

바코드에 있어서 체크숫자를 이용한 오류수정과 바코드의 현재와 미래

량식 대학교 전산수학과 김화준
cellmath@gmail.com

바코드의 역사, 그 체크숫자(check digit)를 이용한 오류수정과 바코드의 현재와 미래에 대하여 조사를 하였다. 이 논문을 통하여, 현재 바코드와 관련하여 진행 중인 연구에 대한 소개와 연구방향에 대하여 약간의 제시를 하고자 한다.

주제어 : 바코드, 체크 숫자, 워터마크, 2D 바코드, RFID, DNA 바코드

0. 서론

바코드(barcode)는 굵거나 가는 바(bar)와 스페이스(space)의 조합에 의하여 숫자 또는 특수기호를 광학적으로 판독하기 쉽게 부호화한 것이다([12]). 이 bar들을 이용하여 정보의 수집이나 표현, 그리고 해독을 가능하게 하여 준다. 이 bar와 space들의 조합은 이진수 0과 1의 비트(bit)로 바뀌게 되고, 이것은 바코드 인식장치에 빛의 반사를 이용하여 데이터(data)를 재생 시키며 재생된 데이터를 수집, 전송한다.

이 바코드는 매출정보, 재고관리나 카운터(counter)에서 기다리는 시간의 절약 같은 편리함 때문에 오늘날 모든 제품에 부착되어 있으며, 요즈음은 바코드의 데이터 표현의 확대를 위하여 점(dot)이나 원(circle)으로 만든 1)2차원 바코드(2D barcode)도 많이 사용되고 있다([4]). 이 2차원 바코드는 1980년대 중반부터 사용되기 시작하여 지로공과금(giro)등에 사용되고 있다. 우리나라는 현재 2차원 바코드 4종이 KS규격으로 사용되고 있다.

현재 많이 사용되고 있는 바코드 시스템(system)은 AIDC(Automated identification and data capture)인데, 이것의 장점은 바코드 표를 읽는 속도가 빠르고 정확도가 높다는 것이고, 그것보다 더 좋은 점은 2)RFID(Radio Frequency Identification) 시스템

1) 2D 바코드: X축과 Y축의 양 방향으로 데이터(data)를 배열시켜 평면화한 바코드로 고용량의 데이터를 처리하기 위한 목적에서 개발되었고, 데이터를 구성하는 방법에 따라서 stacked code와 matrix code로 나뉜다.

과 비교해 유지비용이 저렴하고 안전성이 우수하다는 것이다.

그 원리는 누구나가 다 알고 있듯이 바코드 스캐너(scanner)가 바코드를 읽으면 그것이 바코드 판독기(barcode reader)에 전달하여 글씨로 변환되는 원리이다. 물론 이 바코드 판독기는 광학적으로 표현된 바코드 기호(symbol)를 컴퓨터가 수용할 수 있는 디지털 데이터로 변환하는 기능을 가진 장비를 말한다.

이 논문은 바코드의 체크숫자를 이용한 오류수정과 현재 바코드와 관련하여 진행 중인 연구에 대한 소개와 연구방향에 대하여 약간의 제시를 하고자하며, 새로운 사실을 밝힌다기보다는 그 내용을 정리하는데 의미를 두고 싶다.

1. 바코드에 대한 역사

오늘날 사용되는 바코드는 슈퍼마켓(supermarket)의 관리효율을 높이기 위하여 고안되었다. 이 바코드는 1923년 Wallace Flint의 Harvard 대학교의 '슈퍼마켓의 계산자동화'에 대한 논문을 최초로 보는 견해도 있으나, 그의 시스템은 랙(rack)과 천공카드(punched card)를 이용한 단순한 내용이고 경제성이 없어서 이것을 바코드의 기원으로 보기에 는 무리가 있다.

이런 이유로, 바코드의 기원을 Bernard Silver와 Norman Joe Woodland를 드는 것이 타당하다 하겠다. 이 바코드에 대한 연구는 Drexel 대학교의 공학연구소의 대학원생이던 Silver가 연구소의 소장(Dean)과 대형 식품회사의 경영자와의 통화를 우연히 듣게 된 데서 시작되었다. 그 내용은 슈퍼마켓의 카운터(checkout counter)에서 상품정보의 데이터(data)수집에 관한 연구 프로젝트(a research project)에 관한 것이었고, 이때가 1948년이였다. 그러나 이 연구소의 소장은 이 연구에 관심이 없었고, 이 전화통화를 우연히 들은 Silver는 이 내용을 Drexel 대학교의 강사이면서 대학원생이었던 Woodland에게 이야기하게 된다. 물론 Woodland는 이것에 많은 관심을 가지게 되었고, 이 때부터 바코드에 대한 두 사람의 공동연구가 시작이 되었다. 그 후 이 두 사람은 이 주제에 대한 많은 연구를 진행하였다. 처음에는 자외선아래에서 잉크(ink)를 사용하는 방식을 연구하였으나, 잉크의 안정성과 인쇄비용의 비경제성으로 실패하고, 이 실패를 밑거름으로 사운드트랙(soundtrack)과 모스 부호(Morse code)를 사용하는 방

2) RFID(Radio Frequency Identification): 무선 주파수(RF, Radio Frequency)를 이용하여 대상(물건, 사람 등)을 식별할 수 있는 기술로서, 상품에 데이터(data)가 입력된 태그(tag)를 부착하고 여기에 판독기에서 방사된 특정한 주파수가 닿으면 데이터가 전송되어 식별이 이루어지는 방식으로 교통카드, 주차관리, 도서관리, 출입통제용 카드, 동물식별, 하이패스용 카드 등에 응용되고 있음. 인체 이식용 RFID tag가 암의 발병률을 높인다는 연구결과가 있으며[7], 가격경쟁력이 확보된다면 바코드를 대신할 미래의 기술로 전망되고 있음.

식, bar와 space가 원모양의 소형 사격 표적형태를 띠는 황소의 눈 코드(bull's eye code)를 이용한 방식 등 여러 시도가 있었고, 마침내 기능적인 바코드를 개발하게 되었고 이것으로 1949년 10월에 미국 특허를 출원하여 1952년에 등록(US patent no. 2,612,994)이 되었다([15]). 판독기(reader)로는 500 와트(watt) 백열전구와 영화 음향 시스템(movie sound system)인 RCA 935 photo-multiplier tube를 이용하였다.

그 후, Silver는 38세의 젊은 나이로 요절하였으나, Woodland는 이 특허로 인하여 Drexel 대학교에서 자리를 얻게 되고 1992년에는 Bush 대통령에 의하여 'National Medal of Technology'를 받게 된다.

이 특허에 IBM과 Philco가 이들의 특허에 관심을 기울였고, 이 특허는 좋은 조건을 제시한 Philco에 1962년에 넘기게 된다. 이 특허는 다시 RCA(Radio Corporation of America)로 팔리게 되고, RCA는 이 기술이 결실을 맺을 수 있도록 많은 기술적인 개발을 하게 된다. 1960년대 말에 RCA사는 scanner를 개발하여 1972년에 Cincinnati의 Kroger상점에 설치하여 18개월을 시험 운영하여 비용절감과 시스템개선에 효과를 보게 된다.

1970년대 중반에는 미국 슈퍼마켓특별위원회(US Supermarket Ad Hoc Committee)가 결성되어 식료품 업계의 표준코드와 심벌(symbol)을 채택하고자 하였으며, 시험기간을 거쳐 1973년 UPC(Universal Product Code) 심벌을 식료품업계의 표준으로 결정하였다. 이 UPC 바코드가 1974년 Marsh's 슈퍼마켓에서 처음으로 사용하게 되는 것을 시작으로 미국과 캐나다의 슈퍼마켓의 전역으로 확대되었고, 그 효용성에서 큰 성공을 하게 되었다([15]). 이에 자극을 받은 유럽에서도 1976년 EAN((European Article Number) 코드(code)와 심벌을 채택하게 된다. 우리나라는 1988년 EANA(European Article Numbering Association)으로부터 국가코드 880을 부여받아서 사용하고 있다.

1980년대 들어서 이 바코드 기술은 비약적인 발전을 하였고 기술적 경제적으로 그 효용성이 입증되자, 모든 비즈니스(business)에 이 기술이 사용되었다. 이 바코드가 상업적으로 성공을 한 것은 1981년 9월 1일에 미국 안보국이 군부대로 들어오는 모든 물건에 대하여 39개의 코드의 사용을 채택하면서 본격적인 시대로 들어서게 되었으며, 바코드 심벌이 차지하는 면적을 줄이기 위한 노력도 이 시기에 시작이 되었다.

요즈음은 수학을 이용한 체크숫자(check digit)와 에러수정 알고리즘(error correcting algorithm)으로 인하여 바코드판독기의 정확도가 99,999/100,000까지 이르고 있다([15]).

이 분야와 관련된 주요 수학 논문들은 근본적으로 바코드의 디지털(digit)들을 대수적 구조를 가진 군(group)의 원소로 해석하고 있으며, Carlsson([2], [3])와 Ghrist([5],

[10])는 그의 논문에서 바코드를 호모로지(homology) 이론으로 설명하고 있다.

2. 바코드에 있어서 체크숫자를 이용한 오류수정

바코드의 체크숫자의 도입이유는 에러(error)를 인지하고 수정하기위하여 도입되었다. 이는 주민등록증이나 신용카드의 체크숫자의 의미와 같이, 정보가 잘못 읽혀지는 것을 사전에 방지하는 일종의 견제 장치로 표현할 수 있다. 아래의 1)과 2)의 내용은 일반적으로 알려진 평범한 이야기이지만 전체적인 맥락에서 필요하다 생각되어 정리하였으며, 체크 숫자를 이용한 오류수정에 중심을 두고 이 단원을 전개하고자 한다.

1) 바코드에 적합한 숫자의 의미

현재 바코드의 국제 표준은 EAN-13 코드((European Article Number-13 code)인데 13자리의 숫자로 구성되어 있고 다품목 코드, 표준형 B와 의약품 코드의 3종류가 있다.

EAN-13	국가코드	제조업체 코드	상품코드	체크 숫자
자릿수	2-3	5	4	1
부여기관(한국)	880	한국유통물류진흥원	제조업체	컴퓨터에서 자동적으로 계산

<표 1>. EAN-13

위에서 보는 것과 같이 이 코드는 국가코드, 제조업체 코드, 상품코드와 마지막 숫자인 체크 숫자(check digit/ 컴퓨터에서 자동적으로 계산)로 구성이 되어 있다. 이 체크숫자는 앞의 12자리의 숫자에 의하여 결정이 되며, 국가 코드는 1982년 이전에 EAN-international에 가입한 나라는 2자리이고 그 후에 가입한 국가코드는 세 자리가 부여되어 있다.

제조업체 코드와 상품코드의 자릿수는 국가마다 차이가 있으며, 우리나라의 경우는 각각 5와 4자리다. 바코드와 관련된 여러 곳에서 제조업체 코드와 상품코드의 자릿수를 각각 4와 5로 기재되어 있는 곳이 많으나, 제조업체 코드 부여기관인 한국유통물류진흥원의 기준인 5와 4자리를 기준으로 하는 것이 타당하리라 판단된다.

마지막 숫자인 체크 숫자는 scanner에 의하여 잘못 읽혀지는 것을 방지하기위해서 만든 코드로서 modulus 10방식으로 계산이 된다.

2) 바코드에 적합한 체크숫자의 의미

a) EAN-13

이 방법은 앞의 12자리의 숫자 중 홀수 자리(odd positions)의 수들은 그대로 더하

고, 짝수 자리(even positions)에 있는 수들은 더한 후 3을 곱한다. 이 두 값에 체크 숫자까지 더하면 10의 배수가 되도록 구성되어 있다. 여기서, 짝수 번째 자릿수에 곱한 3은 가중치(weight)를 의미한다. 이것은 두 개의 인접한 바코드가 바뀌었을 때, 그 차이를 구별하지 못하는 오류를 피하기 위함이다.

예를 들어, 다음과 같은 바코드 숫자의 경우를 살펴보자.

9780471007104 ... *)

먼저 홀수 번째 자리의 수들을 더하면 $9+8+4+1+0+1=23$ 이고, 짝수 번째 자리의 수들을 더하여 3배를 하면 $3 \times (7+0+7+0+7+0)=63$ 이 된다.

이 2개의 합을 더하면 86이기 때문에 총합이 10의 배수인 90이 되기 위해 마지막에 오는 체크 숫자는 4가 되어야 한다.

위의 예는 서적의 경우이고, 이 서적의 경우에 있어서 표시 방법은 앞의 페이지(page)와 달리 국제표준도서번호(ISBN, International Standard Book Number)에 의하여 결정된다. 자세한 내용은 ISBN에서 다루기로 한다. 그 내역은 다음과 같다.

EAN-13	978	0	471	00710	. 4
내역	서적의 경우 접두어	국가 번호	제조업체 코드	상품코드	체크 숫자

<표 2>. EAN-13의 예

위의 표에서, 접두어는 단행본 서적이면 978을, 연속 간행물이면 979로 표시하며, 이 숫자는 국가번호가 아니면서 국가번호가 할당되었기 때문에, 이것을 종종 '책나라'(bookland)라 부른다([16]).

b) UPC(Universal Product Code)

UPC는 2004년까지 미국과 캐나다에서 사용되었던 코드로서 12자리의 표준형과 8자리의 단축형이 있는데 현재는 사용되고 있지 않다.

12자리 숫자의 UPC의 체크숫자는 12번째 자리의 수이다. 이것은 EAN-13과 비슷한 방법으로 체크숫자를 구성하는데, 짝수 자리의 수들은 그대로 더하고, 홀수 자리에 있는 수들은 더한 후 3을 곱한다. 이 두 값에 체크 숫자까지 더하면 10의 배수가 되도록 해 놓았다.

그러나 이 방식은 현재의 국제표준방식이 아니므로 이 정도의 소개로 그치고자 한다.

c) ISBN

출판물 및 문헌정보 유통을 효율화 하기위하여 각종 도서나 전자출판물에 국제적으로 표준화된 방법으로 고유번호를 부여하며, 보통 책의 뒷면 오른쪽 아래에 표시한다.

이 ISBN은 10자리로 구성되어 있고 5자리의 부가번호를 포함할 수 있으며, 10자리 ISBN의 공간의 부족으로 2005년부터 13자리를 사용하고 있고(본격적으로는, 2007년부터 시행), ISBN-13으로 표시되고 있다. 이 ISBN-13은 EAN-13의 번호를 그대로 ISBN의 번호로 사용하기 때문에 바코드와의 통합으로 생각할 수 있으며, 체크숫자는 바뀌지만 이미 존재하는 바코드 체계를 그대로 활용할 수 있다는 장점이 있다([16]).

앞에서 예로 들은 바코드번호 9780471007104위에는 ISBN 0-471-00710-2(90000)가 적혀있다.

그 내역은 다음과 같다.

ISBN	0	471	00710	-	2	90000
내역	국가번호	발행자번호 (출판사번호)	서명식별번호		체크숫자	부가번호(내용분류/발행형태/대상독자)

<표 3>. ISBN의 예

ISBN이 위 자릿수에 의하여 표시되기 어려울 때는 하이픈(hyphen, -)을 이용하여 구별한다. 국가번호는 반드시 한 자리만을 사용하지는 않으며, 출판사번호는 각 국가의 국가별 ISBN의 기관에서 할당을 담당하고 있고, 항목은 각 출판사에서 임의로 정할 수 있다. 국가번호는 다음과 같이 결정된다.

언어권	영어권	프랑스어권	독일어권	일본	러시아	중국	한국	멕시코
국가번호	0	2	3	4	5	6	89	970

<표 5>. ISBN의 국가번호

이 ISBN은 바코드로 나타나 있지 않으므로 scanner로 직접 읽을 수가 없기 때문에 EAN-13으로 바꾼 바코드를 함께 표시하기로 1980년 IAAEAN(International Article Association EAN)과 국제 ISBN 관리기구와의 협정에 의하여 정하여졌다. 그 방법은 앞에서 언급한바와 같이, 단행본 서적이면 978을, 연속 간행물이면 979를 맨 앞에 표시하고 ISBN의 9자리 숫자를 그대로 적은 다음 체크 숫자를 붙이는 형태이다. 그런 까닭에, 위의 ISBN 0-471-00710-2은 EAN-13의 9780471007104로 표시된다. 여기서 체크 숫자는 EAN-13에서 정한 방식으로 결정이 된다.

이 ISBN의 체크숫자는 다음과 같은 방식으로 정해져 있다. 앞의 9자리에 차례대로 10부터 2까지를 곱한 후 더한 후, 여기에다가 체크숫자를 더하면 11의 배수가 되게 구성되어 있다. 예를 들어, 다음과 같은 경우의 ISBN을 보자.

0-471-00710-2

인데, 먼저 체크숫자를 제외한 앞의 9자리의 수를 위와 같은 방법으로 계산하면

$$0 \times 10 + 4 \times 9 + 7 \times 8 + 1 \times 7 + 0 \times 6 + 0 \times 5 + 7 \times 4 + 1 \times 3 + 0 \times 2 = 130$$

이다.

이 숫자 130에 체크숫자를 더한 값이 11의 배수인 132가 되어야하므로, 마지막에 위치한 체크숫자는 2가 되어야한다. 이 계산은 modulus 11의 방식으로 계산이 되므로 체크숫자가 10인 경우가 생기는데, 이 경우, 10을 X로 표시하기로 정하여져 있다.

우리나라의 경우 ISBN은 국립중앙도서관 한국문헌번호센터에서 부여하며, 이 ISBN은 OCR(Optical Character Recognition) B문자로 표시하고, 부가번호는 바코드로 표시한다.

이 밖에도 EAN-14 코드, 쿠폰코드, SSCC(Serial Shipping Container Code), UCC/EAN-128 코드가 있다. 여기서 UCC는, uniform code council의 약자이다.

위와 같은 이유로 바코드를 이용한 작업은 체크숫자 덕분에, 오류가 생기는 경우를 피할 수 있다.

3) 바코드의 수식 표현

a) EAN-13

국제표준인 EAN-13 코드를 기준으로 바코드의 체크숫자를 구성했을 때를 기준으로 표현해보자. 바코드의 각 자리수를 $x_1 x_2 \cdots x_{12} c$ 로 나타냈을 때, 체크숫자 c 는

$$c = 10n - \left[\sum_{k=1}^6 x_{2k-1} + 3 \sum_{k=1}^6 x_{2k} \right] \pmod{10}$$

(n 은 정수)으로 표시된다.

2)의 a)의 예에서 []안의 숫자는 $23 + 63 = 86$ 이므로 체크숫자는 $90 - 86 = 4$, $100 - 86 = 14$, $80 - 86 = -6$, ... 등과 같이 나타난다. 따라서 10을 법(modulus)으로 하였을 때, 체크숫자 c 는 4가됨을 알 수가 있다.

b) ISBN

바코드의 각 자리수를 $x_1x_2 \cdots x_9c$ 로 나타냈을 때, 체크숫자 c 는

$$c = 11n - \sum_{k=1}^9 x_k(11 - k) \pmod{11}$$

(n 은 정수)으로 표시된다.

4) 체크숫자를 이용한 오류 수정

바코드는 일부가 손상되더라도 다른 부분을 읽어서 정상적인 정보를 얻을 수 있는 장점을 가지고 있다. 1차원 바코드는 복원을 위하여 심벌(symbol)과 별도로 아래에 육안으로 인지할 수 있는 정보(숫자)를 표시하고, 2차원 바코드는 오류정정 기능을 자체적으로 가지고 있어서 오류 수정이 자동으로 이루어지나, 얼룩이나 더러움에 약한 약점을 가지고 있다([13]).

여기에서는 간단한 예로 1차원 바코드에서 한 개의 바코드(a single bar)가 손상되었을 때와 두 개가 손상되었을 경우에, 체크숫자를 이용한 오류 수정으로 정상적인 정보를 얻는 경우를 생각해보기로 한다.

A) 한 개의 숫자가 손상된 경우

a) EAN-13

먼저 공식을 생각해보자.

경우 1) l 번째 자릿수(홀수 번째)가 손상되었을 경우;

$$\text{손상된 한 개의 바코드 } x = 10n - \left[\sum_{k=1}^6 x_{2k-1} - x_l + 3 \sum_{k=1}^6 x_{2k} + c \right] \pmod{10},$$

여기서 c 는 체크숫자이다.

예를 들어, 앞의 예 9780471007104에서 일곱 번째 숫자 1이 손상되었다고 가정하자. 그리고 숫자를 978047 x 007104로 표시하자.

위의 공식을 이용하면,

$$x = 10 \times 9 - (22 + 3 \times 21 + 4) = 1$$

임을 쉽게 알 수 있다.

경우 2) l 번째 자릿수(짝수 번째)가 손상되었을 경우;

손상된 한 개의 바코드를 x 라 할 때,

$$3x = 10n - \left[\sum_{k=1}^6 x_{2k-1} + 3 \sum_{k=1}^6 x_{2k} - 3x_l + c \right] \pmod{10},$$

여기서 c 는 체크숫자이다.

앞의 예 9780471007104에서 여섯 번째 숫자 7이 손상되었다고 가정하자. 그리고 숫자를 $97804x1007104$ 로 표시하자.

위의 공식을 이용하면,

$$3x = 90 - (23 + 3 \times 14 + 4) = 21$$

이고, 그런 까닭에 $x = 7$ 을 쉽게 얻을 수 있다.

b) ISBN

앞의 예 0-471-00710-2에서 l 번째 자리수가 손상되었다고 가정하고 손상된 한 개의 바코드를 x 라 하면,

$$(11 - l)x = 11n - c - \sum_{k=1}^9 [x_k(11 - k)] + x_l(11 - l) \pmod{11},$$

여기서 c 는 체크숫자이다.

예를 들어 앞의 예에서 세 번째 자리가 손상되었다고 하고, 다음과 같이 표시하자.

$$0-4x1-00710-2$$

위의 공식을 이용하면,

$$8x = 11n - 2 - (0 \times 10 + 4 \times 9 + 1 \times 7 + 0 \times 6 + 0 \times 5 + 7 \times 4 + 1 \times 3 + 0 \times 2)$$

로 표시되고, 이것을 계산하면 $11n - 76$ 이 8의 배수이어야 하므로, $n = 12$ 이고 그런 까닭에 $x = 7$ 이라는 것을 알 수가 있다.

B) 두 개의 숫자가 손상되었을 경우

EAN-13이나 ISDN이나 찾는 논리는 비슷하므로 EAN-13의 경우에 한하여 손상된 숫자를 찾는 것을 생각해보자.

a) 홀수 번째 자리수 두개가 손상되었을 경우;

손상된 두 개의 바코드를 x 와 y 라 할 때,

$$x + y = 10n - \left[\sum_{k=1}^6 x_{2k-1} - (x + y) + 3 \sum_{k=1}^6 x_{2k} + c \right] \pmod{10}$$

로 표시된다. 여기서, c 는 체크숫자이다.

예를 들어, 앞의 예 9780471007104에서 셋째자리와 다섯째자리의 수가 손상되었을

경우에 그 수를

$$97x0y71007104$$

에 의하여 표시하자. 앞의 공식에 의하여,

$$11 + x + y + 3 \times 21 + 4 = x + y + 78 \quad (x + y \leq 18)$$

이 10의 배수가 되어야하므로, $x + y$ 는 2 또는 12가 되어야한다. 이 경우 합이 12가 되는 경우의 수는 10가지이나, 우리가 이미 국가정보나 그 외의 이미 알고 있는 정보가 있기 때문에 손상된 숫자를 찾는 것이 거의 가능하다.

b) 짝수 번째 자릿수 두 개가 손상되었을 경우;

$$3(x + y) = 10n - \left[\sum_{k=1}^6 x_{2k-1} + 3 \sum_{k=1}^6 x_{2k} - 3(x + y) + c \right] \pmod{10}$$

에 의하여 표시되어진다.

c) 홀수와 짝수 자릿수가 각각 한 개씩 손상되었을 경우;

손상된 홀수 번째의 숫자와 짝수 번째의 숫자를 각각 x, y 라 하면

$$x + 3y = 10n - \left[\sum_{k=1}^6 x_{2k-1} + 3 \sum_{k=1}^6 x_{2k} - (x + 3y) + c \right] \pmod{10}$$

에 의하여 표시되어지고, a)와 같은 방법에 의하여 손상된 숫자를 찾는 것이 거의 가능해진다.

세 개의 숫자가 손상되었을 경우에는 이미 알고 있는 정보와 위의 방법에 의하여 접근을 하여 손상된 숫자를 찾는 것이 어느 정도 가능하다.

5) 바코드의 해독 원리

바코드의 해독 방법은 바코드 리더(barcode reader)가 바코드를 읽음으로써 파형이 발생하고 이를 컴퓨터가 해독 가능한 데이터 형태로 변환한 다음, 모니터(monitor)에 글자로 표시(display)되는 원리이다.

이 바코드 리더는 스캐너(scanner)와 디코더(decoder)로 구성되어 있다. 이 바코드 리더가 심벌(symbol)을 해독하여 데이터로 변환되는 과정은, scanner가 스캔(scan)할 때 심벌의 bar와 space는 빛의 반사율이 다르므로, scanner의 포토 센서(photo detector or sensor)에서 이 빛의 양을 감지하여 그 크기에 따라 전기신호를 발생하게 되고, 이 신호를 증폭하여 아날로그(analog or analogue) 신호(전압)를 생성시킨다. 그 후 이 전압이 상한(supremum)에 도달하면 3)래치(latch)가 동작하여 High상태를 유지

하고, 하한(infimum)에 도달하면 latch가 Low상태로 변하게 만든다. 이렇게 생성된 디지털 신호는 decoder 내의 디지털 신호처리 과정을 거쳐 그 비트 패턴(bit pattern)에 할당된 데이터 값을 생성하게 된다. 즉, 레이저 타입(laser type)의 아날로그 파형은 빛(light source)의 반사에 의한 강약을 포토 센서로 받아 전기적인 신호로 변환하고, 디지털(digital) 파형은 아날로그의 신호를 0과 1로 변환하는 4)원리이다.

결론적으로, 이것은 디코드 회로(decode circuit) 및 변환 소프트웨어(transformation software)에 의하여 컴퓨터가 해독 가능한 데이터 형태로 변환된 뒤 일반적인 데이터 전송방식에 의하여 컴퓨터로 전송 처리되는 원리이다.

3. 바코드의 현재와 미래

1) 워터마크(watermark)

워터마크란 어떤 파일(file)에 대한 저작권정보등을 식별할 수 있도록 디지털이미지나 오디오 및 비디오파일에 삽입한 비트 패턴이나 디지털심벌로 지적 재산을 보호하기 위해서 삽입한 심벌을 말한다. 이것은 어느 정도는 볼 수 있게 만드는 것이 일반적이나 전혀 보이지 않도록 설계된 것도 있다. 이 워터마크를 삽입할 때는 파일 전체에 골고루 이 마크가 퍼져있어야 하며, 알고리즘에 의한 축소 같은 그 파일의 일반적인 변경에 대해 충분히 견딜 수 있도록 견고하게 삽입되어야 한다([14]).

이 워터마크를 삽입하는 이유는 바코드가 복사가 되는 단점을 가지고 있기 때문이다. 이 문제점을 극복하기 위하여, 현재 바코드에 제조사의 고유한 워터마크신호를 삽입하여, 이 바코드가 제품의 제조사에서 생성한 바코드인지를 판별할 수 있게 하는 연구가 진행되고 있다. 이미 이 연구는 상당부분 특허가 등록되어 있고([1]), 경제성과 워터마크의 삽입으로 인하여 바코드의 원래 기능에 어떤 영향을 미치는가 하는 안전성에 관한 검토가 진행되고 있고, 또 일부는 이미 상품화되고 있다.

2) RFID

앞의 주석에서 잠시 언급한 바와 같이, RFID는 상품에 5)태그(tag)를 부착하고 여기

3) 래치(latch): 어떤 입력에 대한 출력 상태를 다른 입력이 있을 때까지 유지하는 논리 회로

4) 디코더의 데이터 전송원리: 빛이 반사되는 양에 의하여 이진수체계인 computer로 데이터를 보냄. 즉, 바코드의 검은 부분은 빛을 적게 반사하므로 이진수 1로, 스페이스인 흰 부분은 0으로 데이터가 전송됨

5) 태그(tag)의 작동원리: 기본적으로 안테나(antenna), 배터리(battery)와 간단한 메모리(memory)가 들어있으며, 판독기가 보내주는 전파로 전기를 만들어 저장된 정보를 무선으로 전송하는 원리이다.

에 판독기에서 보낸 특정한 주파수가 닿으면 데이터가 무선으로 전송되어 식별이 이루어지는 방식으로, 안전성과 경제성에서 바코드와의 경쟁성이 확보된다면 바코드를 대신할 미래의 기술로서 6)전망되고 있다. 현재 tag의 가격이 점점 떨어지는 추세이고, 이 RFID의 장점은 통상적으로 10m이내의 상품을 무선으로 판독하기 때문에 카운터에서의 대기시간을 거의 0에 가깝게 줄일 수 있는 효과를 가지고 있다. 아직까지는 이 기술이 투자대비효과(ROI, Return on Investment)를 기업에 주고 있지는 못하지만, 기업 프로세스(process) 혁신차원에서 올해 Ford 자동차의 Land Rover가 공급업체로부터 자동차 부품의 배송을 최적화하기위한 RFID pilot project를 올 1월에 시작하였다([6]).

이미 Bangkok과 우리나라에서 개(dog)에 이식용 tag를 부착하여 개의 관리를 하겠다는 보도가 있었다. 그러나 [7]에서 언급 하고 있는 것과 같이 인체이식용 tag가 암의 발병률을 높인다는 부정적인 연구결과가 있는 만큼, 귀걸이 같은 이식과는 7)다른 방식을 생각한다면 하는 신중한 접근이 필요하다고 생각된다.

3) DNA(Deoxyribonucleic acid) 바코드

a) 생명의 백과사전 프로젝트(project)

생물체는 자신의 고유 DNA를 바코드로 이용할 수 있는데, 동물의 경우, 이의 역할을 하는 유전자가 세포 소기관인 미토콘드리아에 들어있는 시토크롬산화수소 I 유전자(Cytochrome Oxidase I or CO I)이다. 이 유전자는 모계유전을 하므로 대를 거듭해도 종 내에서 잘 보존되는 특징이 있다([11]).

한편 동물이 아닌 식물, 균류 그리고 해조류 등의 바코드는 동물의 CO I처럼 명확히 구별되는 유전자(DNA 바코드 표지)가 무엇인지 아직까지 과학자들 사이에 의견통일이 되지 않았으나, 식물의 경우 올해 1월에 열린 국제 DNA 바코드협회의 학술대회에서 고려대 김 기중교수가 개발한 DNA 바코드 표지가 국제표준으로 채택되었다.

2004년 11월에 캐나다(Canada) Guelph 대학교의 Paul Herbert 교수는 high-volume DNA bar coding에 대한 연구비로 3백만 달러(3 millions dollar)를 Gordon and Betty Moore Foundation으로부터 받아서 이것에 대한 연구를 진행하고 있다. 이 연구내용은 현재 존재하고 있는 동식물에 대한 데이터베이스(database)를 구축하는 것이 포함된다. 이 연구를 통하여, 현재 구분되는 1,200,000개의 동식물 유전자를 10,000,000개까지 식별가능하게 되리라 판단되며, 어느 동식물의 한 조각을 가지고 그 전체에 대한 정보를 얻을 수 있게 되리라 전망되고 있다.

6) IT 시장조사기관인 Gartner社는 RFID 기술 거래로 인한 전 세계 수익이 금년에 약 12억 달러로 작년 대비 약 31%의 성장과, 중국의 금년도 RFID 시장의 매출을 140억 달러로 전망([6]).

7) 목걸이나 팔찌는 강아지의 성장에 따라 상해를 주기 때문에 피해야 함.

b) 나노 DNA 바코드(NDB)

DNA는 유기물이기 때문에 열이나 독성물질에 약하다는 단점을 가지고 있으므로, 8)염기 서열의 조작으로 정보가 저장된 DNA를 나노 캡슐(nano capsule)에 담아서 제품에 뿌리거나 바르는 형태의 바코드를 말한다. 이 바코드는 눈에 안 보이고 극소량만 사용해도 기존바코드에 비교하여 6.3×10^{15} 배 많은 정보를 담을 수 있으나([11]), 아직은 경제성과 나노 입자(nano particle)가 환경이나 인체에 미치는 9)영향을 모르기 때문에 상품화단계까지는 이르지 못하고 있다.

이 바코드 기술은 서울대 최진호 교수연구팀(현재 이화여대)에 의하여 처음으로 2004년 학계에 발표되었다.

c) DNA 바코드 시스템의 기대효과

이 바코드 시스템은 소비자가 필요한 모든 정보를 담고 있기 때문에 불법 정보조작이 불가능해지고, 그런 까닭에, 유통시장이 투명해지며 전염병의 전파경로를 파악하는데 유용하게 쓰일 수 있다. 즉, 쉽게 말해서 지폐에 발라서 위조지폐를 방지할 수 있고 원산지의 위조를 원천적으로 불가능하게 만들어줄 수 있다.

4. 연구방향에 대한 약간의 제시

바코드 기술은 소비자가 계산대에서 기다리는 시간을 절약해줄 뿐만 아니라, 판매자의 상품의 도난 방지나 공급자의 상품분석과 재고관리에 많은 도움을 주고 있다. 바코드와 관련된 연구방향에 대하여 약간의 제시를 하고자 한다.

첫째, 스캐너가 바코드가 읽은 후 디코드에서 신호로 처리할 때, Wavelet De-Noising([8])을 이용하여 noise를 제거하여 그 신호가 읽을 수 있는 범위를 확대시키는데 이 부분에서 우리 수학자에게 필요한 연구가 있으리라 판단된다.

둘째, 현재 바코드를 만드는 비용에 비교하여 경제적이면서도 정보용량의 확대를 가져올 수 있는 새로운 바코드 시스템(barcode system)에 어떤 수학적 도구를 이용하는 것이 가장 효율적이고 경제적인지 연구하는 것도 우리의 몫이라 여겨진다.

2차원 바코드는 그 종류가 현재 수십 종류가 사용되고 있으며 대표적인 data

8) 염기: DNA를 구성하는 adenine(A), guanine(G), cytosine(C)과 thymine(T)이 있으며, 이들의 배열에 의하여 이중나선구조인 DNA가 된다.

9) 현재 나노기술의 위험성에 대한 많은 연구가 진행 중이며 미국은 정부차원의 연구가 진행 중인 것으로 알려져 있다.

matrix는 인치(inch)당 이론상 500만 글자를 인식할 수 있다고 하나, 현실적으로는 500개 정도를 사용하고 있다고 한다([13]). 현재 점(dot)의 굵기로 구별되는 이 matrix를 3×3 행렬의 숫자로 통일하면, 잉크(ink)의 절감효과와 고용량의 데이터를 프린트(print)하는데 곤란한 레이저 프린터(laser printer)의 사용을 가능하게 하여 주는 효과를 가져 올수 있다. 0부터 9까지의 수를 사용한다면, 경우의 수도 10^9 이므로 모든 데이터를 충분히 표현할 수 있다고 판단되며, 복사가 용이한 단점은 어차피 현재의 data matrix도 마찬가지이므로 별 문제가 없을 듯 보이며, 이 문제는 워터마크에 의하여 처리하면 될 것이다. 만약 자릿수의 제한 같은 조건이 있어서, 고용량의 데이터의 표현에 적합하지 않다면 4×4 행렬이나 각 행렬의 성분을 0부터 99까지의 두 자릿수에 의하여 표시하면 부품 번호, 용량이나 일련번호 같은 데이터의 표현에 문제가 없으리라 판단된다. 다만 전제조건으로는 숫자를 빠른 시간에 식별할 수 있는 경제성이 있는 판독기의 개발이 필요하며, 미래에 바코드를 대신하리라고 전망되는 RFID와의 경쟁력이 확보되어야 하겠다.

계속해서 생겨나는 새로운 바코드 시스템에서 우리 수학자들의 역할이 중요한 부분을 차지하기를 기대해본다.

감사의 글 적절한 조언을 주신 심사위원들께 감사를 드립니다.

참고 문헌

1. 이홍구, 김홍신, 이충훈, 바코드 워터마킹 방법, 한국특허(039632), 2003.
2. E. Carlsson, G. Carlsson, and V. de Silva, *An algebraic topological method for feature identification*, Intl. J. Computational Geometry and Applications, 291-314(2006).
3. G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas, *Persistence barcodes for shapes*, Intl. J. Shape Modeling, 11(2005), 149-187.
4. D. E. Gabrielson, *1D finder pattern for 2D bar codes*, US Patent(6102289), 2000.
5. R. Ghrist, *Barcodes: The persistent topology of data*, Amer. Math. Soc. 2007 Current event Bulletin (2007), 1-15.
6. KISTI, 금년도 글로벌 RFID 시장 전망, 글로벌동향브리핑(GTB), 2008-02-28,

-
- <http://www.yeskisti.net/yesKISTI/Briefing/Trends/View.jsp?ct=TREND&clcd=J&clk=&lp=TM&gotoPage=1&cn=GTB2008020664>
7. KISTI, RFID 칩으로 인한 암 유발 가능성, 글로벌동향브리핑(GTB), 2007-09-15, <http://www.yeskisti.net/yesKISTI/Briefing/Trends/View.jsp?cn=GTB2007090348&SITE=KLIC>.
 8. D. Madej, *Matching Wavelets to Signals*, Speckle Noise Filtering in a Laser Bar Code Scanner, Symbol Technologies, 1-24.
 9. J. P. Ruppert, R. C. Fish, T. A. Yap and R. M. Ames, *Portable RFID Tag and Barcode Reader*, US patent(5640002), 1997.
 10. V. de Silva and R. Ghrist, *Coverage in sensor networks via persistent homology*, to appear, Alg. & Geom. Topology, 2006.
 11. <http://blog.chosun.com/blog.log.view.screen?blogId=25920&logId=2448253>
 12. <http://member.okgiro.com/index.jsp>
 13. http://tlashford.com/frames/basics_datamatrix/basics_datamatrix.htm
 14. <http://www.tta.or.kr/Home2003/library/ttasSearch.jsp?rn=1>
 15. <http://en.wikipedia.org/wiki/Barcode>
 16. http://en.wikipedia.org/wiki/International_Standard_Book_Number

Error Correcting using the Check digit on Barcode, and the present and future of Barcode

Dept. of Computational Mathematics, Rangsit University **Hwa Joon Kim**

Barcode technology is becoming an essential tool for every companies, and this makes help us to gain time, analysis of goods, an inventory control, a prevention of burglar and so on.

In this paper, we have treated about the history of barcode, its error correcting using the check digit, and the present and future of barcode.

We wish roles of mathematician on new barcode system which it bring on an economical efficiency and stability.

Key words : barcode, check digit, watermark, 2D barcode, RFID, DNA barcode

2000 Mathematics Subject Classification : 01A99

접수일 : 2008년 3월 16일 수정일 : 2008년 4월 23일 게재확정일 : 2008년 5월 7일