GLCM 을 이용한 다중 베르누이 확률 변수 기반 자동 영상 동적 키워드 추출 방법

*박태준
한양대학교 전자컴퓨터통신공학부
e-mail : sympa@hanyang.ac.kr

Automatically Dynamic Image Annotation Method Based on Multiple Bernoulli Relevance Models Using GLCM Feature

*Tae-Joon Park
Dept. of Electronics and Computer Engineering, Han-Yang University

Abstract

In this paper, I propose an automatic approach to annotating images dynamically based on MBRM(Multiple Bernoulli Relevance Models) using GLCM(Grey Level Co-occurrence Matrix). MBRM is more appropriate to annotate images compare with multinomial distribution. The model is used in limited test set, MSRC-v2 (Microsoft Research Cambridge Image Database). The results show that this model is significantly outperforms previously reported results on the task of image annotation and retrieval.

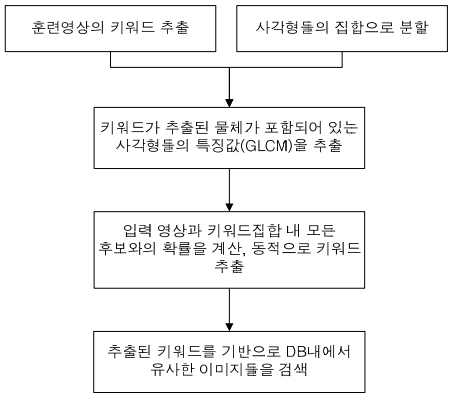
1. 서론

본 논문에서는 문서 분류기법(Text Classification)에서 파생된 다항 분포(Multinomial distribution)와 베르누이 방법(Multi-variate Bernoulli Model)[4] 중 베르누이 방법을 영상에 응용하여 적용시킨 MBRM(Multiple Bernoulli Relevance Models)[1] 과 물체의 질감을 효과적으로 표현할 수 있는 GLCM[2]을 혼합하여 자동으로 영상에서 동적으로 키워드를 추출(Annotation)할 수 있는 방법을 제안한다. 본 논문에서는 MSRC-v2 DB 를 기반으로 실험하였다.

(그림 1) MSRC-v2 DB 이미지와 Ground Truth 의 예

2. 본론

아래(그림 2)에 자동으로 영상의 키워드를 동적으로 추출 하고 검색을 목표로 하는 본 논문에 대한 흐름도를 나타내었다.



(그림 2) 전체 시스템 흐름도

첫째로 훈련영상 내에서 직접 사용자가 판단하여 키워드를 추출한다. 둘째로 훈련영상을 가로·세로 일정한 비율로 임의의 사각형들의 집합으로 분할한다. 이는 키워드 추출의 결과가 Segmentation 의 정확도에 종속되지 않고 알고리즘 속도도 상대적으로 빠르며 또한 정확도 측면에서도 Segmentation 을 사용하는 방법과 큰 차이가 없다는 장점이 있다. 셋째로 추출한 키워드가 있는 사각형들만을 떼어내어 특징값(GLCM)을 추출한다. 이렇게 모든 훈련영상들에 대해서 처리를 한 후 실제 입력 영상이 들어오면 훈련영상에서 했던 방식과 동일하게 임의의 사각형들의 집합으로 분할하고 모든 사각형에 대한 GLCM 을 계산한다. 그 후 분할된 입력 영상과 키워드집합 내 모든 키워드 후보와의 확률을 계산하여 일정 확률 이상의 키워드들을 선택하는데 이 때 추출되는 키워드의 개수는 본 논문에서 제안된 방법을 통해 각 입력 영상에 따라 동적으로 결정된다.

2.1 결합 확률 $P(r_A, w)$

영상과 키워드의 결합 확률 $P(r_A, w)$는 다음과 표시할 수 있다.

$$P(r_A, w) = \sum_{J \in T} \left\{ P_T(J) \prod_{a=1}^{n_A} P_g(g_a|J) \times \prod_{v \in w} P_v(v|J) \prod_{v \notin w} \left(1 - P_V(v|J)\right) \right\} \quad (1)$$

이때 A 는 입력 영상, r_A는 분할된 입력영상, v 는 키워드집합 내의 각각의 키워드, w 는 r_A와의 확률을 구하기 위해 선택된 키워드, n_A 는 임의의 사각형의 개수, T 는 훈련영상집합, J 는 훈련영상집합 내의 각각의 영상, $P_T(J)$는 훈련영상의 개수를 나타낸다. $P_g(g|J)$는 비모수커널추정법에 의한 확률밀도(Non-parametric kernel-based density)로서 가우시안 커널(Gaussian kernel)을 사용하여 다음과 같이 계산한다.

$$P_g(g|J) = \frac{1}{n}\sum_{i=1}^{n}\frac{\exp\{-(g-g_i)^T\Sigma^{-1}(g-g_i)\}}{\sqrt{2^k\pi^k|\Sigma|}} \quad (2)$$

Σ는 공분산 행렬을 의미하며 편의상 $\Sigma=\beta\cdot I$ 로 계산된다. (β는 대역폭(bandwidth)으로서 경험적으로 값을 선택하며 I는 단위행렬을 뜻한다.) n은 J에 있는 GLCM 이 추출된 사각형의 수, g는 분할된 입력영상에서 한 사각형의 GLCM 특징값, g_i는 훈련영상 J 에 있는 한 사각형의 GLCM 값이다.

또한 식(1)에서의 $P_v(v|J)$ 는 다중 베르누이 확률 변수 기반으로 다음과 같이 표현된다.

$$P_v(v|J) = \frac{\mu\delta_{v,J} + N_v}{\mu + N} \quad (3)$$

N은 전체 훈련영상집합 내 영상의 개수, N_v는 v를 포함하는 훈련연상집합 내 영상의 개수, $\delta_{v,J}$는 J 에 v 가 포함된다면 1 값을 갖고 그렇지 않으면 0 값을 갖는다. μ는 경험적으로 값을 선택한다.

2.2 GLCM(Grey Level Co-occurrence Matrix)

총 20 개의 대표적인 GLCM 의 특징값(4 방향, 5 유형)을 사용하였다.

$$\text{Constrast} = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2,\ \text{Energy} = \sqrt{\sum_{i,j=0}^{N-1} {P_{i,j}}^2},$$
$$\text{Entropy} = \sum_{i,j=0}^{N-1} P_{i,j}\left(-\ln P_{i,j}\right),\ \text{Homogenity} = \sum_{i,j=0}^{N-1}\frac{P_{i,j}}{1+(i-j)^2}, \quad (4)$$
$$\text{Dissimilarity} = \sum_{i,j=0}^{N-1} P_{i,j}\,|i-j|$$

식(4)에서 i, j 는 각각 GLCM 행렬의 가로, 세로 픽셀의 위치를 의미하고 $P_{i,j}$는 GLCM 행렬내의 픽셀값을 나타낸다.

2.3 동적으로 결정되는 추출 키워드의 개수

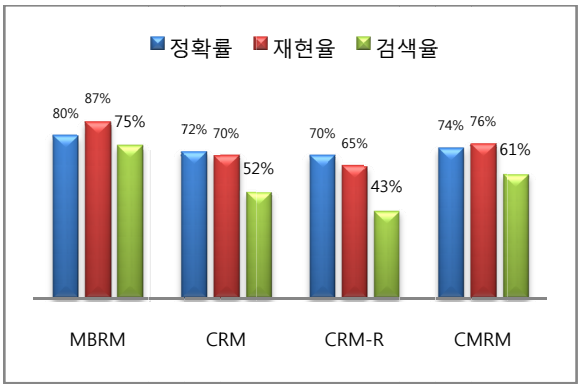
각 입력영상에 따라 추출되는 키워드의 개수는 반드시 동적이어야 하는데 본 논문에서 제안하는 방법은 다음과 같다.

```
I←0;  Max←0;
While(i < n){
        If(Max < D[i] – D[i+1]){                    (5)
                Max←D[i] – D[i+1];
        }
        i++;}
```

식(1)을 통해 각각의 키워드에 대한 확률값이 계산되면 키워드집합 내 키워드의 개수(n)의 크기에 적합한 정렬 알고리즘(병합, 퀵, 버블 등)을 이용하여 확률에 대한 내림차순(또는 오름차순)으로 값을 정렬한다(D[n]). 그 후 가장 높은 확률값을 갖는 i단계와 다음으로 높은 확률값을 갖는 i+1단계의 차이값을 구한 후 i가 n보다 작지 않을 때까지 i를 1씩 증가시키면서 차이값에 대한 최대값(Max)을 구한다. 그 후 최대값이 나오는 지점을 기준으로 키워드 추출의 개수를 결정한다.

3. 실험결과

제안된 기법을 평가하기 위해 Microsoft 사에서 제공하는 MSRC-v2 DB(그림 1)를 기준으로 기존의 여러 방법들(CRM, CRM-R, CMRM 등)과 비교 실험을 하였다.(그림 3) 평가 척도로는 정확률(Precision)과 재현율(Recall), 그리고 추출된 키워드를 이용한 DB 내 다른 영상들과의 검색율을 사용하였다.



(그림 3) 동적 추출 MBRM 과 다른 방법들과의 비교

(그림 3)을 보면 MBRM 이 정확률, 재현율, 검색율 모두 다른 방법들보다 뛰어나다는 것을 확인할 수 있다. 추가적으로 Segmentation 을 사용하는 CRM, CMRM[3]이 예상과는 다르게 영상을 임의의 사각형의 집합으로 분할하는 CRM-R 보다 비슷하거나 더 높게 측정되었는데 이유는 실험 DB 에서 제공하는 이상적인 Ground Truth(그림 1)를 사용하였기 때문이다.

4. 결론

본 논문에서 제안한 방법은 다른 여러 추출 방법들보다 높은 정확률, 재현율, 검색율을 나타내었다. 또한 Segmentation 의 정확도에 종속되지 않기 때문에 제한된 범위의 DB 에서만이 아닌 보다 광범위한 영상에서 사용될 수 있을 것으로 기대된다.

참고문헌

[1] S. L. Feng. et al. “Multiple Bernoulli Relevance Models for Image and Video Annotation”. IEEE Computer Society Conference on CVPR, Vol. 2, No. 2, pp. 1002-1009, 2004
[2] R. Jain. et al. Machine Vision. McGraw-Hill, Inc., 1995
[3] J. Jeon. et al. “Automatic Image Annotation and Retrieval using Cross-Media Relevance Models”. In Proceedings of the 26th Intl. ACM SIGIR Conf., pp. 119-126, 2003
[4] McCallum. et al. “A Comparison of Event Models for Naïve Bayes Text Classification”. AAAI-98 Workshop on Learning for Text Categorization, 1998