3차원 물체 영상 공간에서공간에서 투영

[표 1]에서 나타나듯이 개선된 방법을 이용한 매칭 방법이 기존의 최소 거리 매칭 방법보다는 매칭 성공률이 높음을 알 수 있다. 특히 잘못된 매칭에 대해서는 많은 개선률을 나타내었다.

[표 1] 매칭 방법에 따른 매칭 성공률

매칭 방법	입력 영상	매칭	잘못된	매칭
		성공률	매칭	성공률
기존의 거리계산	모델 구성(O)	10.5 %	11 %	78.5 %
	모델 구성(X)	15.8 %	20.2 %	62 %
개선된 k-Nearest Neighbor	모델 구성(O)	6.1 %	3.7 %	90.2 %
	모델 구성(X)	13.2 %	16.8 %	70 %

5. 결 론

본 논문에서는 매칭 알고리즘인 개선된 k-Nearest Neighbor 방법을 이용하여 단일 물체 영상 단위의 매칭이 아닌 몇 개의 연속적인 입력 영상에 대해 각각 몇 개의 모델 영상들을 인식의 단위로 이용하였다. 개선된 방법은 물체 영상과 입력 영상의 각 영상에 대한 거리를 비교함으로써 단위 시간 내 투영된 입력 영상간 거리가 먼 경우 빠른 회전이며 한 물체 영상의 전체 거리의 합이 클수록 물체 영상의 크기도 크다는 것을 추정할 수 있었다. 그리고 물체 영상 자체를 계산하여 인식하는 게 아니라 물체 영상 공간이라는 고유공간을 구성한 후에 단지 기여도가 큰 8개의 벡터로만 인식을 수행하기 때문에 자원 축소의 효과까지 얻을 수 있었다.

참 고 문 헌

- [1] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison Wesley, 1992.
- [2] Hiroshi Murase and Shree K. Nayar, "Visual Learning and Recognition 3-D object from appearance," international journal of Computer Vision, Vol.14,1995.
- [3] Christopher Eveland, Kurt Konolige and Robert C. Bolles, "Background Modeling for Segmentation of Video-Rate Stereo Sequences", CVPR 1998, pp. 266-271
- [4] Chambers, J. M. "Computational Methods for Data Analysis. New York : Wiley," 1997
- [5] Gnanadesikan, R. "Methods for Statistical Data Analysis of Multivariate Observations. New York : Wiley," 1997
- [6] Ragael C. Gonzalez and Richard E. Woods "Digital Image Processing," Prentice Hall, 2002