

수취인 주소 영역의 영상매칭에 기반한 유사 우편물 추출 방법

정창부⁰, 박상철, 손화정, 김수형
전남대학교 전산학과
cbjeong@iip.chonnam.ac.kr⁰

Postal Image Filtering Method based on ROI Image Matching

Chang-Bu Jeong⁰, Sang-Cheol Park, Hwa-Jeong Son, Soo-Hyung Kim
Dept. of Computer Science, Chonnam National University

요 약

우편물을 배달순서대로 정렬하는 과정을 순로구분이라고 하는데, 우편물의 처리과정에서 가장 많은 시간이 소요되는 부분이 바로 이 순로구분 과정이다. 우정선진국의 순로구분 자동화 시스템은 바코드를 이용하여 순로구분을 자동으로 처리하고 있지만, 국내의 시스템은 순로구분의 전단계까지만 수행하고 나머지 과정은 배달원이 자신의 경험에 의해 수작업으로 처리하고 있다. 본 논문에서는 우편물을 자동 검증하는 방법으로 바코드 대신에 우편영상 특징기반의 Virtual ID 사용을 위한 유사한 우편영상 추출방법을 제안한다. 제안 방법은 질의영상과 후보영상간의 유사성을 판별할 때, 각 영상의 수취인 영역에서 추출된 문자열 Bounding Box들의 대응되는 쌍을 결정하고 그 쌍들의 영역이 겹치는 정도를 유사도로 측정하는 방법이다. 291개의 우편영상에 대하여 실험한 결과, 289개의 영상이 정상적으로 유사한 우편영상으로 추출되었다. 또한, 유사한 우편영상으로 추출된 영상의 평균개수는 3.78개로 비교적 높게 나왔지만 이는 실험 데이터에 동일 DM 발송 우편물이 연속적으로 위치한 경우가 많았기 때문이다.

1. 서 론

최근 인터넷의 급속한 보급과 이용자 증가에 따라 전자메일의 이용이 보편화되었다. 이러한 통신수단의 다양화로 인하여 개인간의 우편물량은 감소하였으나, 광고, 청구서, 카드 결제 등의 DM 발송물량의 증가로 전체적인 우편물량은 증가하는 추세이다. 이러한 우편물량의 증가에 효율적으로 대처하기 위해 우정사업본부에서는 우선 우체국별로 구분하는 발송 및 도착구분 작업을 자동화하여 운영하고 있다. 그러나 이와 같은 작업의 자동화로 인해 효율성은 향상되었으나 순로구분 작업은 아직까지 수작업에 의존하고 있는 실정이다. 집중국에서 발송 및 도착구분 과정을 거쳐 집배국으로 운송된 우편물을 각 집배원별로 구분하고, 이를 다시 집배원의 배달 순서대로 정렬하는 작업을 순로구분이라 하는데, 이와 같은 순로구분을 포함한 집배 과정의 업무는 국내 우편업무 처리의 운가 구성비를 창구, 발착, 운송 및 집배의 네 과정으로 나누어 계산할 경우, 절반 이상을 차지한다. 그러므로, 수작업에 전적으로 의존하는 순로구분은 우편물량이 증가함에 따른 우편물 처리비용 상승의 주요인으로 작용하고, 또한 우편물 자동처리 과정의 가장 커다란 걸림돌로 작용하므로 이를 해결하기 위하여 순로구분의 자동화가 반드시 필요한 실정이다[1].

우편기술 선진국에서는 수많은 연구를 통해 그들의 주소를 자동으로 인식하는 기술을 확보하여 집중국에서 우편물의 인식 결과를 바코드로 변환하고 우편물에 인쇄한다. 그리고 배달국에서의 모든 처리 과정과 순로구분 과정에서 인쇄된 바코드를 활용함으로써 우편물 처리의 자동화와 정보화를 실현하고 있다. 이에 반해, 국내의 경우

는 순로구분 자동화를 실현할 수 있는 수준의 주소인식 기술을 확보하지 못하여 우편번호만을 인식하여 이용하고 나머지는 수작업으로 처리하고 있다. 그래서 정부와 연구기관에서는 우편물 주소의 인식 결과를 바코드로 변환·인쇄하여 이용하는 순로구분 자동화의 구현을 진행하고 있다. 우편물을 자동으로 구별·검증하는 방법으로 위에서 언급한 바코드를 일반적으로 사용하고 있으나, 최근에는 우편물 영상의 특징을 이용한 Virtual ID의 우편물 검증 기술이 제시되고 있다[1][2].

본 논문에서는 Virtual ID를 사용한 우편영상 검증 시스템의 일부로써, 후보 영상군에서 동일 영상과 유사한 영상을 추출하는 방법을 제안한다. 본 논문의 기본적인 목표는 질의영상과 후보영상의 유사 여부를 결정하는 것으로써 우편영상에서 수취인 주소영역(ROI)의 문자열에 대한 BB(Bounding Box) 정보를 비교하는 방법을 이용한다.

2. 관련 연구

2.1 Virtual ID를 이용한 순로구분

순로구분기를 이용한 우편물 순로구분 방법은 우편물의 투입횟수에 따라 구분할 수 있는데, 그림 1은 2Pass 방식을 적용하고 있다. 배달국의 순로구분기에 우편물이 1차 투입(1Pass)되면 수취인 주소를 인식하여 해당 구분칸(Box), 즉 배달 지점별로 분류(1차 정렬)되며 각 우편물의 Virtual ID를 추출하여 저장한다. 그리고 다시 구분칸에 있는 우편물을 순서대로 순로구분기에 재투입(2Pass)하면, 1Pass에서의 Virtual ID 정보와 비교하여 수취인 주소를 결정하고 해당 구분칸에서 재정렬(2차 정

별) 된다. 최종적으로 구분칸과 구분칸의 우편물은 배달원의 배달순서대로 정리되어있기 때문에 효율성을 증대시킬 수 있다[2][3].

Virtual ID를 이용한 우편물 검증 기술은 외국 기업인 Solystic만이 보유하고 있는 기술이다. Solystic은 우편영상의 Virtual ID를 전체 영상에서와 부분 영상(예쉬 영상)에서 각각 구한 특징값들로 구성했는데, 이는 영상의 가로와 세로 크기, 그레이 레벨의 평균과 분산, 엔트로피 등이 있다. 그리고 추가되는 우편영상의 주소열을 구성하는 단어의 개수로 우편물을 최종 검증하고 있다[4]. Solystic에서 제공하는 자료에 의하면 위의 기술로 구현된 자사의 시스템이 95.5%의 성능을 보인다고 하였으나 실제로는 80%정도의 검증 성능을 보인다. 국내에서는 아직까지 Virtual ID를 사용한 우편물 검증 관련 연구는 전무하다.

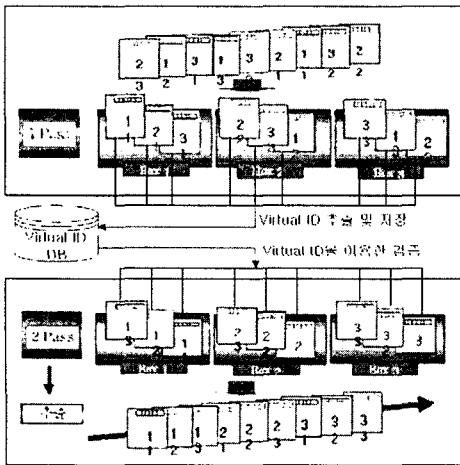


그림 1. Virtual ID를 사용한 순서구분 원리

3. 제안 방법

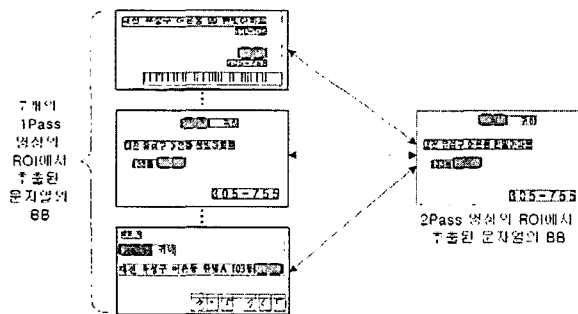


그림 2. 제안 방법의 우편영상 비교 방법

제안 방법은 ROI에서 추출된 문자열 영상들의 정보를 이용하여 2Pass의 영상과 유사한 영상(들)을 1Pass의 후보 영상군에서 추출하는 것이다. 2Pass 영상의 ROI로부터 추출된 문자열의 BB정보가 주어지면 1Pass 영상의

ROI로부터 추출된 문자열의 BB정보와 비교하여 유사도를 계산한다. 유사도 측정은 2Pass 영상과 1Pass 영상에서 각 영상의 기준에 대한 상대적 위치가 유사한 문자열 BB쌍을 결정하고 이들의 영역이 겹치는 정도를 측정하는 것이다. 이렇게 계산된 유사도를 이용하여 1Pass의 후보 영상군에서 유사한 영상들을 추출한다. 본 논문의 성능 실험을 위하여 1Pass와 2Pass의 우편영상의 개수와 순서는 동일한 것으로 가정하고, 1Pass에서의 후보 영상군은 그림 2와 같이 2Pass의 영상과 동일한 영상과 앞뒤로 위치한 각각 3개의 영상, 즉 7개의 영상으로 한다.

3.1 비교할 BB쌍 결정

2Pass와 1Pass의 영상에서 매칭할 BB쌍의 결정을 위하여 각 영상에서의 문자열 BB들의 상대적인 위치를 구한다. 우선 BB들의 상대적인 위치정보를 계산하기 위하여 ROI에서 기준점의 설정이 필요하다. 기준점은 창봉투 밀림과 같은 변이에 덜 민감하여야 하므로, ROI에서 최상단과 최하단에 위치하는 BB의 정보를 이용한다. 그래서 상위 기준점(Base_{Top})과 하위 기준점(Base_{Bottom})을 그림 3과 같이 설정한다. 이때 BB들의 위치를 비교하기 위하여 BB에 대한 대표점(BB_{pos})을 설정하는데, BB의 위치정보(BB_{top}(상), BB_{bottom}(하), BB_{left}(좌), BB_{right}(우))를 이용하여

$$BB_{pos} = \left(BB_{right}, \frac{BB_{top} + BB_{bottom}}{2} \right)$$

와 같이 구한다.

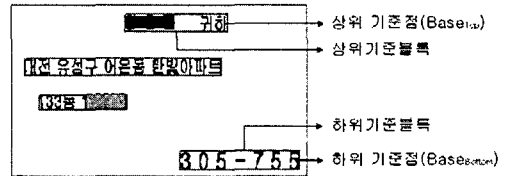


그림 3. 상위 기준점과 하위 기준점의 설정 예

기준점이 결정되면 각 기준점에 대한 각 문자열 BB의 수직거리와 각도를 계산한다. 그림 4는 상위 기준점에 대한 두 번째 BB의 수직거리와 각도를 구하는 예를 보여준다. 그리고 상위와 하위 기준블럭에 해당되는 BB의 수직거리와 각도는 0으로 한다. 이렇게 구해진 BB의 수직거리와 각도의 차이가 최소인 2Pass의 BB와 1Pass의 BB가 비교쌍으로 최종 결정된다.

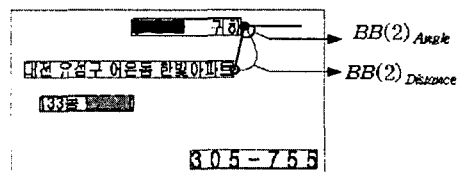


그림 4. 상위 기준점에 대한 수직거리와 각도 계산 예

3.2 유사도 측정

앞절에서 결정된 비교쌍의 BB들은 서로 다른 기준점에 대한 위치정보를 가지고 있으므로 기준점이 같도록 그림 5처럼 BB들의 위치 조정이 필요하다. 2Pass의 하나의 BB에 대한 유사도는 그림 6처럼 BB쌍의 겹치는 영역이 2Pass와 1Pass의 각각 BB영역에서 차지하는 비율들 중에 큰 값으로써,

$$S = \max \left\{ \frac{P(BB_{2Pass} \cap BB_{1Pass})}{P(BB_{2Pass})} \times 100, \frac{P(BB_{2Pass} \cap BB_{1Pass})}{P(BB_{1Pass})} \times 100 \right\}$$

와 같은 식으로 계산된다. 그리고 2Pass 영상과 1Pass 영상의 최종 유사도는 위와 같이 계산된 개개의 BB에 대한 유사도들의 평균으로 계산된다.

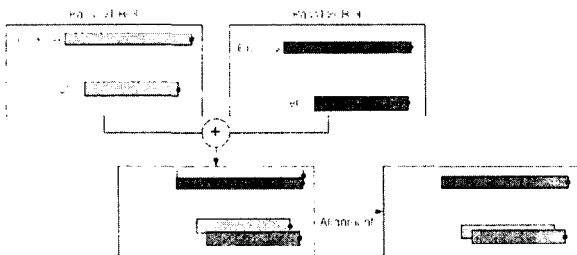


그림 5. BB들의 위치 조정



그림 6. 유사도 측정의 예

3.3 유사한 영상의 결정

7개의 1Pass 후보 영상군에서 2Pass 영상과의 유사도가 가장 큰 영상을 1순위로 선택한다. 그리고 1순위로 선택된 영상의 유사도와의 차이가 10%미만인 영상들을 유사한 영상들로 최종 결정한다.

4. 실험 및 결론

본 논문에서 제안한 방법의 성능을 평가하기 위하여 한국전자통신연구원에서 제공받은 291쌍(1Pass 영상 291개, 2Pass 영상 291개)의 영상을 실험 데이터로 이용하였다. 성능 분석은 2Pass의 영상이 (1) 1순위로 결정된 1Pass의 후보영상과 일치하는 정확률과 (2) 1순위 후보영상의 유사도와의 차이가 10%미만인 후보군에 포함될 확률, 그리고 (3) 유사한 영상으로 추출된 영상의 평균 개수로 측정한다.

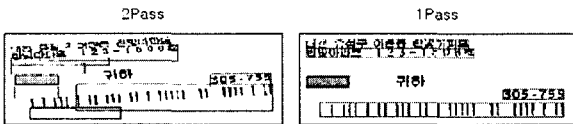


그림 7. 실험 오류의 예

실험결과로 (1)에 대한 성능은 55%의 정확률을 보였고, (2)에 대해서는 99%로써 2개의 2Pass 영상에 대하여 오류가 발생했다. 오류의 원인은 2Pass의 영상이 순

로구분기에서 획득될 때, 그림 7처럼 기울어짐이나 뒤틀림이 발생함으로써 ROI에서의 문자열 BB 정보가 1Pass의 것과는 차이가 나기 때문이다. 그리고 (3)의 결과는 3.78개로써 7개의 후보 영상군에서 절반정도밖에 걸러내지 못했는데, 그 이유는 실험 데이터에서 동일한 DM발송 우편물들이 연속되게 배치되는 경우가 자주 발생하기 때문이다. 표 1은 291개의 영상에 대해 매칭한 결과 얻어진 혼동행렬(confusion matrix)의 일부를 나타내고 있다. 표 1의 결과에서 2Pass의 76번 영상에 대한 유사한 영상이 1Pass의 78번 영상만을 추출함으로써 오류가 발생했는데, 원인은 그림 7과 같다.

표 1. 291개의 영상에 대한 유사도 행렬의 일부

1Pass \ 2Pass	...	73	74	75	76	77	78	79	80	...	결과
73		100	48	67	57						O
74		42	100	63	69	68					O
75		64	71	98	95	85	51				O
76		47	62	31	82	35	98	80			X
77			60	81	63	98	51	43	32		O
78				45	23	44	98	90	43		O
79					27	40	90	98	44		O
80						26	43	27	98		O
...											

5. 결론

본 논문에서는 Virtual ID를 사용한 우편영상 검증 시스템의 일부로써, 후보 영상군에서 동일 영상과 유사한 영상을 추출하는 방법을 제안하였다. 질의영상과 후보영상에서 기준점을 설정하고, 기준점으로부터 계산된 상대 위치정보를 이용하여 대응되는 문자열의 BB쌍을 결정하였다. 그리고 그것들의 영역이 겹치는 정도를 이용하여 유사도를 측정하고, 유사한 영상을 추출하였다.

향후 연구로는 유사한 영상의 평균 개수를 줄이기 위하여 문자열 영상의 수직투영 프로파일을 비교하는 DTW (Dynamic Time Warping)의 적용하고자 한다.

Acknowledgement

본 연구는 한국전자통신연구원의 "Virtual ID 사용을 위한 우편 영상 검증 기술 연구" 과제의 지원으로 수행되었다.

참고문헌

- [1] 정보통신연구진흥원, "국내외 순로구분 자동화 기술 현황," 주간기술동향, 통권 1096호, 2003
- [2] 송재관 외 4명, "자동기계에 의한 우편물 순로구분 동향," ETRI 전자통신동향분석, 제15권 제6호, 2000
- [3] 김수형, 박상철 외 4명, Virtual ID 사용을 위한 우편 영상 검증 기술 연구 과제 중간보고서, 연구보고서, 한국전자통신연구원, 2004
- [4] <http://www.solystic.com/>