

Homogeneity와 Ranklets를 이용한 치밀 유방에서의 종괴(mass)형 암 검출

Detection of mass type-Breast Cancer using Homogeneity and Ranklets on Dense Mammographic Images

박준영*, 천민수*, 김원하**, 김성민***

Jun-Young Park, Min-Su Chon, Won-Ha Kim, Sung-Min Kim

Abstract - In this paper, we propose a new method for detection of mass-type breast cancer in dense mammogram. As the proposed method analyzes texture of the breast tissue using method by fusing Homogeneity and Ranklets, improve problem of traditional method. Homogeneity gives the measure of uniform density, and Ranklets determine orientation selective property at vertical, horizontal and diagonal in mass region. The proposed method is suitable to dense mammogram with tangled normal tissue and cancer tissue. SVM(Support Vector Machine) classifier is used for effective detection of mass-type breast cancer in dense mammogram.

Key Words : breast cancer, mass, dense mammography, texture analysis

1. 서론

유방암을 디지털 영상에서 자동으로 판독하기 위한 많은 영상처리 기술들이 발전되어 왔지만 현재 임상에서 사용되고 있는 최고의 검출율을 가진 시스템도 아직 mass형 유방암의 검출은 80%에 머물고 있다[1]. 특히 한국어성과 같은 동양 여성에게서 많이 나타나는 치밀 유방의 경우 암 조직과 정상 조직이 엉켜져 있어서 기존의 mass형 유방암 진단 시스템을 치밀 유방영상에 적용하였을 경우에는 상당히 낮은 검출율을 나타낸다.

기존의 암 검출 연구에서, Karseemeijer는 의심되는 영역의 화소들의 기울기(gradient)를 통계적으로 구하여 해당 영역의 조직의 texture가 중심으로 향하는 정도로 암을 판정하는 방법을 제안하였다.[2] Brake는 Karseemeijer의 알고리즘을 multi-scale 방식으로 적용하여 mass형 유방암의 검출 성능을 향상시켰다.[3] 위의 방법들은 암 조직과 정상 조직이 엉켜져 있는 치밀 유방영상에 적용할 경우, 국지적인 gradient의 변화가 심하여 gradient가 통계적으로 mass의 중심부를 향하는지에 대한 판정이 쉽지 않다. Mudigonda등은 일정한 크기의 window에 속한 화소 값들의 gradient의 상관관계(correlation)를 계산하여 주된 방향성이 있는 영역을 검출하였다. 이들은 검출된 영역들에서의 texture를 분석하여 유방암을 검출하는 방법을 제시하였다[4]. 그러나 위의 방법도 치밀 유방영상에 적용시킬 경우에는, 암 조직과 정상 조직이 엉켜져 있어서 mass 영역에서 gradient의 상관관계(correlation)를 찾기가 어렵다.

본 논문에서는 기존의 방법들의 문제점을 개선하기 위하여 유방 조직의 texture를 Homogeneity와 Ranklets를 융합한 기법으로 분석하여 치밀 유방영상에서 mass형 유방암을 검출하는 방법을 제안한다. 1단계에서는 모든 mass의 중심 영역에서 Homogeneity를 측정하여 중심 영역의 조직이 균일하고, 응집되어 있는 mass들을 추출한다. 2단계에서는 추출된 mass 영역의 중심에서 Ranklets를 계산하여 조직의 texture의 방향이 중심으로 향하는지를 판별하여 mass형 유방암을 검출한다. Ranklets은 기존의 gradient 측정 방법보다 간단하며 신뢰성 있는 texture의 방향을 계산한다. 따라서 기존의 방법보다도 치밀 유방에서 효과적으로 mass형 유방암을 검출한다. 또한 SVM(Support Vector Machine) 분류기를 사용하여 Homogeneity 계수와 Ranklets 벡터를 최적화하였다[5]. 본 논문에서 제한하는 기법은 삼성의료원에서 수집한 치밀 유방으로 이루어진 142개의 영상 데이터에 적용하였으며, mass형 유방암의 검출 성능에 있어서 우수한 결과를 얻었다.

본 논문은 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 Homogeneity와 Ranklets 기법을 치밀 유방에 적용하여 texture의 상태를 분석한다. 3장에서는 최적화된 데이터 분류를 위해 Homogeneity 계수와 Ranklets 벡터를 SVM에 적용한다. 4장에서는 본 논문의 결론에 대해서 논한다.

2. 유방조직의 texture 분석

일반적으로 mass형 유방암은 암 조직이 중심에서 방사형으로 성장하는 특징을 가진다. 따라서 mass형 유방암의 중심부에 위치한 조직들은 균일한 밀도를 가지면서, texture의 방향이 중심으로 향한다. 본 장에서 제안하는 기법은 mass의 중심영역에서 Homogeneity를 측정하여 조직이 균일하게 응집되어 있는 mass들을 추출하고, 추출된 mass 영역에서 Ranklets를 계산하여 조직의 texture의 방향이 중심을 향하는지를 판단하여 mass형 유방암을 검출한다.

* 박준영, 천민수 : 경희대학교 전자공학과 석사과정

** 김원하 : 경희대학교 전자공학과 부교수

*** 김성민 : 건국대학교 의공학과 부교수

2.1 Homogeneity를 이용한 균일 mass 조직 검출

Mass형 유방암은 중앙이 중심에서 주변조직으로 침투하기 때문에 mass 중심부에서 균일한 texture 형태로 나타난다. Homogeneity는 분석하려는 영역에서 texture의 균일성을 측정하는 계수이다[6]. 그러므로 치밀 유방영상에서 섬유 조직과 같이 단순히 영겨있는 정상 조직과 암 조직은 Homogeneity 계수에 의해 분리될 수 있다. Texture의 균일성을 측정하는 Homogeneity 계산은 다음과 같다.

$$H = \sum_{i \leq M} \sum_{j \leq M} \frac{P(i,j)}{1+|i-j|} \quad (1)$$

여기서 P(i,j)는 수평, 수직 거리가 각각 1이 되는 Co-occurrence 행렬이고, M은 texture의 균일성을 검사할 mass 영역 중심부의 크기를 나타낸다. 본 논문에서 M은 mass 영역 전체 크기의 20%가 되도록 하였고, Co-occurrence 행렬은 5x5로 하였다.

<그림 1>은 암 조직을 포함한 mass 영역과 정상 조직으로 이루어진 mass 영역에서 Homogeneity 계수들의 예를 보여 준다. 그림에서 보듯이 mass 영역의 중심부에서 측정된 Homogeneity 계수가 클수록 그 영역의 texture가 균일하므로 단일조직을 가지는 암 조직에 가깝다. 반대로, Homogeneity 계수가 작으면 그 영역의 texture가 불균일하여 여러 조직이 영겨져 있는 조직에 가깝다. Homogeneity의 최적화된 계수는 SVM(Support Vector Machine) 분류기를 이용하여 결정하였다. Homogeneity의 계수 결정은 제3.2절에서 설명한다.

2.2 Ranklets를 이용한 texture 방향성 측정

Ranklets은 2002년 Smeraldi에 의해 개발된 알고리즘으로서, 얼굴인식에서부터 3D 구조와 유동성 물체의 움직임을 추정하는 데 응용되고 있다. Ranklets은 픽셀 값의 상대적인 등급에 따라 수직, 수평, 대각선으로 texture의 방향성을 추정하는 벡터이다[7]. 본 논문에서는 Ranklets를 mass 중심을 포함하는 영역에 적용하여 수직, 수평, 대각선에서 유방 조직의 texture의 방향성을 추정한다. 추정된 방향성을 근거로 조직의 texture의 방향이 mass의 중심으로 향하는지를 판단하여 mass형 유방암을 검출한다.

<그림 2>는 Ranklets 벡터를 구하려는 영역에서 수직, 수평, 대각선에 따라 상대적인 밝기의 등급을 구하기 위한 영역들의 구분을 보여준다. 상대적 밝기 등급 계수 W_{yx}^i 는 T_i 영역

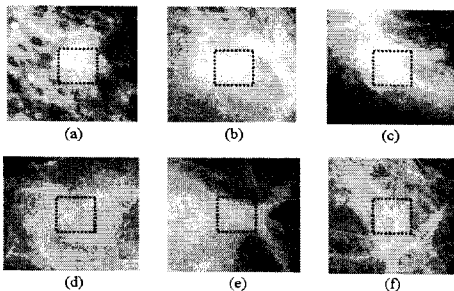


그림 1. 중심부에서의 Homogeneity를 측정하기 위한 mass(a,b,c)와 non-mass(d,e,f)를 나타낸 영상과 영상의 중심(window). 각 영상은 (a)235×235, (b)219×219, (c)133×133, (d)137×137, (e)97×97, (f)155×155이다.

표 1. 유방 조직과 정상 조직을 포함한 영역의 Homogeneity

	(a)	(b)	(c)	(d)	(e)	(f)
Homogeneity	0.4934	0.5354	0.5250	0.3533	0.3776	0.3667

에 포함된 화소들이 C_i 영역에 대응되는 위치에 있는 화소들보다 더 밝은 화소들의 개수가 많을수록 크게 된다. W_{yx}^i 는 다음과 같이 구할 수 있다.

$$W_{yx}^i = z \in T_i \left\{ \pi(x) - \frac{N(N+1)}{2} \right\} \quad (2)$$

여기서 x 는 T_i 에 속하는 화소들의 값이고 N 은 전체화소의 개수이며, π 는 rank transform으로 화소 값들을 상대적인 등급으로 변환한다. 다음은 rank transform, π 의 한 예이다.

$$\begin{pmatrix} 55 & 95 & 25 & 153 \\ 26 & 75 & 92 & 200 \\ 21 & 64 & 88 & 154 \\ 101 & 190 & 199 & 222 \end{pmatrix} \xrightarrow{\pi} \begin{pmatrix} 4 & 9 & 2 & 11 \\ 3 & 6 & 8 & 15 \\ 1 & 5 & 7 & 12 \\ 10 & 13 & 14 & 16 \end{pmatrix} \quad (3)$$

그러면, 수직, 수평, 대각선에 해당하는 Ranklets 벡터는 다음과 같은 식으로 계산된다.[7]

$$R_i = \frac{2 W_{yx}^i}{N^2/4} - 1, \quad i = v, h, d \quad (4)$$

본 논문에서 Ranklets를 계산하는 영역은 mass의 중심으로부터 mass 영역 전체 크기의 40%가 되도록 하였다. 또한 100um/pixel 해상도의 영상에서는 해상도가 높아서 근접하는 화소들 사이에 변화가 미세하여 texture를 관찰하는 것이 비효율적이다. 따라서 본 논문에서는 영상을 downsample하여 해상도를 400um/pixel로 낮추었다. 400um/pixel의 영상 해상도에서는 화소간의 변화가 비교적 뚜렷해서 texture를 관찰하는 것이 효율적이며, 계산량도 20%이하로 감소된다.

<그림 3>은 암 조직을 포함한 mass 영역과 정상 조직으로 이루어진 mass 영역에서 Ranklets 벡터들의 예를 보여 준다. 그림에서 보듯이 암 조직을 포함한 mass 영역 중심부에서 Ranklets 벡터의 값은 모든 방향에 대하여 작게 나타난다. 이는 mass 중심부에서 texture의 방향이 중심을 향하고 있어서 texture의 수직, 수평, 대각선 방향성이 작기 때문이다. 반면에 정상조직으로 이루어진 mass 영역 중심부에서 어느 한 부분이상에서 Ranklets 벡터의 값이 크게 나타난다. 이는 mass 영역의 중심부에서 texture의 방향이 중심을 향하고 있지 않아서 texture의 특징한 방향성이 나타나기 때문이다. SVM(Support Vector Machine) 분류기를 사용한 최적화된 Ranklets 벡터 결정은 3.2절에서 설명한다.

3. SVM(Support Vector Machine)

최근에 패턴분류에 있어서 각광을 받고 있는 SVM 분류기는 1995년 Vladimir Naumovich Vapnik에 의해 개발된 통계적 학습이론으로 VC(Vapnik-Chervonenkis)이론을 통해 데이

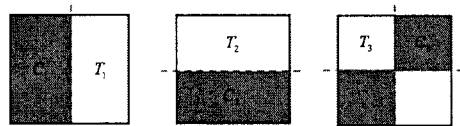


그림 2. 수직, 수평, 대각선에 따라 상대적인 밝기의 등급을 구하기 위한 영역

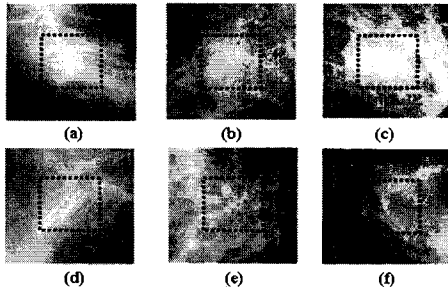


그림 3. 중심부에서의 Ranklets를 측정하기 위한 mass(a, b, c)와 non-mass(d, e, f)를 나타낸 영상과 영상의 중심(window). 각 영상은 (a)153×153, (b)105×105, (c)181×181. (d)91×91, (e)83×83, (f)217×217이다.

표2. 유방 조직과 정상 조직을 포함한 영역의 Homogeneity

Ranklets	(a)	(b)	(c)	(d)	(e)	(f)
수직	0.17	0.12	0.04	0.51	0.61	0.66
수평	0.25	0.20	0.25	0.07	0.22	0.14
대각선	0.26	0.35	0.18	0.66	0.18	0.12

터 분류를 최적화한다. 본 논문에서는 제안하는 Homogeneity와 Ranklets를 융합한 기법으로 mass형 유방암 검출을 최적화하기 위하여 SVM 분류기를 사용하였다. SVM 분류기의 판별식은 다음과 같다[5].

$$f(x) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i (x_i^T x) + b\right) \quad (5)$$

α_i 와 y_i 는 각각 support vector x_i 에 해당하는 가중치 벡터(weight vector)와 클래스레벨(class label)을 나타내고, x 와 N 은 입력벡터와 support vector의 수를 나타낸다.

그림 4는 Homogeneity 계수와 Ranklets 벡터를 이용한 SVM의 학습과정을 보여준다.

3.1 texture의 균일성에 의한 mass 검출

최적화된 Homogeneity 계수 결정을 위해 암 영역과 정상 영역에서 얻은 50개의 Homogeneity 계수 데이터를 Training Set(I)으로 하여 Homogeneity를 위한 SVM 분류기를 학습하였다. 학습을 통해 나타난 Homogeneity 모델을 검증데이터에 적용하여 균일한 texture를 가지는 mass를 검출한다. 검증데이터는 학습데이터와는 별개로 암 영역(71개)과 정상 영역(71개)에서 얻은 Homogeneity 계수를 계산한 것이다. Homogeneity 모델에 의해 검출된 texture가 균일한 mass는 69개의 True Mass(96%)와 43개의 False Mass(61%)이다.

3.2 texture의 방향성에 의한 mass 검출

최적화된 Ranklets 벡터의 결정을 위한 SVM 분류기의 학습은 학습데이터를 제외하고 Homogeneity와 동일하게 이

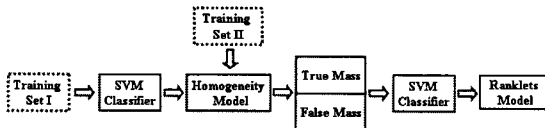


그림 4. Homogeneity 계수와 Ranklets 벡터의 SVM Training

어졌다. Ranklets 모델은 Homogeneity 모델에 의해 추출된 균일한 texture를 가지는 mass에서 계산한 Ranklets 벡터를 Training Set(II)으로 하여 학습되었다. 또한 Homogeneity 모델에 의해 추출된 texture가 균일한 mass만을 검증데이터로 사용하였다. 이를 통해 Ranklets 모델에 의해 texture의 방향이 중심을 향하는 mass는 62개의 True alarm(89%)과 24개의 False alarm(56%)이다.

본 논문에서는 mass 중심부의 texture의 균일성과 방향성을 판단하는 최적화된 Homogeneity 계수와 Ranklets 벡터를 결정하여 치밀 유방영상에서 얻은 142개의 mass 데이터에서 87%(62/71)의 검출율과 33%(24/71)의 오검출율을 보였다.

4. 결론

본 논문에서는 삼성의료원에서 수집한 치밀 유방영상에서 Homogeneity와 Ranklets를 융합하여 mass형 유방암을 검출하는 기법을 제안하였다. 제안하는 Homogeneity와 Ranklets를 융합한 기법은 영클어진 정상 조직과 암 조직을 가지는 치밀 유방에서 mass 영역의 texture를 분석하여 균일한 조직을 가지는 mass를 추출하고, texture의 방향성을 판정하여 mass형 유방암을 검출한다. 제안하는 기법은 치밀 유방에 적용시켰을 때, 기존의 연구 방법에서 문제가 되는 국지적인 gradient의 변화에 의한 영향이 적기 때문에 안정적이라는 장점을 가진다. 본 논문에서 제안하는 기법은 치밀 유방에서 True alarm의 검출율을 높이면서 False alarm의 검출율을 낮추는 우수한 성능을 보였다.

감사의 글

본 연구는 보건복지부의 의료공학 융합기술개발(02-PJ3-PG6-EV06-0002) 사업 지원으로 수행되었음.

참고 문헌

- [1] R. M. Rangayyan, Biomedical Analysis, CRC press, 2005.
- [2] N.Karssemeijer, "Detection of masses in ammograms," Image-Processing Techniques for Tumor Detection (ed. R. N. Strickland), Marcel Dekker, Inc., New York. 2002. pp. 187-212.
- [3] Brake, G. M., Karssemeijer, N., "Single and multiscale detection of masses in digital mammograms", IEEE Trans. Medical Imaging, vol. 18, no. 7, July (1999) 628-639.
- [4] N.R.Mudigonda, "Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-Field Analysis", IEEE Trans.Med.Imag., vol. 20, no. 12, December 2001.
- [5] Chang, C.-C. and C.-J. Lin(2001). LIBSVM: a library for support vector machine. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [6] R. Jain, R. Kasturi and B. Schunck, Machin Vision, McGraw Hill, 1995.
- [7] F.Smeraldi, "Ranklets:a Complete Family of Multiscale, Orientation Selective Rank Feature", Computer Science, september 2003.