

우편물 자동처리 성능 향상을 위한 우편용 바코드 오류교정에 관한 연구

박 문 성[†] · 우 동 진^{††} · 김 혜 규^{†††}

요 약

현재 대부분의 우편집중국에서는 우편봉투의 우편번호 및 주소를 광학문자판독기에 의하여 인식하고 3 of 5 형광 바코드를 인쇄하고 판독하여 구분기에 의하여 처리하고 있다. 약 31%~35% 정도의 우편물은 자동구분 처리가 불가능한 우편불이며, 대부분 우편주소 및 바코드의 인쇄 품질의 저하로 인하여 발생된다. 이에 따라, 우편물을 100% 바코드로 자동구분 처리하여 우편고객에게 최고 품질의 우편배달 서비스가 제공될 수 있도록 우편용 바코드 판독율의 향상을 목표로 하였다.

본 논문에서는 우편물 자동구분 처리시스템의 성능을 향상시키기 위하여 우편용 3 of 5 바코드 오류교정 방법을 제안한 것이며 우편번호와 체크문자로부터 자동으로 오류패턴 규칙을 생성한 후 이를 이용하여 오류를 교정한다. 제안한 방법으로 시험한 결과에 의하면 약 21.09%~50%의 정확도로 3 of 5 바코드 오류를 교정할 수 있게 된다.

A Study on the Error Correction of Postal Barcode to Increase the Performance of Automatic Processing in Mail Items

Moon-Sung Park[†] · Dong-Chin Woo^{††} · Hye-Kyu Kim^{†††}

ABSTRACT

In most mail automatic processing centers, after facing and canceling, letter mails is passed through an Optical Character Recognition/Barcode Sorter (OCR/BS) to read the postal code and 3 of 5 fluorescent (luminescent) barcode is applied. Normally, 31%~35% of these mails are rejected. The main reasons for reading failures are poor printing quality of addresses and barcodes, script printing, and failure to locate the address. Our goal is to provide mailer with the highest quality of delivery service and customer barcode service as we move toward 100% barcoding automation of letter mails.

In this paper, we propose a method of correcting 3 of 5 barcode errors for increasing the performance of automatic processing system in mail items. 3 of 5 barcode errors are corrected by using error pattern rules, which are automatically extracted from parity bits and postal codes. The experimental result shows that our system can correct barcode errors in accuracy of 21.09%~50%, respectively.

† 정 회 원 : 한국전자통신연구원 우정자동화팀 연구원
†† 정 회 원 : 한국전자통신연구원 우정자동화팀장 책임기술원
††† 정 회 원 : 한국전자통신연구원 우정기술연구부장 책임연구원
논문접수 : 1999년 3월 23일, 심사완료 : 1999년 8월 5일

1. 서 론

우정 업무에 있어 우편물 구분 및 배달작업은 인력의 의존도가 가장 높은 부분으로 80%이상을 차지한다. 대부분의 국가에서는 이러한 인력 의존도를 최소화하고, 안정적인 우편배달 서비스를 제공하기 위한 목적으로 기계적인 처리방법을 도입하여 적용하고 있다[1, 8, 9]. 우편물 자동구분 처리는 우편물 자동 이송장치와 이동 중인 우편물상에 기재된 우편번호 및 주소정보를 광학문자판독기에 의하여 인식하고, 판독된 정보를 기계 바코드(형광 바코드)로 인쇄하며, LSM(Letter Sorting Machine)에서는 우편물상에 인쇄된 바코드를 판독하여 자동으로 구분하는 과정으로 이루어진다[5]. 만일, 광학문자판독과정에서 오류인 우편물은 해당 우편물의 이미지를 캡처(capture)하며, 비디오 코딩 시스템을 통하여 우편번호를 입력하고 재구분 처리하는 방법으로 구성하여 운영되고 있다[3, 14].

광학문자판독기(OCR; Optical Character Reader)에 의하여 이동하는 우편물상에 표기된 우편번호를 인식하는 과정에서 존재하는 인식되지 않는 오류우편물은 비디오 코딩 방법에 의하여 처리하고 있으며, 비디오 코딩에 의한 처리량을 줄이고 보다 효과적인 우편물 자동구분 처리를 위하여 고객 바코드 인쇄제도를 도입하여 적용하고 있다[4, 6, 13]. 고객 바코드 인쇄제도는 OCR에 의하여 판독된 정보를 기계 바코드로 인쇄하는 방법을 대신하여 고객이 기계 바코드와 동일한 규격으로 흑백 바코드를 인쇄하게 하고, LSM에서는 이 흑백 바코드 판독하여 처리토록 함으로써 OCR 및 비디오 코딩의 양을 축소하고, 우편물 자동구분을 할 수 있도록 하는 방법이다[7, 8, 9]. 고객 바코드 인쇄는 대부분 다량의 우편물을 취급하는 업체에 의하여 우편물상에 고객 바코드를 인쇄하여 우편집중국에 접수하게 되는데, 이 우편물을 모두 검사할 수 없으므로 고객 바코드 인증제도에 의하여 인증된 업체를 대상으로 접수하여 처리하고 있다. 그러나, 우편물 접수과정에서는 모든 우편물을 검사할 수 없으므로 샘플 우편물을 대상으로 검사하고 있다. 인증된 다량의 우편물 취급업체는 다량의 우편물을 인쇄하여야 하므로 프린터의 인쇄 품질이 일정하게 유지될 수 없는 문제점을 가지고 있다. 이와 같은 문제점을 해결하기 위해서는 기계에서 인쇄되는 형광 바코드와 우편고객에 의하여 인쇄되는 흑백 바코드의 판독율을 향상시키기 위한 노력이 필요하다.

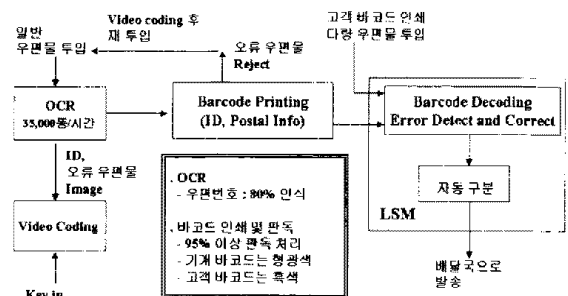
본 논문은 현재 사용중인 3 of 5 바코드의 오류 교정의 범위를 확장하여 바코드의 판독율 향상시켜 오류 우편물의 최소화 및 자동처리 촉진을 연구의 일환으로 수행된 결과이다. 본 논문의 내용은 제 2장에서 3 of 5 바코드의 개요와 연구 배경과 3 of 5 바코드의 오류교정 범위를 확장하기 위한 요구사항 및 방법을 다루고, 제 3장에서는 오류교정의 범위를 확장시키기 위한 방법을 제시하였으며, 오류교정페턴 생성방법을 설계하였다. 제 4장에서는 본 논문에서 제시한 방법에 의하여 오류교정 시스템을 설계하였으며, 제 5장에서는 오류교정 알고리즘에 실제로 사용하는 우편번호 정보를 적용하여 분석하였다. 제 6장에서는 적용한 결과를 요약하고, 추후연구에 대하여 다루고자 한다.

2. 우편용 3 of 5바코드의 개요

2.1 연구 배경

우편물 자동구분 처리를 위해 사용되는 3 of 5 바코드는 OCR에 의하여 인식된 우편번호를 이동하는 우편물상에 고속으로 인쇄하기 용이한 구조로 매우 간략한 형태로 구성되어 있으며, 이 바코드를 판독하여 자동구분하고 있다. 시간당 35,000통을 처리하는 이 시스템의 구성은 다음과 같은 흐름에 의하여 우편물을 자동으로 구분하여 처리하고 있다.

- 1) 우편물 투입
- 2) 운송용 벨트에 의한 우편물 이송
- 3) 이동중인 우편물의 우편번호를 OCR에 의하여 80% 이상을 판독하여 처리
- 4) 판독된 우편번호 및 체크문자를 생성하여 해당 우편물에 3 of 5 형광 바코드 인쇄
- 5) 이동 중인 우편물의 바코드(흑백 및 형광색)를 판독(95%)하여 해당 구분 칸(bin)으로 전달[11]



(그림 1) 우편물 자동구분처리의 흐름도

(그림 1)과 같은 과실로 이루어지는 우편물 자동 구분과정에서 OCR에서의 인식되지 않아 오류우편물로 구분되는 경우(20%)에는 비디오 코딩을 위하여 이동중인 우편물의 이미지를 OCR에서 캡처하여 저장하고, 오류우편물에 비디오 이미지의 ID를 우편봉투 뒷면 하단 혹은 정해진 영역에 인쇄하여, 오류우편물 구분 칸으로 전달한다. 이 오류우편물의 이미지와 해당 ID는 비디오 코딩 시스템에 전달되며 운영자에 의해 해당 우편번호가 입력되고, 이 입력 정보를 자동구분처리 시스템에 해당 ID와 우편번호가 전달하고, 이 작업이 완료되면, 오류우편물을 자동구분처리 시스템에 재투입하여 ID값과 입력된 우편번호에 의해 자동으로 구분한다. 이 오류우편물량의 증가로 비디오 코딩을 위한 인력 의존도가 높아지고 있다.

이러한 문제점들을 해결하기 위한 방법으로 다량의 우편물을 취급하는 업체를 대상으로 기계 바코드와 동일한 규격으로 검은색/흰색을 조합으로 하는 바코드를 인쇄토록 하고, LSM에 흑백 바코드 판독기를 추가하여 OCR에 의한 인식 과정과 바코드 인쇄 과정을 생략하고 자동 구분하게 함으로써 오류우편물의 최소화 및 자동처리의 효율을 극대화 시키고 있다[7,8,10,14]. 그러나, 고객이 사전에 인쇄한 바코드 품질의 저하로 인한 문제점이 발생될 수 있다.

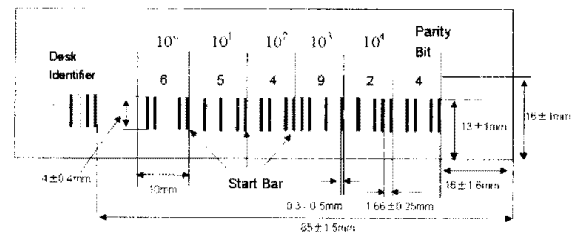
고객 바코드의 인쇄품질 저하는 고객 바코드 검증시 완벽하게 인쇄된 우편물에 의하여 검증을 받아야 하며 인증된 다량우편물 취급업체는 우편물에 고객 바코드를 인쇄하여 접수하게 되는데 인쇄용 잉크의 불충분, 일부분의 인쇄 상태불량 등으로 바코드 판독율이 저하되는 문제점이 있다. 이에 따라, 바코드문자의 부분적인 훼손을 보정하기 위한 방법이 요구되고 있다[12]. 이에 따른 기계 바코드 및 고객 바코드의 판독율을 향상시키기 위한 노력이 요구되고 있다. 바코드 판독율을 향상시키기 위한 방법의 하나로 바코드 오류교정 비율을 향상시키기 위한 연구가 필요하다. 또한, 우편물 자동구분 처리시스템에 적용하기 전에 실제 바코드 오류교정 비율이 분석되어야 할 것이다.

2.2 우편용 3 of 5 바코드의 규격

기계 바코드의 규격과 동일한 규격으로 고객 바코드를 인쇄하여 자동으로 구분하는 방법이 적용되고 있다. 이에 따른, 바코드의 규격은 숫자 0~9까지의 바코드문자를 표현할 수 있으며, 바코드문자는 동기화를

위한 시작 바(start bar)를 모두 포함하고 있다. 또한, 바코드문자 당 10mm의 공간을 사용하며, 하나의 바코드문자에 흑색 바 3개와 백색 바 2개 그리고 시작 바로 구성되며, 각 바의 위치에 따른 바의 가중치는 0, 1,2,4,7을 적용하고 있다. 또한, 표현된 바코드문자의 오류를 검출하고 보정하기 위하여 체크문자(parity bit)를 1문자 적용하여 사용한다. 현재, 사용되는 바코드의 문자는 우편번호 6자리와 체크문자 1자리를 적용하여 70mm의 길이에 표현하고 있다.

하나의 바코드문자의 길이가 10mm이므로, 우편봉투의 최소 길이 140mm를 기준으로 할 경우에 바코드문자를 인쇄할 수 있는 영역은 좌측여백 35mm, 우측여백 20mm를 제외한 나머지 값에 의하여 85mm 이므로 8개의 바코드문자를 인쇄할 수 있으며, (그림 2)에서의 Desk Identifier인 우편물 자동구분 처리시스템을 구분하기 위한 1자리를 추가로 인쇄하여 사용되는 것을 포함하면 8개의 바코드문자가 적용되고 있다. 그리고, Desk Identifier와 우편번호 및 체크문자의 사이에 1개 바코드문자 길이 만큼의 공간이 적용되므로 실제 추가로 바코드문자를 수록하기 위한 공간이 없다. 따라서, 하나의 체크문자에 의한 하나의 바코드문자를 교정하는 방법이 적용된다.



(그림 2) 3 of 5 바코드 규격

체크문자의 계산방법은 우편 번호가 305-505일 경우, Parallel balance는 $18(3+0+5+5+0+5) = 18/10 = 1r8$ 이고, 체크문자는 $10-8(나머지)$ 에 의해 2의 값이 된다 (10 modulo)1, 2, 4.

2.3 오류교정을 위한 요구사항

우편용 3 of 5 바코드는 우편물 자동구분 처리를 위하여 사용되며, 숫자 정보 6자리와 체크문자 하나를 사용하고 있다. 이러한 바코드의 구조는 N Modulo 방법에 의한 체크문자를 생성하므로 바코드문자 판독과정에서 하나의 바코드문자가 오류인 경우에만 오류를

교정하여 처리할 수 있고, 2개 이상의 바코드문자가 오류인 경우에는 체크문자에 의하여 오류교정이 불가하므로 오류우편물로 분류하여 재 구분작업을 해야 한다. 재 구분작업의 경우 이미 우편봉투에 인쇄되어진 바코드문자를 사용할 수 없으므로 대부분 수작업 구분에 의존된다[9, 11, 12, 13]. 이에 따른 문제점을 최소화하기 위해서는 오류교정 비율을 향상시키거나 오류를 최소화하기 위하여 바코드 인쇄를 정해진 규격으로 정확하게 하여야 한다. 바코드 인쇄는 고객 바코드 인쇄제도에 정해진 규격에 의하여 정확하게 인쇄되었는지 사진에 검사하는 고객 바코드 인증제도를 시행함으로써, 오류 발생을 부분적으로 최소화할 수 있으며, 항상 모든 우편물을 검사하여 처리할 수 없으므로 오류교정을 추가적으로 할 수 있도록 해야 한다. 또한, 바코드문자의 구조를 변경하여 오류교정 방법을 추가하는 것보다는 현재 사용되고 있는 바코드문자의 규격을 그대로 유지시키면서 우편물 자동구분 처리시스템에 일부 기능만 추가하여 오류교정이 가능하도록 방법이 요구된다. 현재 사용되고 있는 우편번호는 도, 시, 군, 구, 동 등에 관한 지역을 구분하는 특징이 있다. 또한, 체크문자는 우편번호 값에 의하여 계산되는 값이다. 이 두 가지 조건을 이용하여 오류교정 패턴을 생성하는 방법이 고려되어야 한다.

3. 우편용 3 of 5 바코드의 오류교정

본 장에서는 2장의 요구사항을 기반으로 바코드 판독 오류교정의 범위를 확장하기 위하여 우편번호의 특징과 체크문자에 의한 오류패턴을 생성하여 부분적으로 삭제되어도 오류교정이 가능한 방법을 설계한 것이다.

3.1 체크문자에 의한 오류교정

바코드문자의 규격에 따른 오류교정 방법은 바코드문자의 값과 체크문자의 값을 비교하여 하나의 바코드문자를 교정할 수 있다. 오류교정 방법으로는 체크문자의 값이 판독되고 우편번호에 해당하는 바코드문자 판독과정에서 오류인 바코드문자의 값을 0으로 설정하고 판독된 우편번호의 값을 모두 더하여 체크문자의 값과 차를 구하는 방법으로 오류인 바코드문자를 교정할 수 있다. 예를 들면, 우편번호가 305-350이면, 체크문자는 우편번호의 합인 $3+0+5+3+5+0=16$ 값의 full 십진수인 20에서 16을 빼면, 체크문자는 4가 된다.

1) 정상 판독

정상적인 경우에는 체크문자 4의 값이 판독된 후 우편번호 3, 0, 5, 3, 5, 0 순으로 판독된다.

- 우편번호 합은 $3+5+0+3+5+0 = 16$ 이 된다.
- 체크문자의 값을 계산하면 4가 된다.
- 판독된 체크문자와 우편번호 판독에 의하여 계산된 체크문자의 값이 동일하면 정상으로 처리된다.

2) 오류교정 : 하나의 바코드문자 오류

- 체크문자 4와 값이 판독된 후 *, 0, 5, 3, 5, 0의 순으로 판독
- 1번째 자리의 값을 0으로 오류 바코드문자의 수를 1 그리고 오류 위치 값을 1로 설정한다.
- 우편번호의 합은 $0(*)+5+0+3+5+0 = 13$ 이 된다.
- 따라서, 체크문자의 값을 계산하면 7이 된다.
- 판독된 체크문자와 우편번호 판독에 의하여 계산된 체크문자의 값이 동일하지 않고, 오류 바코드문자의 수가 1인 경우에는 판독된 체크문자의 계산 값 7에서 체크문자 판독 값 4를 빼다.
- $7-4 = 3$ 이므로 1번째 오류 바코드문자의 값을 3으로 교정하여 처리한다.

판독과정에서 2개 이상의 바코드문자 값이 오류인 경우에 위와 같은 방법으로 계산하면, 오류인 바코드문자의 값들의 합이 계산되는 특성이 있다. 따라서, 바코드문자는 6개의 바(흰색 바 포함)가 하나의 바코드문자를 구성하므로 16.67%의 정확도로 오류를 교정할 수 있게 된다. 2개 이상의 바코드문자의 오류교정이 불가능하므로 또 다른 방법을 적용하기 위한 연구를 필요로 한다.

3.2 오류패턴 생성 및 오류교정 알고리즘 설계

본 절에서는 우편번호의 특성과 체크문자의 배열에 따른 오류패턴을 생성하여 추가로 오류교정할 수 있는 방법을 제시한 것이다. 오류교정을 위한 참조테이블의 생성방법은 우편번호는 6자리이고, 체크문자 1자리를 사용하는 특성과 우편번호 앞의 3자리는 대 구분 즉, 소도시 이상의 영역을 구분하기 위한 번호이며, 뒤에 3자리는 배달구역별로 분류하기 위한 코드로 사용되고 있다. 이러한 특성을 고려하여 우편번호 앞의 3자리 혹은 4자리 단위로 구분하고 우편번호 및 체크문자의 값으로 영역으로 정렬하여, 체크문자의 값이 정렬된 영역에서 유일한 값을 가지는 패턴들을 저장한 후, 바

코드 문자 판독과정에 적용하는 방법으로 설계하였다. 이 방법을 적용한 이유로는 체크문자 및 바코드문자의 추가 또는 변경하지 않고 오류패턴 정보만을 추가적으로 적용하기 위한 방법을 고려했던 것이다. 다음과 같은 방법을 절차에 의하여 오류교정을 위한 오류패턴을 생성하였다. 우편번호 데이터베이스에서 우편번호를 바코드로 변화하기 위한 우편번호 바코드 문자 참조 테이블을 생성(체크문자의 값 포함)한다.

1) 생성된 우편번호 바코드 테이블의 값을 체크문자의 값을 기준으로 하는 오류패턴 테이블 생성한다. 체크문자의 값을 chk_{val} , 바코드문자의 값을 $\{p_1, p_2, p_3, p_4, p_5, p_6\}$ 로 정의하고 체크문자와 바코드문자에 의한 패턴 값을 기준으로 수식(1)에서 수식(6)을 적용하여 $\{c_1, c_2, c_3, c_4, c_5\}$ 의 값이 생성하였으며, 이 값을 오류패턴 생성의 기준 수식으로 적용하였다. 체크문자의 계산 값은 항상 정수가 되도록 하며, 이 정수 값을 10으로 나눈 후, 나머지 값을 적용한다.

$$chk_{val} - \sum_{i=1}^6 P_i / (10 + 1) \cdot 10 - \sum_{i=1}^6 P_i \mid 10 \text{mod} \dots (1)$$

$$c_1 = |10 - chk_{val} - P_6| \dots (2)$$

$$c_2 = |c_1 - P_5| \dots (3)$$

$$c_3 = |c_2 - P_4| \dots (4)$$

$$c_4 = |c_3 - P_3| \dots (5)$$

$$c_5 = |c_4 - P_2| \dots (6)$$

- 2) 이와 같이 계산된 오류패턴 값들을 우편번호 앞의 3자리 $\{P_1, P_2, P_3\}$ 혹은 4자리 단위 $\{P_1, P_2, P_3, P_4\}$ 로 정렬한다.
- 3) 상기 절차에 의하여 정렬된 우편번호의 테이블에서 체크문자 값 chk_{val} 을 기준으로 우편번호 영역별로 정렬한다.
- 4) 각 영역 별로 체크문자에 의해 구분된 값들 중에서 단 하나의 체크문자 값에 해당하는 패턴과 동일한 체크문자의 값이 발생되면 $\{c_1, c_2, c_3, c_4, c_5\}$ 값들을 비교하여 일치되지 않는 바코드문자의 값에 대한 우편번호, 체크문자, $\{c_1, c_2, c_3, c_4, c_5\}$ 값을 대상으로 하는 오류교정 패턴들을 생성한다.

상기 과정에 의해 생성된 바코드 오류교정 패턴은 삭제된 바코드문자를 교정하거나 바코드 문자 판독방법을 개선하기 위하여 적용하고자 한다. 오류패턴 정보를

원래 같은 방법에 의하여 생성한 예전 데이터(표 1)와 같다.

<표 1>은 우편번호가 469-***인 영역을 대상으로 생성한 예로 체크문자의 순서로 배열된 값들의 특성을 비교하면 체크문자 7인 값을 제외하고 단일한 체크문자 값을 가지게 된다. 따라서, 체크문자와 469-8까지 판독되면 체크문자가 7인 값을 제외하고 오류패턴 값에 의하여 우편번호를 판독할 수 있게 된다.

<표 1> 오류패턴 참조 테이블

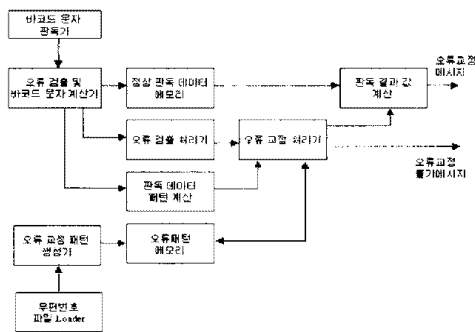
우편 번호						체크 문자	오류패턴				
P1	P2	P3	P4	P5	P6		C1	C2	C3	C4	C5
4	6	9	8	3	0	0	6	0	9	1	2
4	6	9	8	2	0	1	5	1	8	0	2
4	6	9	8	1	0	2	4	2	7	1	0
4	6	9	8	0	0	3	3	3	6	2	2
4	6	9	8	9	0	4	2	4	5	3	6
4	6	9	8	8	0	5	1	5	4	4	4
4	6	9	8	7	0	6	0	6	3	5	2
4	6	9	8	1	5	7	1	5	4	4	3
4	6	9	8	6	0	7	1	5	4	4	2
4	6	9	8	5	0	8	2	4	5	3	2

체크문자의 값이 7인 값을 제외하고 오류패턴 값은 모두 다르기 때문에 판독되는 바코드 문자, 체크문자, 오류패턴 값과 바코드 문자 단위로 판독되는 과정에서 계산되는 오류패턴 계산 값과 비교하면 2개 이상의 우편번호가 불량일 경우에도 오류교정이 가능하게 된다. 또한, 오류패턴 참조 테이블을 469-***영역을 체크문자 순으로 정렬하여도 똑 같은 결과가 됨을 알 수 있었다. 체크문자가 469-***영역에서 하나의 값만 가지는 경우에는 뒤 3자리가 모두 삭제되어도 오류패턴 참조 테이블 값에 의하여 판독될 수 있다. 오류문자가 하나 발생되었을 경우에는 체크문자 값에 의하여 어느 위치와 관계없이 오류교정이 가능하다. 이와 같은 방법들을 적용하면 우편용 3 of 5 바코드 판독율이 향상될 수 있을 것으로 기대된다. 이와 같은 오류교정 방법을 적용할 경우에 최대 3개의 바코드 문자를 교정할 수 있으므로 50%까지도 오류교정이 가능하다. 그러나, 오류패턴이 존재하는 우편번호만 추가로 오류교정이 가능하므로 실제적으로는 우편번호의 사용범위에 중복되어 적용된 체크문자 수가 적어야 효과적이다. 이러한 문제점을 해결하는 방법으로는 우편번호와 체크문자의 값이 단일 값을 갖는 우편번호로 지정되어야 하며, 단

일한 체크문자의 값이 없을 경우에는 이미 중복된 체크문자에 해당하는 우편번호를 배정하면 오류교정의 비율을 안정적으로 유지시킬 수 있을 것으로 기대된다.

4. 우편용 3 of 5 바코드 오류교정 시스템

우편용 3 of 5 바코드 오류교정 시스템은 앞 장에서 설계한 방법을 적용할 수 있도록 우편번호 파일을 읽어 자동으로 오류패턴을 생성하고, 오류교정 시스템의 오류패턴 메모리에 생성된 오류패턴을 저장하여 바코드 문자 판독시 검색할 수 있도록 구성하였다. 또한, 바코드 문자 판독과정에서 오류가 발생되는지 검출하는 방법은 바코드문자의 값이 정상상태로 바가 판독되고, 바코드 문자 값이 올바르게 판독되는지 확인하는 과정과 오류패턴 정보가 존재하는 우편번호인 경우에는 오류패턴 값을 계산하여 메모리에 존재하는 오류패턴 값과 비교할 수 있도록 구성하였다. 체크문자와 우편번호 2~3자리가 판독되면 체크문자와 판독중인 우편번호와 일치하는 오류패턴 값이 존재하는지 검색한다. 만일, 오류패턴의 값이 존재하면 해당하는 오류패턴의 우편번호로 판독된 것으로 처리될 수 있도록 구성한 것을 특징으로 한다.

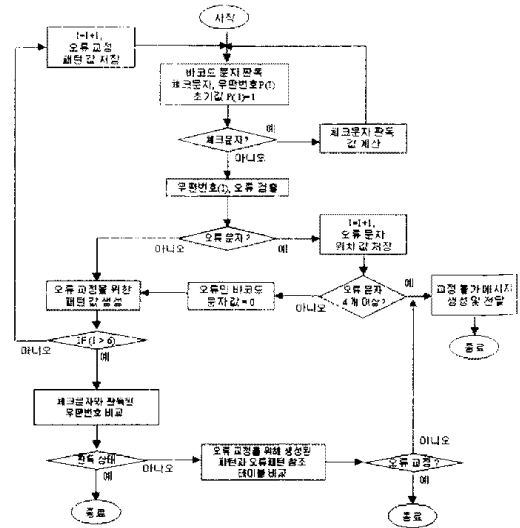


(그림 3) 3 of 5 바코드 오류교정 시스템

오류 바코드문자의 오류교정은 체크문자에 의한 오류교정을 수행함과 동시에 오류교정 패턴 값이 존재하는 우편번호에 대한 오류패턴 메모리의 값을 검색하고 비교하여 오류교정이 되도록 하였으며, 오류교정 시스템의 구성도는 (그림 3)과 같다.

(그림 4)는 오류교정을 위한 흐름을 보인 것으로 오류인 바코드문자의 수가 4개 이상이면 오류교정이 불가능한 것과 체크문자 및 오류교정 패턴에 의하여 오류교정이 불가능한 우편물은 오류우편물 구분 칸으로 분류될 수 있도록 하는 기능을 포함하는 것을 특징으로 한다. 또한, 오류 바코드문자의 정보는 오류 바코드

문자의 위치, 오류 바코드문자의 수 그리고 오류 바코드 문자 값을 0으로 설정하는 과정이 포함된다.



(그림 4) 오류교정의 동작 흐름

5. 오류교정 알고리즘의 적용

본 장에서는 앞에서 정의한 오류패턴에 의한 오류교정 방법과 체크문자에 의한 오류교정 방법을 실제 우편번호를 적용할 수 있도록 우편번호 파일을 생성하고, 오류패턴을 생성하여 적용하였다.

우편물의 자동구분은 우편번호 앞의 3자리를 판독하여 해당지역으로 운송하기 위한 구분작업과 수취인에게 배달하기 위하여 우편번호 6자리 모두를 판독하여 구분하는 배달구분 방법으로 구분된다. 이에 따라 발송 구분된 우편물 혹은 발송 및 배달구분 작업을 동시에 수행하여야 하는 경우에 적용되는 오류교정 알고리즘으로 우편집중국 단위로 오류교정패턴을 생성하여 적용되어야 하므로 각 지역별로 구분하여 우편번호의 추가 및 변경 시 오류교정 비율을 최대한 유지하기 위한 기반 정보로 활용되어야 한다.

이에 따라 서울특별시, 광역시, 각 도별로 구분하여 현재 사용중인 우편번호 8,177개를 대상으로 분석하였다. 예를 들어, 대구광역시의 경우에 700-010부터 711-840까지 영역이 배정되어 있고 서구의 경우에는 700-010부터 700-600까지 배정되어 있다. 대구광역시 서구에 배정되어 사용되는 우편번호의 수는 27개이다.

이러한 값들을 기준으로 우편번호의 배정 비율을 계산하여 추가로 부여 가능한 우편번호의 양을 계산할

는 있도록 우편번호 배정율을 보였다. 나머지 우편번호 공간에 추가 혹은 변경사항이 발생되어 우편번호를 배정하려면, 다음과 같은 조건이 만족되도록 우편번호를 배정하여 오류교정 비율을 유지하거나 향상시킬 수 있는 방법을 적용하여야 한다.

- 1) 우편번호 배정영역에서 현재 사용중인 우편번호중에서 사용하지 않고 있는 체크문자(0~9 사이의 값)를 찾는다.
- 2) (1)항에서 중복되지 않는 체크문자가 존재하면, 배정영역에서 우편번호 뒤의 3자리 중에서 인접된 우편번호 중에서 우편번호와 차이가 적은 값을 체크문자들을 생성하여 유일한 값을 가지는 체크문자인지 확인한 후, 이 우편번호에 대한 오류교정패턴을 생성하여 적용한다.
- 3) (1)항의 검색 결과에 의한 값이 존재하지 않으면, 2개 이상의 체크문자가 있는지 확인한다.
- 4) (3)항에 의하여 체크문자가 존재하면, 배정 영역에서 우편번호 뒤의 3자리 중에서 인접된 우편번호의 구역의 우편번호와 차이가 적은 값을 선택의 체크문자들을 생성하여 2개 이상이 존재하는 체크문자 값과 동일한 우편번호를 적용한다.
- 5) 상기 항목들에 해당되지 않는 경우에는 배정 영역에서 우편번호 뒤의 3자리 중에서 인접된 우편번호의 구역의 우편번호와 차이가 적은 값을 선택의 체크문자를 생성하고, 중복될 체크문자에 대한 오류교정 패턴 값을 제거한 후 적용한다.

이와 같은 우편번호 배정방법에 따라 오류교정 비율의 값이 증가 혹은 감소되는 현상이 발생될 수 있다.

오류교정 비율의 계산은 가장 먼저 관측된 체크문자의 판독 이후 우편번호에 대한 오류가 발생된 경우를 가정하였다. 또한, 오류패턴을 포함하지 않고, 오류교정이 가능한 바코드분자는 하나이므로 6자리 우편번호에 대하여 16.67%의 오류교정이 가능하다.

본 논문에서 제시한 오류교정 방법을 적용하면 오류교정패턴 정보가 존재하는 우편번호만 추가로 오류교정이 가능하므로 우편번호의 설정 수와 오류교정패턴의 수를 해당 영역별로 구분하여 적용하였다. 또한, 오류교정패턴 값에 의하여 추가로 오류를 교정할 수 있는 최대 바코드문자의 수는 3개이다. 이러한 조건은 우편번호 뒤 3자리가 모두 삭제되었을 경우에 가능한 조건이다.

이에 따라 사용중인 우편번호의 수를 T_{p_n} , 오류교정패턴이 있는 우편번호의 수를 E_{p_n} 라 하면, 오류교정패턴이 없는 우편번호의 수는 (수식 7)과 같다. 상기와 같은 조건에 의하여 오류교정패턴에 의한 교정 가능한 바코드분자가 1자리 추가된 경우(최소 오류교정율)는 (수식 8)과 같이 정의되며, 우편번호 뒤 3자리가 모두 삭제되어도 오류교정이 가능한 경우에 최대 오류교정율은 (수식 9)와 같이 정의된다

$$N_{p_n} = T_{p_n} - E_{p_n} \dots\dots\dots (7)$$

$$Error_Correct_1 = \frac{E_{p_n} + T_{p_n}}{6T_{p_n}} \dots\dots\dots (8)$$

$$Error_Correct_2 = \frac{N_{p_n} + 3E_{p_n}}{6T_{p_n}} \dots\dots\dots (9)$$

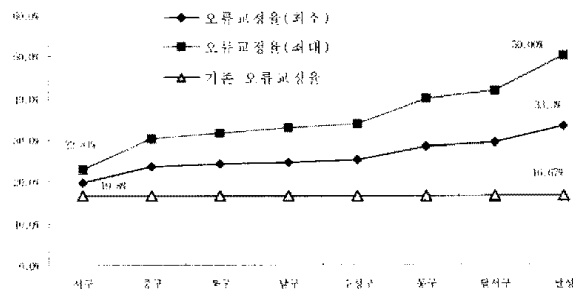
<표 2>는 상기와 같은 수식들을 실제 우편번호에 적

<표 2> 서울특별시, 광역시, 각 도별 오류교정 비교표

지역 구분	우편번호 범위	사용중인 우편번호	오류교정 패턴생성	오류교정율(최소)	오류교정율(최대)	우편번호 배정율
서울	100012 ~ 158774	2073	275	18.9%	21.09%	10.90%
부산	600010 ~ 618450	501	144	21.5%	26.25%	4.75%
광주	500010 ~ 506600	298	101	22.3%	27.96%	8.24%
인천	400011 ~ 407753	292	126	23.9%	31.05%	5.92%
대전	300100 ~ 306791	282	132	24.5%	32.27%	7.81%
대구	700010 ~ 711840	335	175	25.4%	34.08%	6.63%
울산	680010 ~ 689930	113	81	28.6%	40.56%	4.05%
강원	200010 ~ 269850	387	283	28.9%	41.04%	5.79%
경기	409350 ~ 487930	966	682	28.4%	40.20%	4.30%
경남	621010 ~ 678970	632	489	29.6%	42.46%	5.85%
경북	712010 ~ 799800	628	599	32.6%	48.46%	5.63%
전남	513810 ~ 556907	469	403	31.0%	45.31%	6.81%
전북	560010 ~ 597860	461	380	30.4%	44.14%	5.80%
제주	690010 ~ 699900	108	43	23.3%	29.94%	6.51%
충남	312700 ~ 360765	324	292	31.7%	46.71%	4.27%
충북	361100 ~ 395900	308	207	27.9%	39.07%	7.61%

용된 결과를 보인 것으로 서울특별시의 경우는 우편번호의 배정 비율은 낮지만 다른 도시보다는 많은 우편번호가 배정되어 있고, 오류교정 비율은 21.09%이다. 인천광역시의 경우에는 우편번호 사용 가능한 범위는 많으나 실제적으로 사용되는 우편번호가 적으며 오류교정 비율은 30.05%이다. 울산광역시의 경우에는 우편번호 사용 가능한 범위가 적고, 실제 우편번호 사용 비율도 낮으나 오류교정 비율은 40.56%이 된다. 이 방법에 의하여 오류검출 및 오류교정을 할 경우에 서울특별시 및 광역시들의 평균 오류교정 비율은 30.39%이며, 기존의 방법보다 평균 13.72%를 추가로 오류교정을 할 수 있게 된다.

각 도의 경우는 우편번호의 배정범위가 넓게 설정되어 있고 실제로 사용되고 있는 우편번호의 비중은 평균 5.81%이다. 각 도의 오류교정의 평균치는 41.51%로 많은 대부분의 오류를 교정할 수 있을 것으로 기대된다. 전국적으로 오류교정의 비율은 평균 24.84%를 추가로 오류교정이 가능함을 확인하였다.



(그림 5) 대구광역시의 오류 교정율

또한, (그림 5)는 대구광역시에서 사용중인 우편번호를 구 영역단위로 구분하여 오류교정율을 보인 것이며, 최소 값과 최대 값을 나타낸 것이다. 또한, 우편번호 배정비율 값이 큰 경우에 오류교정비율이 낮음을 확인할 수 있었으며, 실제 우편번호 사용 양과 사용 가능한 범위가 충분함에도 불구하고, 임의로 우편번호를 부여함에 따라 오류교정패턴이 적게 생성되는 문제를 내포하고 있다. 이러한 오류교정 알고리즘을 적용함에 있어 오류교정을 보다 많이 할 수 있도록 하려면 우편번호 배정방법에 오류패턴생성을 위하여 사용한 체크문자의 분포와 우편번호에 따른 우편물 발생 정도를 고려하여 우편번호를 배정하면 모든 지역을 만족시킬 수 없지만 오류교정 비율을 높일 수 있을 것으로 기대된다. 또한, 바코드문자의 판독의 과정에 소요되는 시간 및 오류교

정 시간이 축소될 수 있을 것으로 기대된다.

6. 결 론

본 논문에서 제시한 우편용 3 of 5 바코드의 오류검출 및 오류교정 방법을 적용할 경우에 21.09%~50%까지 오류를 교정할 수 있으며, 판독시간을 부분적으로 단축시킬 수 있다. 추가적인 오류교정 비율은 적지만 우편물 처리량을 기준으로 산출하기 위하여 35,000(통/시간)을 처리하고 OCR에 의한 우편번호의 인식 비율이 80%, 바코드 판독 95%로 가정하면[11], 오류우편물량은 시간당 하나의 자동처리 장치에서 OCR에 의하여 7,000(통/시간)이 발생되고, 바코드판독에 따른 오류가 5%이므로 350(통/시간)이 발생된다. 따라서, 하나의 자동처리 시스템에서 7,350(통/시간)의 오류우편물(총 우편물량의 21%)이 발생된다. 바코드 오류교정 비율을 40%로 가정하면, 350(통/시간)의 바코드판독 오류우편물 중에서 약 140(통/시간)에 해당하는 우편물을 추가로 처리할 수 있는 효과가 있다. 7,210(통/시간)가량이 오류우편물(총 우편물량의 20.6%)로 구분되게 된다. 오류우편물의 규모를 축소하기 위하여 50%가 고객 바코드 인쇄제도에 의하여 접수되는 우편물이라 가정할 경우에 OCR 인식 과정을 통과한 오류우편물은 3,500(통/시간)이며, 이 우편물 중에서 바코드판독 오류우편물은 175(통/시간)이고, 오류교정 알고리즘을 적용하면 70(통/시간)을 줄일 수 있어 3,605(통/시간)이 발생된다. 그리고, 고객 바코드에 의하여 처리되는 우편물은 17,500(통/시간) 중에서 바코드판독 오류우편물은 875(통/시간)이고, 오류교정 알고리즘을 적용하면 350(통/시간)을 추가로 처리하게 되므로 525(통/시간)의 오류우편물이 발생된다. 즉, 고객 바코드와 OCR 인식 처리의 비중을 1/2로 할 경우에 4,130(통/시간)으로 모든 우편물량은 7,210(통/시간)이므로 기존의 오류우편물량을 기준으로 약 42.72%정도까지 오류우편물이 축소됨을 알 수 있다.

우편물량이 많이 발생하는 경우(예. 크리스마스, 연말)에는 우편물 자동구분처리 시스템의 가동시간이 증가되므로 이에 따른 오류교정 비율이 증가될 것으로 기대된다. 우편번호 영역에 따른 오류패턴 정보를 지역별로 존재하게 될 우편집중국에 적용하여 해당 지역에 따른 오류교정기능 및 패턴 값에 의한 판독방법을 추가할 경우에 판독율을 높일 수 있을 것으로 기대된다.

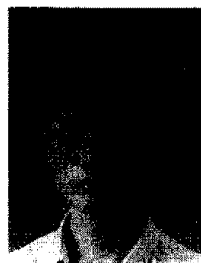
우편번호의 배정방법도 무작위로 배정하기 보다는 오류패턴이 많이 생성될 수 있도록 우편번호 값에 의하여 생성되는 체크분자 값을 기준으로 하여 우편번호를 배정하고 오류교정의 비율을 높이거나 유지하기 위한 노력이 요구된다.

또 다른 문제점으로는 우편번호의 오 기재에 따른 오 구분의 비율이 30~40% 이상에 이르고 있는데, 이러한 문제점은 행정동(예. 가양 1동), 다량 수신처, 사서함에 따라 부여된 우편번호의 사용에 대한 인식 부족으로 대부분 법정동(행정동이 포함된 우편번호, 예. 가양동) 우편번호를 사용하기 때문에 발생되고 있다. 이러한 문제점을 해결하기 위해서는 행정동이 포함된 법정동의 우편번호를 사용하지 않도록 하며, 바코드 인쇄 규격 및 품질을 향상시키기 위하여 개발된 True Type Font를 우편고객에 제공하고 우편 고객에게 홍보를 지속적으로 하여 고객 바코드 인쇄 제도가 효과적으로 수행될 수 있도록 많은 노력이 요구되고 있다.

본 논문에서 제시한 오류교정 알고리즘을 적용하기 위해서는 우편번호 배정 방법의 개선과 변경되어진 우편번호의 오류패턴을 일정한 기간 단위로 적용될 수 있도록 하기 위한 절차를 마련하고 현재 운영중인 LSM에 적용하기 위한 노력이 요구되고 있다.

참 고 문 헌

- [1] 김현수, "우편 업무 바코드 시스템 활용 방안", 우정기술 및 서비스 workshop, pp.43-78. 1996. 6.
- [2] 박문성, 송재관, 우동진, "우편물 자동처리 촉진을 위한 3 of 5 고객 바코드 검증 시스템", 한국정보과학회 추계학술발표대회, pp.496-498. 1998. 11.
- [3] 유중돈, 권인소, "우편 자동화 기술 바코드 인식 기술에 관한 연구", 우정기술 및 서비스 workshop, pp.3-23. 1996. 6.
- [4] 정보통신부 우정국, "고객에 의한 우편번호 바코드 인쇄 제도", 정보통신부, 1998. 11.
- [5] 한국전자통신연구원, "우편 순로 구분의 자동화", 우정기술정보 Letter, 제98-1호, pp.2750. 1998. 4.
- [6] Donald Clysdale, "Canada Post Corporation's Point of Call Identifier," Advanced Technology Conference, Vol.2, pp.771-782, December 1992.
- [7] Hitoshi Uehara, Yoshiaki Nakamura, Masataka Nakagawa, Terutaka Tanaka, Akira Ohzawa, Ichiro Isawa, Hiroshi Miyoshi, "Research on Bar-codes for Mechanized Mail Processing," <http://www.iptp.go.jp/>, July 1995.
- [8] J. Strohmeyer, J. Nice, "Carrier Sequence Bar Code Sorter," Advanced Technology Conference, Vol.2, pp.1061-1074, December 1992.
- [9] KB Redersen, Hans Gartner, Walter S. Rosenbaum, "New Applications and Technology Trend in Postal Service," Advanced Technology Conference, pp.823-836. Vol.2, December 1992.
- [10] Masataka Nakagawa, Hiroshi Miyoshi, "Barcodes for use in Mechanized Mail Processing," <http://www.iptp.go.jp/>, June 1995.
- [11] MPT, "The New Postal Code System in Japan," pp.1-18. February 1998.
- [12] Royal Mail, "Mailsort Customer Barcoding Technical Specification," OXFORD, pp.1-11. April 1995.
- [13] Shoji Watanabe, Shunkichi Isohara, "Simulation Analysis of a New Mail Processing System: Development of a Mail Processing Model--," <http://www.iptp.go.jp/>, August 1995.
- [14] Teruo Takahashi, Iwao Kawahara, Shigeki Toyama, Katsumi Ohsuga, Yoshiaki Nakamura, Ikuo Yamashita, "Research on Mechanized Processing of Large-sized Mail and International Mail," <http://www.iptp.go.jp/>, June 1996.



박 문 성

e-mail : mspark@etri.re.kr

1991년 숭실대학교 전자 및 컴퓨터공학과 졸업(공학석사)

1983년~현재 한국전자통신연구원 우정자동화팀 연구원

관심분야 : Data Capture & Carrier, 우편 자동화 기술, 신호처리, 데이터 통신망



우 동 진

e-mail : djwoo@etri.re.kr

1972년 연세대학교 분헌정보학과
졸업(학사)

1986년 정주대학교 분헌정보학과
졸업(석사)

1976년~1982년 국방과학연구원 신
입기술원

1983년~현재 한국전자통신연구원 우정자동화팀장 책
임기술원

관심분야 : Data Capture & Carrier, 정보검색



김 혜 규

e-mail : hkim@etri.re.kr

1973년 서울대학교 공과대학 응용
물리학과(학사)

1985년 서강대학교 경영대학원 경
영학과(석사)

1994년 서강대학교 공공정책대
학원 정보처리(석사)

1979년~현재 한국전자통신연구원 우정기술연구부장 책
임연구원

관심분야 : 정보산업정책, 멀티미디어, 인식 기술