DBpia

HPV DNA 칩의 영상처리 알고리즘 구현

김 종 대**, 연 석 희**, 이 용 업**, 김 종 원*

Implementation of the Image Processing Algorithm for HPV DNA chip

Jong-Dae Kim, Seok-Hee Yeon, Yong-Up Lee, Jong-Won Kim*

요 약

본 논문은 여성 자궁경부암의 원인이 되는 인두유종 바이러스(papillomavirus)를 진단하기 위한 HPV-DNA 칩의 영상으로부터 탐촉자(probe)의 위치를 찾아내고 각 탐촉자에서의 교잡반응(hybridization) 여부를 결정하기 위한 영상처리 방법의 구현에 관한 것이다. HPV-DNA 칩은 22종의 HPV를 알아내기 위한 탐촉자들과 환자 샘플과 항상 반응하는 마커들로 구성되어 있다. 칩 영상에서 탐촉자 들의 위치는 마커와 상대적으로 고정되어 있어서 도터(dotter)와 스캐너(scanner)의 정밀도에 따른 오차만을 갖는다. 한편 각 탐촉자는 진단신뢰도를 높이기 위하여 4번 씩 인쇄된다. 본 논문은 마커들간의 상대거리와 탐촉자의 중복 인쇄라는 사전정보를 템플릿 정합방법과 결합하여 안정적으로 마커를 찾고, 교잡반응 여부를 정확히 분류하는 방법을 제시한다. 정규화된 상관도를 정합도(matching measure)로 채택하여, 마커들의 상대적 위치에 대하여 평균을 취하면 안정적으로 마커를 찾을 수 있다는 것을 실험을 통하여 보인다. 또한 이미 계산된 정합도를 특징(feature) 값으로 사용하여 4개의 중복 탐촉자에 대한 평균 특징 값을 분류(classification)하면 탐촉자의 교잡반응 여부를 정확히 알아낼 수 있다는 실험 결과를 보인다.

Keywords: HPV, DNA chip, prior knowledge, template matching, normalized correlation

ABSTRACT

This paper addresses an image processing technique for the human papillomavirus (HPV) DNA chip to discriminate whether the probes are hybridized with the target DNA. HPV DNA chip is designed to determine HPV geno-types by using DNA probes for 22 HPV types. In addition to the probes, the HPV DNA chip has markers that always react with the sample DNA. The positions of probe-dots in the final scanned image are fixed relative to the marker-dot locations with a small variation attributable to the accuracy of the dotter and the scanner. The probes are quadruplicated to enhance diagnostic fidelity. Prior knowledge including the marker relative distance and the replication information of probes is integrated into the template matching technique with normalized covariance measure. It was demonstrated that the employment of both of the prior knowledges can be accomplished by simply averaging the template matching measures over the positions of the markers and probes. The resulting proposed scheme yields stable marker locating and probe classification.

* (주)바이오메드랩, ** 한림대학교 정보통신공학부
논문번호 : 030108-0317, 접수일자 : 2003년 3월 17일
※이 논문은 2003년도 한림대학교 학술연구 조성비에 의하여 연구되었음.

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

Ⅰ. 서 론

DNA 칩(chip) 기술은 생체 내 유전자의 교잡반응 형태(hybridization pattern)를 분석하거나 돌연변이가 발생된 특정 유전자를 대량으로 탐색할 수 있어서 이를 신약개발, 유전병의 진단, 그리고 SNP(Single Nucleotide Polymorphism)등의 분석에 광범위하게 사용되고 있다.

DNA 칩이란 유전자 검색용으로서 엄청나게 많은 종류의 DNA를 고밀도로 붙여놓은 생화학 반도체이다. 즉 수cm 정도의 좁은 면적의 금속이나 유리 표면에 수백 개에서 수만 개 정도의 올리고 뉴클레오티드(oligonucleotide)나 cDNA를 고 밀도로 부착시키고, 이들 DNA와 상보적으로 교잡반응(hybridization)하는 유전자를 고속으로 분석하는 장치이다. 이러한 DNA 칩이 대체 할 수 있는 기존의 대표적인 유전공학 방법으로는 Southern과 Northern blot, 돌연변이 검색 그리고 DNA sequencing 등이 있다. 이와 같은 기존의 방법들과 가장 큰 차이점은 동시에 많은 양의 유전자를 빠른 시간 안에 검색할 수 있다는 것이다.

DNA 칩은 붙이는 유전물의 크기에 따라 cDNA 칩과 올리고 뉴클레오티드 칩으로 구분할 수 있다. cDNA 칩은 최소한 500 염기쌍(base pair) 이상의 유전자가 붙어져 있고, 올리고 뉴클레오티드 칩에는 15-25개의 염기들로 이루어진 올리고 뉴클레오티드가 놓여있다. 본 논문은 올리고 뉴클레오티드 칩의 일종인 HPV-DNA 칩의 자동 분석을 위한 영상처리 알고리즘에 관한 것이다.

HPV DNA 칩은 여성 자궁 경부 암의 주원인이 되는 인두유종 바이러스(Human Papilloma Virus, HPV) 감염을 진단하기 위하여 개발된 것으로 현재 여러 곳에서 임상 중에 있다[1]. HPV DNA 칩의 구조를 살펴보면 그림 1과 같다. 1개의 칩은 환자 4명을 처리 할 수 있도록 4개의 chamber로 구성되어 있으며, 각 chamber는 진단의 신뢰도를 높이기 위해서 2개의 똑같은 탐촉자 집합(probe set)으로 구성되어 있다. 한 개의 탐촉자 집합은 4개의 마커와 22쌍의 HPV 유형의 올리고 뉴클레오티드 탐촉자들로 구성되어 있다. 환자의 표적(target) DNA는 PCR 증폭되고, 형광으로 라벨이 붙여진 후에 탐촉자에서 교잡반응을 일으킨다. 이후 몇 가지 후 처리를 거치고, 형광 분석기로 얻은 영상에서 형광으로 발현된 탐촉자를 찾음으로써 표적 DNA를 알 수 있는 것이다.

HPV DNA 칩 영상을 자동으로 분석하기 위하여, 각 chamber안의 마커를 먼저 찾은 후, 마커를 참조로 하여 발현된 각 탐촉자의 위치를 찾아 교잡반응 여부를 판단하는 일련의 영상처리 과정이 필요하다. 본 논문에서는 자동분석을 위하여 두 가지 문제를 제시한다. 첫째는 각 탐촉자 집합에서 4개의 마커를 찾는 것이고, 둘째는 각 탐촉자가 교잡반응을 하였는지 정확하게 판단하는 문제이다.

DNA 칩 영상은 불규칙한 배경을 갖거나, 교잡반응을 포함하는 후 처리 과정동안 비전문적인 사용자로 인해서 많은 잡음이 발생할 수 있다. 그림 2는 불규칙한 배경과 많은 클러스터 잡음(cluster noise)을 포함하고 있는 영상의 예이다. 칩 영상은 불규칙한 배경과 클러스터 잡음 때문에 일반적인 이진 영상처리로는 마커를 찾는데 부적당 하며, 수직 또는 수평 적사(projection) 방법을 적용 하기에는 마커나 탐촉자의 수가 너무 적다. 여러 가지 방법을 실험한 후, 마커를 찾기 위해서 템플릿 정합법(template matching method)을 적용하였다[3, 5]. 템플릿 정합법은 국소 연산자에 해당 하기 때문에 전체 영상에서 마커의 위치를 찾기 위하여 4개의 마커가 고정된 거리에 있다는 사전 정보를 이용할 수 있다[3]. 사전 정보를 템플릿에 융합하여 4개의 마커를 포함하는 큰 템플릿를 고려할 수도 있고, 한 개의 마커만를 포함하는 작은 템플릿으로 정합도(matching measure)를 구한 후에 적용할 수도 있다. 정합도로는 차이합(sum of difference), 상관도(covariance), 모폴로지 연산자(morphological operator)와 같은 방법 등을 사용할 수 있다. 이중에서 공분산(covariance) 이 가장 널리 사용되고 있으며[2, 3 ,4], 이것은 일반적인 공분산(Ordinary covariance)과 정규화 된 공분산(Normalized covariance)의 2가지 식으로 표현될 수 있다. 본 논문에서는 공분산을 사용했으며, 일반적인 공분산[2]과 정규화 된 공분산 방법을 각각 적용하여 실험하고 비교 분석하였다.

탐촉자를 분류하는 데 있어서도 같은 유형의 탐촉자가 4번 중복하여 인쇄되었다는 사전 정보를 이용하였다. 선술한 바와 같이 각 탐촉자 집합에 같은 유형의 탐촉자가 두번씩 인쇄되어 있으며, 한 chamber당 두개의 탐촉자 집합이 있으므로 한 환자당 한 유형의 HPV를 4번 실험하는 효과가 있다. 이와 같이 동시에 여러 번 실험할 수 있도록 탐촉자를 중복하여 인쇄하는 것은 보편적인 방법이다[9].

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

2장에서 템플릿 정합방법과 사전정보를 융합하여 마커를 찾고 탐촉자를 분류하는 알고리즘을 제안한다. 3장에서는 실제 환자 표본으로부터 얻은 칩 영상에 대한 결과를 제시하고 4장에서 결론을 맺는다.

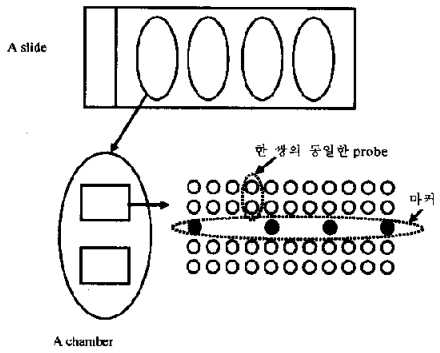
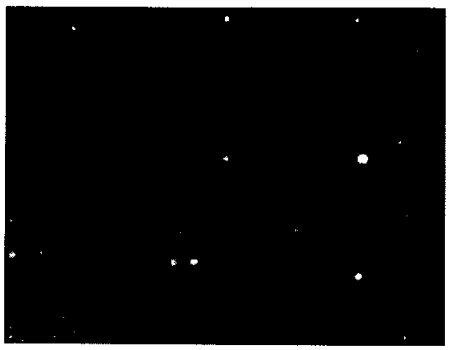
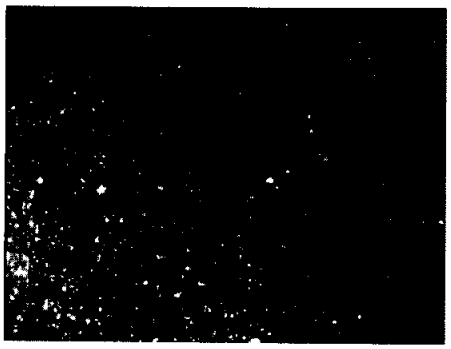


그림 1. HPV DNA chip의 구조



(a) 불규칙한 배경을 갖는 영상



(b) 부적당한 처리로 잡음이 많은 경우

그림 2. 배경이 불규칙하거나 잡음이 많은 영상의 예.

Ⅱ. 실험 및 방법

템플릿에서 물체에 해당하는 부분은 원래 도터의 모양과 같은 원형과 연산을 빠르게 할 수 있는 사각형 모양의 두 가지 모양을 채택하였다. 칩 영상에서는 탐촉자나 마커가 도터의 모양과 같을 것이라는 가정과 구분하고자 하는 모양이 원형의 물체가 있는 지의 여부만을 판단하면 되기 때문에 사각형의 템플릿도 원형 물체의 존재 유무를 판단하기에 무리가 없다고 판단된다. 한편 사각형 템플릿의 경우 빠른 연산 알고리즘을 적용할 수 있는 장점이 있다[13].

템플릿 정합도는 물체와 물체 주변과의 차이를 정량화하기 때문에 탐촉자의 분류를 위한 특징값으로 사용될 수 있다. 본 논문에서는 정합도를 사전정보와 융합하여 탐촉자 분류를 위한 특징값을 제안하였다. 제안된 특징값을 사용하여 잘 알려진 베이스 분류기(Bayses classifier)를 사용하여 정확도를 구하였다.

1. 마커 찾기

마커를 찾기 위해서 2가지 사전 정보를 사용할 수 있다. 첫째는 사용하는 마커는 같은 직경을 갖는 원 모양이고, 둘째는 아래 그림에서 보듯이 마커들 사이의 거리가 고정되어 있다는 사실이다. 마커의 모양이 원 모양이지만 전술한 바와 같이 칩 영상에서는 원형 물체의 존재 여부만을 파악하면 되기 때문에 사각형으로 물체를 근사할 수 있다. 한편, 그림 3의 (a)와 (b)에 보인 것처럼 마커들간의 배치에 대한 사전정보를 템플릿의 배경에 융합할 수 있고, 혹은 (c)와 (d)처럼 국소적인 배경을 택하고 후처리 과정에서 사전정보를 융합하는 방법을 채택할 수도 있다. 템플릿 (a), (b)는 마커들 사이의 화소(pixel) 값이 마커 안의 값보다 작다고 가정하지만, 템플릿 (c), (d)는 이 가정을 따르지 않는다. 그러나 불규칙한 배경의 영향으로 인해 이 가정들은 유효하지 않다.

실험 결과 국소적인 배경을 갖는 템플릿의 우수성이 입증되었으므로 본 논문에서는 템플릿(c)와 (d)만을 고려하였다. 템플릿함수는 일반성을 잃지 않고 배경에서 -1이고, 물체내부에서 1의 값을 갖도록 할 수 있다.

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

(a) 전체적인 배경을 갖는 원 템플릿

(b) 전체적인 배경을 갖는 사각 템플릿

(c) 국소 배경을 갖는 원 템플릿

(d) 국소 배경을 갖는 사각 템플릿

그림 3. 여러 가지 템플릿 구성

마커들간의 상대거리가 고정되어 있다는 사전정보를 융합하여 마커를 찾기 위하여 다음과 같은 체계적인 접근 방법을 사용한다.

(1) 모든 탐색영역에 대해 정합도 m(i,j)를 계산한다.

(2) 1/4{m(k,j)+m(k+dx1,j)+m(k+dx2,j)+m(k+dx3,j)}의 값이 최대가 되는 위치 (k,j)를 찾는다.

일반적인 공분산(C)과 정규화 된 공분산(NC)를 적용하는 정합도에 관한 식은 다음과 같다.

$$C(i,j) = \sum_{k,l \in Template} T(k,l)I(i+k,j+l) \qquad (1)$$

$$NC(i,j) = \frac{\sum_{k,l \in Template} T(k,l)(i+k,j+l)}{\sigma_T \sigma_{I(i,j)}} \qquad (2)$$

σ_T와 $\sigma_I(i,j)$는 각각 템플릿 함수와 영상에서 템플릿과 정합되는 부분의 화소 값에 대한 표준편차이다. 템플릿의 내부와 배경이 각각 1과 -1이므로 공분산 C는 템플릿의 내부와 배경사이의 평균차이(mean difference)와 같게 된다.

2. 탐촉자의 분류

탐촉자 분류 성능을 판단하기 위하여 모든 실험영상의 각 탐촉자는 전문가에 의해 분류되어 탐촉자 분류 알고리즘으로 분류한 결과를 비교하였다. 그림 4는 전문가에 의해 분류된 탐촉자의 예를 보인다. 교잡반응이 전혀 안된 것은 class 0, 확실히 된 것은 class 1, 확실치 않은 것은 class 2의 값을 부여하였다. 그림 4에서 볼 수 있듯이 원 안의 탐촉자는 class 2(uncertain), 사각형 안의 probe는 class 1(hybridized), 나머지 probe들은 class 0(not-hybridized)로 분류하였다.

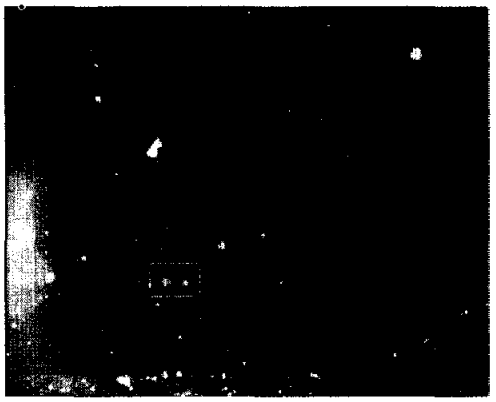


그림 4 probe 분류의 예 (원: Uncertain class (Class 2), 사각형: hybridized (Class 1), 나머지 probe: not-hybridized (Class 0))

각 탐촉자는 마커로부터 일정한 거리에 위치하고 있으므로, 한번 마커를 찾으면 탐촉자의 근사 위치를 알 수 있다. 서론에서 언급하였듯이 탐촉자를 각 클래스로 분류하기 위해서 마커를 찾을 때 계산된 템플릿 정합도를 재사용 할 수 있다. 마커로부터 각 탐촉자의 근사 위치가 결정되면 그 위치를 중심으로 검색 윈도우 범위 내에서 이웃 화소를 검색한다. 검색된 공분산 값 중에서 최대 값을 선택하여 탐촉자의 특징값으로 사용하였다. 그림 1에서 보듯이 한가지 탐촉자당 4개의 특징값을 가지므로 여러 가지 조합이 가능하다.

그림 5는 탐촉자 쌍에서 전술한 바와 같이 구한 공분산 값의 분산도이다. 그림에서 보듯이 대각선 방향으로 분포함을 알 수 있다. 다시말해 각 탐촉자의 공분산 값을 평균한 값을 분류 특징값으로 사용해도 2차원에서 분류하는 것과 비슷한 성능을 보일 것으로 추론된다. 이것은 각각의 탐촉자를 사용해서 독립적으로 실험하게 된다는 것을 보인다. 마찬가지로 두 개의 탐촉자 집합에서 전부 4개의 독립된 실험을 가정 할 수 있기 때문에 4개의 검색된 공분산의 평균 값을 사용하여 1차원에서 분류로 사용하였다.

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

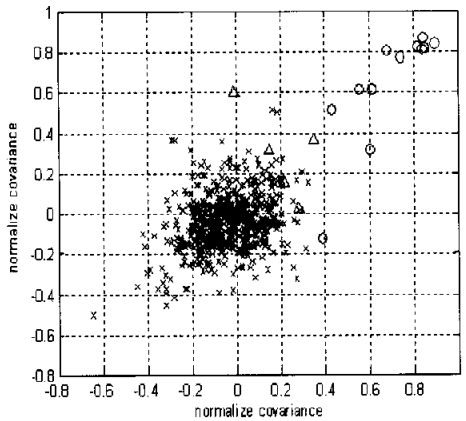


그림 5. 탐촉자 쌍에서의 정규화된 공분산을 벡터그림으로 나타낸 경우 (×: class 0, △: class 2, ○: class 1, 검색창의 크기=0)

정확도를 계산하기 위하여 평균 공분산이 그림 6과 같은 가우시안(Gausssian)분포를 갖는다고 가정하고 베이즈 규칙(Bayes Rule)에 근거한 다음과 같은 선형 분류함수를 정의하였다[5].

$$g_{i(x)} = -\frac{(x-m_i)^2}{2\sigma_i^2} + \log p(w_i), i=0,1,2. \quad (3)$$

여기서 $p(w_i)$, x와 σ_i는 각 클래스가 발생할 확률, 평균 정규화된 공분산, 그리고 각 클래스의 분산이다. 분류경계(decision boundary)는 각 클래스의 분류함수를 같게 하는 평균 공분산 값이 된다. 분류경계가 정해지면 정확도는 식 (4)과 같이 표현될 수 있다.

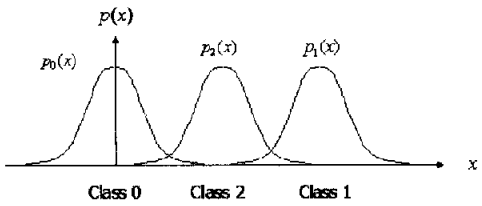


그림 6. 가우시안(Gaussian) 형태의 확률밀도 함수로 표현되는 각 클래스의 분포도

$$\text{정확도} = \int_{-\infty}^{B_{02}} p(w_0)G(m_0,\sigma_0)dx + \int_{B_{02}}^{B_{21}} p(w_2)G(m_2,\sigma_2)dx + \int_{B_{21}}^{\infty} p(w_1)G(m_1,\sigma_1)dx \quad (4)$$

여기서 $G(m_i, \sigma_i)$는 각 클래스의 Gaussian 분포함수이며, B_{02}, B_{21}는 각각 class 0와 class 2사이의 분류경계, class 2와 class 1과의 분류 경계이다. 식 (4)로 표현되는 정확도를 측정하기 위하여 사전 확률 $p(w_i)$는 주로 가정에 의해 결정되고, 각 class의 분포 $G(m_i, \sigma_i)$는 평균과 분산을 각 class의 측정값으로 추정(estimate)하여 정한다[5].

Ⅲ. 결과 및 고찰

1. 마커 찾기

실제 환자 표본으로 교잡 반응된 3장의 슬라이드를 스캔하여 24개의 탐촉자 집합 영상을 얻었다. 각각의 영상은 4개의 마커와 22쌍의 탐촉자들로 구성되어 있다. 표 1은 정합도로써 일반적인 공분산과 정규화 된 공분산을 적용하여 전 장에서 제안한 알고리즘으로 마커를 찾아서 주관적으로 성공여부를 판단하였을 경우의 성공률을 보여주고 있다. 표에서 보듯이 정규화된 공분산을 사용하여 원과 사각형 템플릿에서 100% 성공함을 알 수 있었다. 그림 7은 일반적인 공분산을 적용했을 경우, 높은 밝기의 잡음의 영향으로 마커를 찾는데 실패한 예이다. 공분산 측정은 마커 내 object의 픽셀 값과 배경사이의 평균 차이와 같기 때문에, 높은 세기(intensity)의 클러스터 잡음은 마커보다 공분산 값이 크게 나올 수 있다. 이런 문제점 때문에 공분산의 정규화는 불가피함을 알 수 있었다.

표 2는 사각형 템플릿과 원형 템플릿을 사용했을 때 영상 한 장 당 평균 수행 시간을 보였다. 사각형 템플릿은 가로 세로가 일정한 크기의 템플릿을 사용하므로 한번 계산된 템플릿의 공분산 값은 다음 열과 행의 공분산 값을 계산할 때 재사용이 가능하다. 즉 다음 화소에서의 열과 행의 공분산 계산 시 템플릿 전체에 대하여 다시 계산을 수행할 필요가 없으므로 연산시간이 많이 줄일 수 있는 것이다[12].

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

표 1 : 원과 사각 템플릿의 마커 찾기 성공률

	성공(No. of sets)	실패(No. of sets)	성공률(%)
일반적인 공분산	5	19	20.83
정규화된 공분산	24	0	100

표 2 : 마커 찾기의 영상 한 장당 평균수행시간 (Pentium III*2 800MHz CPU)

	사각형 템플릿	원 템플릿
일반적인 공분산	1초	30초
정규화된 공분산	1초	30초

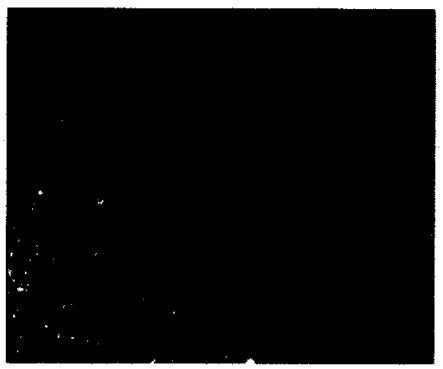


그림 7. 일반적인 공분산을 적용하여 마커를 찾는데 실패한 경우의 예.

평균 정규화 된 공분산 측정값이 최고인 지점을 찾아 마커의 위치로 추정하기 때문에 그 값과 다른 화소 위치에서의 측정값과의 관계를 알아보는 것은 의미가 있다. 그림 8은 24개의 샘플 영상 각각에서 평균 정규화 된 공분산의 마커 위치에서의 측정값과 그 외 다른 위치에서의 측정값의 관계를 나타낸 것이다. X축은 영상번호를 나타내고 Y축은 각 영상에서의 평균 정규화 된 공분산 값의 측정값을 의미한다. "maximum" 으로 표시된 점선 및 실선 쌍은 마커 위치에서의 평균 정규화 된 공분산의 측정값이고, "mean" 그리고 "mean+3*std" 로 표시된 선의 쌍들은 평균 정규화 된 공분산 측정값의 영상 전체 화소에 관한 평균과 표준편차의 3배를 나타낸다.

마커 위치에서는 측정값은 대부분이 0.5보다 크며, 가장 작은 값도 다른 위치에서의 평균 공분산 측정값의 분포보다 표준편차의 3배 이상 떨어져 있다. 이것은 기존의 템플릿 정합법과 마커들의 상대위치에 대한 정보를 효율적으로 이용하는 제안한 알고리즘이 안정적으로 마커를 찾는 다는 것을 보인다. 연산시간이 상대적으로 매우 작은 사각형 템플릿도 원형 템플릿과 견줄만한 성능을 보임을 알 수 있다.

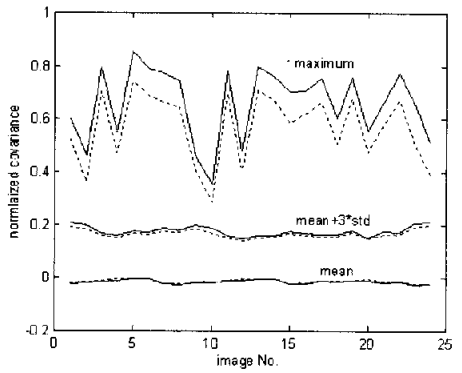


그림 8. 마커에서의 평균 정규화 된 공분산 값과 그 외 화소에서의 평균과 표준편차의 3배 (실선: 원형 템플릿을 사용했을 경우, 점선: 사각형 템플릿을 사용했을 경우)

2. 탐촉자 분류

각 탐촉자의 위치는 마커로부터 고정된 거리에 위치하고 있지만, 영상 획득 및 처리 과정에서 오차를 가질 수 있다. 검색된 마커로부터 탐촉자의 초기 위치를 정하고 이웃화소를 조사하여 주어진 검색 윈도우 크기 내에서 공분산 최대값을 찾는다.

그림 9는 검색창의 크기에 따라 각 클래스의 정규화 된 공분산을 전 장에서 제안한 바와 같이 평균했을 때의 분포도를 나타낸다.

그림에서 볼 수 있듯이 각 클래스가 명확하게 분리되는 것을 알 수 있다. 그림에서 알 수 있듯이 평균 정규화 된 공분산 값 0.5는 마커를 찾을 때 중요한 의미를 갖는다. Class 2는 대부분 0.5 부분에 분포되어 있으며 class 0, class 1은 class 2로부터 멀리 떨어져 있다. '+' 표시는 식(3)에 의하여 구한 class 경계값이다. 한편 검색창의 크기에 따라 분류 정확도가 변할 것이라는 것을 추론할 수 있다.

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

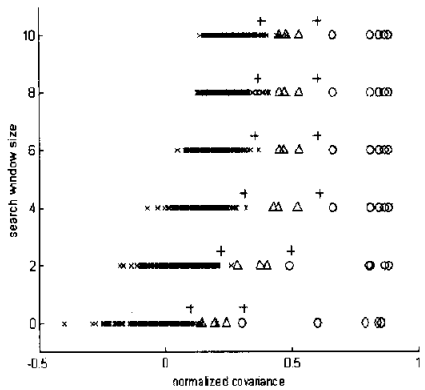


그림 9. 평균 정규화된 공분산을 사용하였을 때의 검색창 크기에 따른 각 class의 분포 (×: class 0, △: class 2, ○: class 1)

식(4)로 표현되는 정확도를 측정 하기 위하여 사전 확률 $p(w_i)$을 가정하여야 한다. 그러나 정확도의 추세는 사전 확률에 따라 크게 변화 하지 않기 때문에 다음과 같은 두개의 간단한 가정으로부터 결정하였다. 첫째는 모든 교잡반응 시에 항상 단일 유형의 HPV에 감염된다는 것이고, 둘째는 교잡반응이 확실히 된 것 (class 1)과 불확실하게 된 것 (class 2)은 같은 빈도로 발생한다는 것이다. 이와 같은 가정으로부터 HPV가 22개의 유형이므로 $p(w_0)=21/22$, $p(w_1)=1/44$ 그리고 $p(w_2)=1/44$ 이 된다. 실제로 여러 가지 사전확률 조합에도 같은 정확도의 추세를 보였기 때문에 본 논문에서는 위의 가정에 대한 결과만을 보였다. 한편 각 class는 정규분포를 가정하였으므로 각 class의 평균과 분산을 측정하여 각 분포를 결정하였다.

아래 그림 10은 식(4)에 위 절에서 가정한 사전정보와 측정된 각 class의 평균과 분산을 대입하여 구한 정확도가 검색창의 크기에 따라 변화하는 것을 나타낸다. 예상했던 것처럼 원이나 사각 템플릿 모두 정확도가 증가하다가 다시 감소함을 보인다. 흥미 있는 관찰은 사각 템플릿을 사용하면 보다 정확히 분류할 수 있다는 사실이다(그림의 점선). 그러나 정확한 분석을 위하여 더 많은 데이터가 필요할 것으로 사료된다.

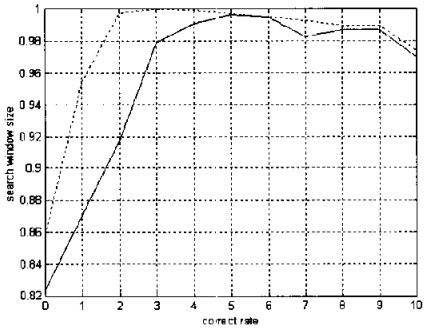


그림 10. 검색창의 크기에 따른 정확도의 변화 (실선: 원 템플릿, 점선: 사각 템플릿)

Ⅳ. 결 론

템플릿 정합법과 사전정보를 조합하여 HPV DNA칩 영상을 해석하는 알고리즘을 제시하였다. 제안한 알고리즘은 마커를 찾는 문제와 탐촉자의 교잡반응 여부를 판정하는 문제를 다루었다. 각 탐촉자를 위한 템플릿으로 정합도를 계산한 후에 마커들의 상대위치에 대한 평균 정합도가 최대인 곳을 찾는 방법을 제시하여 성공적으로 마커를 찾을 수 있음을 보였다. 정합도로는 일반적인 공분산보다는 정규화 된 공분산을 적용하여야 하며, 마커가 원형이지만 연산 시간을 줄이기 위하여 사각형 템플릿으로도 견줄만한 성능을 낼 수 있다는 것을 보였다.

같은 유형의 탐촉자가 4번 중복 인쇄되었다는 사전 정보를 이용한 분류 특징값을 제안하여 교잡반응 여부를 정확히 알아내는 것을 보였다. 마커 탐색결과로 얻은 탐촉자의 근사 위치에 여러 가지 크기의 검색 창을 설정하고, 정규화된 공분산의 최대값을 찾아, 같은 유형의 탐촉자 4개에 대한 평균을 취해 분류 특징값으로 사용하였다. 검색창의 크기에 따른 정확도를 보임으로써 원과 사각형 템플릿 모두 정확히 분류할 수 있는 최적 검색창의 크기를 구할 수 있었다.

제안한 일련의 영상처리 기법들은 다른 종류의 진단용 DNA 칩의 자동 분석이나, 같은 모양의 마커를 사용하는 여러 가지 컴퓨터비젼 응용분야에도 적용 가능하리라 사료된다.

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr

참 고 문 헌

[1] N. C. Cho, H. J. An, J. K. Jeong, S. Kang, J. W. Kim, Y. T. Kim, and T. K. Park, "Genotyping of 22 huaman papillomavirus types by DNA chip in Korean women: Comparison with cytologic diagnosis," Am. J. Obstet. Gynecol., vol 188, No. 1, pp56-62, Jan. 2003.

[2] Y.Chen, E.R.Dougherty, and M.L.Bittner, "Ratio-based decisions and the quantitative analysis of cDNA microarray images," Journal of Biomedical Optics 2(4), pp.364-374, Oct 1997.

[3] J.C.Russ, The Image Processing Handbook, CRC Press, 1999.

[4] D.Schonfeld, "On the relation of order -statistics filters and template matching: optimal morphological pattern recognition," IEEE Trans. Image processing, vol. 9, pp.945-949, May 2000.

[5] R.O.Duda and P.E.Hart, Pattern Classification and Scene Analysis, John Wiley & Sons Inc., pp.1-39, 1973.

[6] N.Braendle, H.Bischof, H.Lapp, "A Generic and Robust Approach for the Analysis of Spot Array Images", SPIE Proceedings, vol 4266-2, 20 - 26 January 2001.

[7] Z.Zhou, J.Stein, Q.Ji, "GLEAMS: A Novel Approach to High Throughput Generic Micro-Array Image Capture and Analysis", SPIE Proceedings, vol 4266-3, 20-26 January 2001.

[8] T.Bergemann, F.Quiaoit, J.Delrow, L.Zhao, "Statistical Issues in Signal Extraction from Microarrays", SPIE Proceedings, vol 4266-4, 20-26 January 2001.

[9] L.Kegelmeyer, L.Tomascik-Cheeseman, M.Burnett, P.van Hummelen, A.Wyrobek, "A groundtruth approach to accurate quantitation of fluorescence microarrays", SPIE Proceedings, vol 4266-6, 20-26 January 2001.

[10] R.Nadon, P.Shi, A.Skandalis, E.Woody, H.Hubschle, E.Susko, N.Rghei, P.Ramm, "Statistical Inference Methods for Gene Expression Arrays", SPIE Proceedings, vol 4266-7, 20-26 January 2001.

[11] W.Liu, R.Mei, D.Bartell, X.Di, T.Webster, T.Ryder, "Rank-based algorithms for analysis of microarrays", SPIE Proceedings, vol 4266-8, 20-26 January 2001.

[12] G.Delenstarr, H.Cattell, C.Chen, A.Dorsel, R.Kincaid, K.Nguyen, N.Sampas, S.Schidel, K.Shannon, A.Tu, P.Wolber, "Estimation of the confidence limits of oligonucleotide array-based measurements of differential expression", SPIE Proceedings, vol 4266-17, 20-26 January 2001

[13] I. Pitas, Digital image processing algorithms and application, John Wiely & Sons Inc., 2000.

김 종 대(Jong-Dae Kim) 정회원

1982년 2월: 서울대학교 전자공학과(공학사)
1984년 2월: 한국과학기술원 전기 및 전자공학과(공학석사)
1990년 2월: 한국과학기술원 전기 및 전자공학과(공학박사)
1988년 3월~2000년 2월: 삼성전자주식회사 중앙연구소 수석 연구원
1994년 12월~1995년 12월: UCSB 방문연구자
2000년 3월~현재: 한림대학교 정보통신공학부 조교수.

<주관심분야> 신호 및 영상처리, 멀티미디어 통신

연 석 희(Seok-Hee Yeon)
2001년 2월: 한림대학교 전자공학과(공학사)
2003년 2월: 한림대학교 컴퓨터공학과(공학석사)
2003년 1월~현재: (주)오스템 근무

<주관심분야> 의료영상, 멀티미디어 통신

Copyright (C) 2002 Nuri Media Co., Ltd. www.dbpia.co.kr