

한국형 디지털 마모그래피에서 SVM을 이용한 계층적 미세석회화 검출 방법

권주원¹, 강호경¹, 노용만¹, 김성민²

¹한국정보통신대학교 멀티미디어그룹 영상 및 비디오 시스템 연구실

²건국대학교 의공학과

(Received July 11, 2006. October 2, 2006)

A Hierarchical Microcalcification Detection Algorithm Using SVM in Korean Digital Mammography

Ju Won Kwon¹, Ho Kyung Kang¹, Yong Man Ro¹, Sung Min Kim²

¹Image and Video Systems Lab., Information and Communications University

²Department of Biomedical Engineering, School of Medicine, KonKuk University

Abstract

A Computer-Aided Diagnosis system has been examined to reduce the effort of radiologist. In this paper, we propose the algorithm using Support Vector Machine(SVM) classifier to discriminate whether microcalcifications are malignant or benign tumors. The proposed method to detect microcalcifications is composed of two detection steps each of which uses SVM classifier. The coarse detection step finds out pixels considered high contrasts comparing with neighboring pixels. Then, Region of Interest(ROI) is generated based on microcalcification characteristics. The fine detection step determines whether the found ROIs are microcalcifications or not by merging potential regions using obtained ROIs and SVM classifier. The proposed method is specified on Korean mammogram database. The experimental result of the proposed algorithm presents robustness in detecting microcalcifications than the previous method using Artificial Neural Network as classifier even when using small training data.

Key words : 컴퓨터지원진단(CAD), support vector machine(SVM), 미세석회화, mammography

1. 서 론

현재 유방암은 유럽과 미국을 비롯해 전 세계적으로 증가하고 있는 여성 암 사망률을 높이는 요인 중 하나이며, 우리나라에서도 유방암 사망률이 급증하고 있다. 실제로 지난 83년 이후 인구 10만 명당 유방암 사망률을 조사한 결과 20년 새 2.2배(220%) 가량 증가했음은 물론, 최근 10년간 국내 25-49세 여성 유방암 환자 사망률의 증가 추세는 36.1%로 전 세계에서 가장 높은 것으로 연구 조사 결과 나타났다 [1].

이러한 유방암의 위험성을 줄일 수 있는 치료 방법으로 수년간 많은 연구가 진행되어 왔다. 그 중, 유방촬영술인 마모그래피에 관

련된 연구는 유방암 조기 진단을 통해 사망률을 낮추는 효과적인 방법으로 사용되고 있다 [2,3]. 특히 마모그래피의 연장 선상이라 할 수 있는 컴퓨터지원진단(Computer Aided Diagnosis: CAD)의 개발은 조직 검사와 같은 불필요한 검사는 물론, 사람의 주관적 판단에 의한 진단 오류 등을 줄일 수 있다는 관점에서 많은 연구가 진행 중에 있다.

미세석회화 검출에 적합한 CAD 시스템의 구현을 위해서 기존에는 Artificial Neural Network(ANN)을 분류기로 사용한 시스템이 연구되어 왔다. 그러나 이 방법은 충분한 학습 데이터를 요구하는 분류기의 특성 때문에 미세석회화 부분의 데이터 수집이 어려운 환경에서 최적의 성능을 얻어내는 데는 한계가 있다. 본 논문에서는 적은 양의 데이터만으로 데이터 분류 최적화에 용이한 Support Vector Machine(SVM)의 사용을 제안한다. SVM은 뛰어난 성능과 더불어 조작성이 간단하여 최근 패턴인식분야에서 각광 받고 있는 이진 분류기이다.

본 논문은 보건복지부 의료공학융합기술 개발사업 디지털 유방 촬영 시스템 개발 과제(02-PG6-EV06-0002)에 의하여 지원되었음.

Corresponding Author : 노용만
대전광역시 유성구 문지동 103-6 한국정보통신대학교 영상및비디오시스템연구실
Tel : 042-866-6129 / Fax : 042-866-6245
E-mail : yro@icu.ac.kr

본 논문은 다음과 같이 구성된다. 2절에서는 미세석회화의 특징과 기존의 미세석회화 검출을 위한 CAD 시스템 관련 연구를 살펴보고, 3절에서는 SVM의 기본적인 원리 및 CAD 시스템에의 적용에 대해 알아본다. 4절에서는 제안하는 알고리즘인 SVM을 이용한 효과적인 미세석회화 검출을 위한 CAD 시스템에 대하여 설명하고, 제 5절에서는 실험 및 결과에 대해서 고찰하고, 마지막으로 6절에서는 결론 및 향후 연구 방향을 정리한다.

II. 관련연구

A. 미세석회화 특징

석회화란 유방 조직 내에 칼슘 성분으로 구성된 작은 조각 덩어리를 뜻하는 것으로, 상황에 따라 조기 유방암의 존재를 나타내기도 한다. 보통 석회화는 크기가 작아 초음파상으로는 발견이 어려울 뿐만 아니라 환자의 자각증상이 거의 없다. 따라서 유방암 초기 검진을 위해 마모그래피를 이용한 석회화의 검출은 필요하다.

일반적으로 석회화의 80%가 양성이고 20% 정도가 유방암과 관련이 있다. 이 석회화의 양성 유무는 석회의 모양, 크기, 분포 및 개수 등과 밀접한 관계를 갖고 있다. 양성 석회화의 경우 석회질의 입자가 크고 굵으며(macrocalcification), 암과 직접적인 연관이 있는 악성 석회화의 경우 그림 1에서와 같이 크기가 0.1mm-1mm 내외의 밝은 점인 미세석회화로 보통 1cm×1cm의 면적 안에 4-5개 이상의 분포 특성을 보이고 있다 [4].

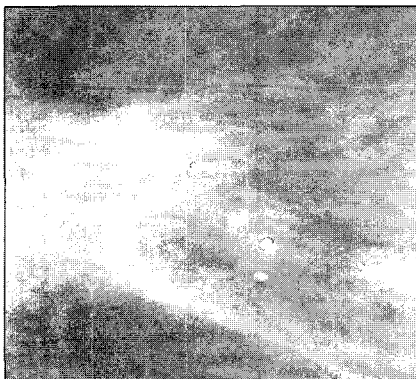


그림 1. 마모그램에서 나타나는 미세석회화.
Fig. 1. Microcalcification in mammogram.

이러한 군집성 미세석회화는 입자의 크거나 모양이 다양하고 불규칙하여 검출이 어렵다. NMD(ACR(American College of Radiology) National Mammography Database)는 이러한 미세석회화를 종류와 분포도 등의 특성에 따라 분류하고, 이들의 양성 유무에 대해서 기술하고 있다 [5]. 실제로 조기 유방암의 60-70%가 미세석회화라는 사실을 감안할 때, 미세석회화의 조기 검출은 유방암 사망률을 줄일 수 있는 효과적인 방법으로 고려되고 있다.

B. Computer Aided Diagnosis(CAD)

CAD는 의료용 영상을 디지털화하여 컴퓨터로 분석함으로써 병의 진단에 도움을 주는 기술로 1999년부터 도입되었으며, 유방암의 X선 촬영 외에, 폐의 X선 단층촬영, 암 조직의 현미경 사진 등 여러 가지 다양한 의료 영상을 대상으로 연구가 진행되고 있다.

유방촬영 분야에서 CAD 시스템은 주로 미세석회화 소견과 종괴 진단과 같은 발견과 분류에 중점을 두고 있다. 유방암의 약 15-20%가 유방촬영술만으로는 진단이 어렵기 때문에 이중 판독을 필수로 하며, 이때 CAD는 제2의 의견기법으로 사용될 수 있다. 또한, 간혹 놓칠 수 있는 병변을 발견함으로써 진단의 민감도를 높일 수 있음은 물론, 위양성 판독을 줄여 불필요한 조직 검사 등을 줄일 수 있어 방사선과 의사를 보조함으로써 업무의 효율성을 높일 수 있다. 이러한 관점에서 유방암 사망률을 줄이기 위해 미세석회화를 효과적으로 검출하기 위한 CAD 시스템의 개발은 활발한 연구 중에 있다 [6-8].

C. 미세석회화 검출을 위한 CAD

미세석회화 검출을 위한 CAD 시스템은 대용량의 마모그램에 대하여 영상처리 및 패턴인식 기법을 사용하여 의심 영역을 검출한다. 따라서 적절한 분류기의 사용은 가장 중요한 부분이라 할 수 있다. 기존에 제안되었던 ANN을 이용한 CAD 시스템 [9]은 분류기의 특성상 충분한 학습 데이터를 요구함은 물론, 분류기를 학습하는 데 오랜 시간이 소모된다. 따라서 정상 영역에 비해 상대적으로 학습 데이터 수집이 어려운 미세석회화의 경우, 불충분한 학습 데이터로 인해 검출 성능의 한계를 가져왔다.

더욱이 한국형 마모그램의 경우 서구형에 비해 유선 조직이 치밀한 Dense-granular의 형태를 띄고 있어 [10] 국내 임상 실험 결과, 현재 상용된 CAD 시스템은 한국형 마모그램을 비롯한 동양계 마모그래피 데이터베이스에서는 적합하지 않다는 문제가 제기되어 왔다.

III. Support Vector Machine

A. Support Vector Machine(SVM)

SVM은 1995년 Vapnik에 의해 Vapnik-Chervonenkis(VC) 이론을 기반으로 제안된 지도학습 알고리즘의 일종으로, 현재 데이터 마이닝 분야는 물론 다양한 패턴인식 응용 분야에 적용되고 있는 이진 분류기이다. 기존에 사용되었던 Multi-Layer Perceptron (MLP)을 비롯한 통계적 패턴인식 방법과 같은 전통적인 학습방식들이 대부분 학습오류를 최소화시키는 경험적 위험 최소화를 기반으로 하는 것에 반해 SVM은 전체의 학습 집단을 하위 집단으로 세분화한 뒤, 이들 각각의 경험적 위험도를 최소화시키는 구조적 위험 최소화를 모델로 한다 [11].

일반적으로 분류기는 학습에 사용되지 않은 새로운 데이터 표본에서도 올바른 결과를 보일 수 있는 일반화 성능이 뛰어나야 한다.

이러한 관점에서 SVM은 특징 공간에서 데이터를 분류할 수 있는 최적의 초평면을 선택함으로써 학습 데이터에만 국한된 성능을 보이는 데이터 과적합현상을 방지함으로써 일반적인 오류를 최소화시켜 좋은 성능을 얻을 수 있다 [12,13].

SVM은 두 클래스를 선형으로 분류할 수 있는 많은 함수 중에서 그림 2에서와 같이 각 클래스에 속해 있는 학습 데이터까지의 최소 거리를 최대화시키는 초평면을 찾고, 이를 결정 경계로 정한다. 그리고 결정 경계에서 가장 가까이 놓여 있는 두 클래스의 학습 데이터들을 Support Vector라 하며, 이 Support Vector들에 가중치를 두어 혼란시킴으로써 적은 양의 데이터만으로도 최상의 학습 효과를 누릴 수 있도록 한다.

가. 선형 SVM

두 클래스에 속해 있는 데이터들이 선형으로 분리 가능할 경우 초평면은 다음과 같은 수식으로 표현될 수 있다.

$$w^T \cdot x + b = 0, \tag{1}$$

단, w 는 가중치 벡터, x 는 입력 벡터, b 는 바이어스 값이다. 각 클래스의 라벨이 $y_i \in \{+1, -1\}$ 에 포함되는 N 개의 데이터를 분류할 수 있는 선형 판별식은 다음과 같다.

$$y_i \cdot (w^T \cdot x_i + b) - 1 \geq 0, \quad \forall i = (1, \dots, N), \tag{2}$$

그러나 대부분의 분류 문제의 경우, 위와는 다르게 그림 2에서와 같이 각 클래스의 데이터들은 비선형적으로 분포되어 있어 선형 경계로 분리할 경우 오분류를 발생한다. SVM에서는 이러한 문제의 해결책으로 슬랙 변수(ξ)를 도입한 유동적인 분류기를 사용하여 일반화 오류를 최소화시킨다. 양의 값을 갖는 슬랙 변수를 도입하여 비선형적으로 분포하는 데이터들을 식별할 수 있는 이진분류 방정식은 다음과 같다.

$$\begin{cases} (w^T \cdot x_i + b) \geq +1 - \xi_i, & \text{for } y_i = +1 \\ (w^T \cdot x_i + b) \leq -1 + \xi_i, & \text{for } y_i = -1, \end{cases} \tag{3}$$

단, ξ 는 선형 경계로 분류 시에 발생하는 모든 학습 에러 값을 의미하며 양의 값을 갖는다.

이러한 부등식 제약조건을 가지고, 최대 마진을 만족시키는 초평면을 찾기 위해서는 아래와 같이 마진의 역수가 최소화되도록 하는 최적화 문제로 해결할 수 있다.

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i, \tag{4}$$

이때, 비선형 분류 문제를 해결하기 위해 페널티 함수 개념을 사용한다. 분리되지 않은 데이터에 오분류와 성능 간의 트레이드 오프 관계의 비용 변수인 C 를 도입하여 C 의 값이 커지면 오분류 오차가 적어지고, 반면 C 가 0의 값에 수렴하게 되면 최대 마진폭을 갖는 초평면이 생성된다.

식 (3)과 식 (4)를 각각 제약식과 목적식으로 이용하는 SVM의 최적화 문제는 볼록 최적화로, 전역 최소화 값을 갖는다. 최적의 가중치 벡터는 볼록 최적화 문제의 필요충분조건인 Karush-Kuhn-Tucker 조건을 만족하는 해를 구함으로써 얻을 수 있다.

$$\hat{w} = \sum_{i=1}^N \alpha_i y_i x_i, \tag{5}$$

단, N_s 는 Support Vector의 개수이며 α_i 는 양의 값인 라그랑제 승수이다. 최적 가중치 벡터를 이용한 이진분류기 SVM의 판별식은,

$$f(x) = \text{sgn}(\hat{w}^T x + b), \tag{6}$$

이다. 따라서 위의 식 (5)에서 얻을 수 있는 Support Vector를 사용하여 얻을 수 있는 최적 가중치 벡터를 이용한 SVM의 판별식

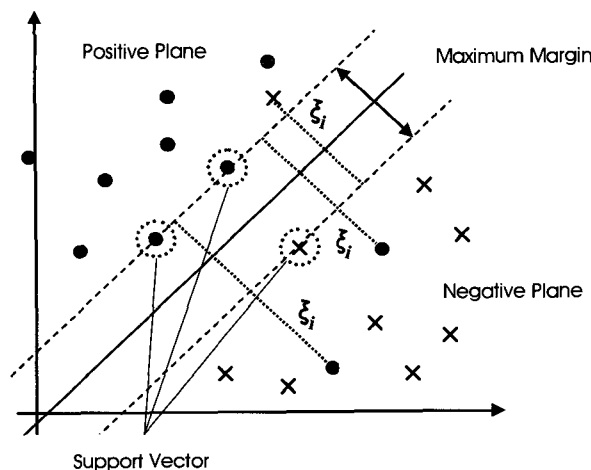


그림 2. SVM의 초평면과 Support Vector.
Fig. 2. SVM with a maximum margin hyper-plane and Support Vector.

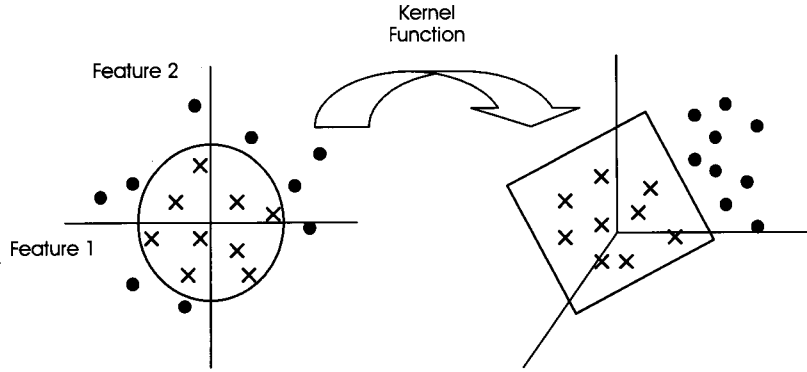


그림 3. 고차원 특징 공간 사상의 예.
Fig. 3. Feature space mapping by kernel function to higher dimension.

은 식 (7)과 같이 표현할 수 있다.

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n \hat{\alpha}_i y_i (\mathbf{x}^T \mathbf{x}_i) + b\right) \quad (7)$$

나. 비선형 SVM

일반적으로는 그림 3의 왼쪽 그림에서와 같이 데이터를 선형 결정 경계로 구분하기에 불가능한 경우가 대부분이다. SVM 분류기는 이러한 문제를 갖고 있는 입력 벡터를 비선형 사상 함수인 커널 함수 $K(\cdot)$ 를 사용하여 고차원 특징 공간으로 변형시켜 선형으로 분리가 가능한 평면을 찾는다.

SVM은 커널 트릭을 사용하여 대표적인 커널 함수인 Polynomial, Radial Basis Function, Sigmoid를 사용하여 고차원 특징 공간으로 사상시킬 수 있다 [14].

SVM 분류기의 판별함수인 식 (7)은 입력 벡터 \mathbf{x} 의 고차원 사상 값인 $\Phi(\mathbf{x})$ 을 사용하여 다음과 같이 표현된다.

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n \hat{\alpha}_i y_i \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) + b\right) \quad (8)$$

그리고 커널 함수는 다음과 같이 고차원 특징 공간상으로 사상된 데이터 간의 내적으로 표현 가능하다.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi^T(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \quad (9)$$

따라서 식 (8)과 식 (9)를 이용하여 선형 분리가 가능한 SVM 분류기의 판별식은,

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n \hat{\alpha}_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (10)$$

와 같다.

B. SVM을 이용한 미세석회화 검출

유방암 조기 진단을 위한 CAD 시스템에서 가장 중요한 부분은

일반 유방 세포 부분과 악성 석회화인 군집성 미세석회화 부분을 분류하고 이를 검출하는 일이다. 따라서 검출에 사용되는 적절한 분류기의 선택은 CAD 시스템의 성능을 결정하는 데 중요하다. 이러한 관점에서 SVM을 이용한 CAD 시스템은 다음과 같은 장점을 갖고 있다.

첫째로, SVM은 구조적 위험을 최소화하여 일반화 오류를 최소화 시킴으로써 기존에 사용되었던 다양한 학습 방법들과 비교하여 전반적으로 좋은 성능을 보인다. 또한, 학습 데이터가 적은 환경에서도 분류 능력이 강하다. 이는 정상 영역에 비해 병변 영역 데이터 수집이 상대적으로 어려운 의료 영상에서 적합하다. 따라서 유방 영상에서 미세석회화 검출에 탁월한 성능을 보일 수 있다.

둘째로, 미세석회화의 경우 다양한 크기와 분포도 등의 특성 때문에 검출에 필요한 다양하게 추출되는 특징 벡터 간의 긴밀한 상관 관계에 의존하여 진단하게 된다. 그렇기 때문에 고차원 입력에도 우수한 능력을 보일 수 있는 분류기를 사용하여야 한다. 이러한 관점에서 SVM은 다양한 커널을 사용하여 비선형 분류 문제에 대해서도 탁월한 성능을 보이며, 커널 함수와 관련된 변수의 개수가 3개 이하이므로 성능 최적화에 관련된 변수의 수가 노드의 수에 비례하는 ANN에 비해 최적화가 용이하다.

이러한 SVM의 용이성은 기존의 연구들을 통해서 확인할 수 있다. Bazzani [8]가 제안한 알고리즘은 2개의 평행 구조의 특징 추출 단계로 구성되어 있으며, Wei [15]가 제안한 논문은 8개의 특징을 단일 구조로 추출한다. 이들 논문에서는 미세석회화의 대비 특성 및 에지 특성 등을 기반으로 한 특징들을 추출하고 이를 이용하여 미세석회화의 존재 유무를 결정하는 검출 단계에서 MLP, LDA(Linear Discriminant Analysis), SVM과 같은 다양한 분류기를 사용한 비교 실험 결과를 통해 미세석회화 검출을 위한 CAD 시스템에서 SVM 분류기의 우수성을 보이고 있다.

IV. SVM을 이용한 미세석회화 검출 방법

본 절에서는 제안하는 미세석회화 검출을 위해 SVM을 사용한 CAD 시스템을 설명한다. 제안하는 알고리즘은 그림 4와 같이 계

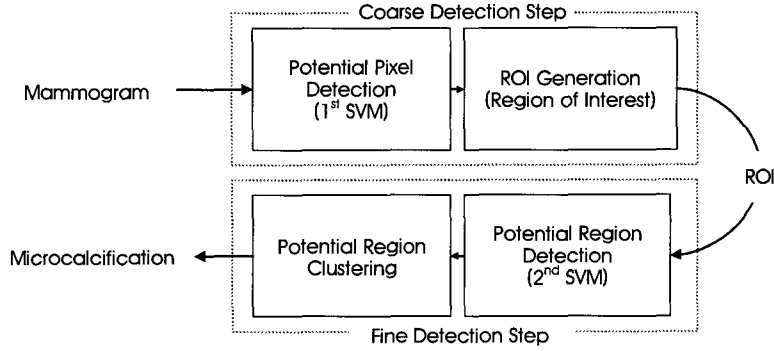


그림 4. 제안하는 미세석회화 검출 CAD 시스템.
Fig. 4. The proposed CAD system for microcalcification detection.

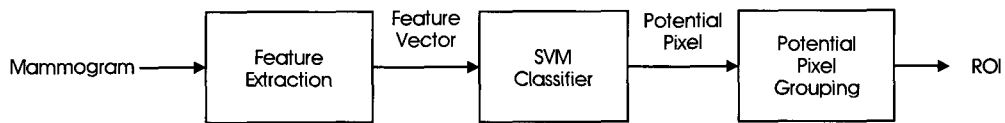


그림 5. Coarse 검출 단계(ROI 형성 단계).
Fig. 5. Coarse detection step(ROI generation).

총적인 두 단계로 구성된다.

첫 번째 Coarse 검출 단계에서는 입력받은 디지털 마모그래피 영상으로부터 의심 가는 픽셀들을 검출하고, 이들의 집합인 Region of Interest(ROI)를 만든다. 두 번째 단계는 Fine 검출 단계로 형성된 ROI로부터 의심 가는 영역들을 검출하고 최종적으로 미세석회화를 결정한다.

A. Coarse 검출 단계

미세석회화는 마모그래프에서 밝고 작은 점들의 형태로 존재한다. Coarse 검출 단계에서는 미세석회화의 크기(0.1mm-1mm 내외), 분포(보통 1cm×1cm의 면적 내에 4-5개 이상 존재)와 주변 영역에 비해 밝은 픽셀 값의 특성을 기반으로 의심 가는 픽셀의 집합인 ROI를 형성한다. 그림 5는 이러한 과정을 블록도로 나타내고 있다.

가. Coarse 검출을 위한 특징 추출

Coarse 검출 단계에서는 의심 픽셀 검출에 적합한 특징 추출을 위해 입력 영상의 모든 픽셀에 2개의 대비도 특성인 대비 중간 값(Median contrast)과 대비 대 잡음비(Contrast to noise ratio)를 나타내는

$$\text{Median contrast}(i, j) = p(i, j) - \text{median}(W), \quad (11)$$

$$\text{Contrast to noise ratio}(i, j) = \frac{p(i, j) - \text{mean}(W)}{\text{std}(W)}, \quad (12)$$

를 이용한다. 이 때, i, j 는 픽셀의 x, y 축 위치를 나타내며 p 는 해당 픽셀의 밝기, W 는 윈도우를 의미한다. $\text{mean}, \text{median}, \text{std}$ 는 해당 윈도우 내에 픽셀 밝기의 평균, 중간 값, 표준편차이다. 윈도우 W 는 전체 영상으로부터 국부 영역의 대비도 특성을 효과적으로 추출할 수 있도록 5×5 와 21×21 두 개의 크기로 설정되었다. 영역의 크기가 5×5 인 작은 윈도우의 경우 지역적인 대비 특성은 잘 나타내지만 노이즈에 약한 반면 21×21 의 큰 윈도우는 지역 대비도의 특성은 약하지만 노이즈에 강한 특성을 갖는다. $\mathbf{F}_{\text{median}_1}^{1st}$ 과 $\mathbf{F}_{\text{CNR}_1}^{1st}$ 는 각각 식(11)과 식(12)에 21×21 윈도우를 사용, $\mathbf{F}_{\text{median}_2}^{1st}$ 와 $\mathbf{F}_{\text{CNR}_2}^{1st}$ 는 5×5 윈도우를 사용하여 입력 영상으로부터 추출된 특징 벡터이다. 따라서 Coarse 검출 단계에서 사용하는 특징 벡터는 $\mathbf{F}^{1st} = \{\mathbf{F}_{\text{median}_1}^{1st}, \mathbf{F}_{\text{median}_2}^{1st}, \mathbf{F}_{\text{CNR}_1}^{1st}, \mathbf{F}_{\text{CNR}_2}^{1st}\}$ 이다.

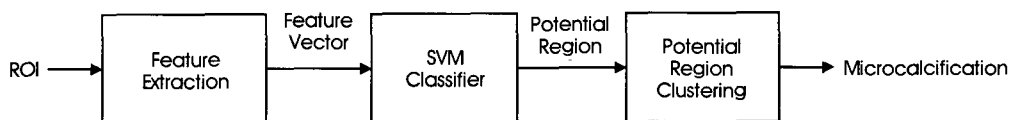


그림 6. Fine 검출 단계(미세석회화 검출 단계).
Fig. 6. Fine detection step(microcalcification decision).

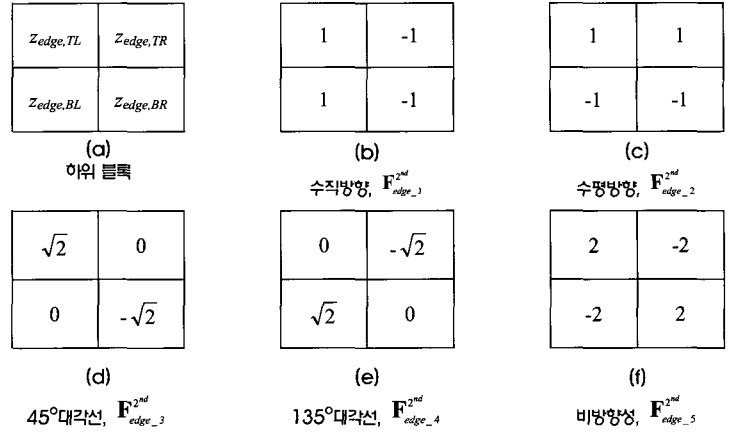


그림 7. 2×2 방향 및 비방향성 에지 마스크.
Fig. 7. Directional and non-directional edge mask.

나. 의심 픽셀 검출

주변 영역에 비해 밝은 명암을 갖는 의심 픽셀을 검출하기 위해 위의 단계에서 추출된 특징 벡터들은 각각 개별적으로 사용하는 것보다는 이들의 상호 관계에 의한 값을 사용하는 것이 노이즈를 최소화시키면서 적절한 대비값을 얻을 수 있어 훨씬 우월한 성능을 보인다. 이를 위해 고차원의 입력 벡터에서도 우수한 성능을 보이는 SVM을 사용하였다. 위에서 추출한 4차원 특징 벡터를 분류기의 입력 벡터로 사용하여 두 개의 클래스를 분류하는 선형 분류 방정식은 식 (13)으로 표현할 수 있다.

$$f(\mathbf{F}^{i'}) = \sum_{i=1}^N \hat{\alpha}_i y_i K(\mathbf{s}_i, \mathbf{F}^{i'}) + b, \tag{13}$$

단, \mathbf{s}_i 는 support vector이다. 입력 마모그램의 모든 픽셀에 대해 식 (13)이 적용되어 의심 픽셀이 검출된다.

다. ROI 형성

검출된 의심 픽셀들은 미세석회화의 크기와 분포 특성에 의해 주변의 의심 픽셀들과 다음과 같은 기준으로 그룹화를 수행한다.

$$Dis((i, j)_{potential}, (i', j')_{potential}) < T_d, \tag{14}$$

단, Dis()는 해당 픽셀 간의 거리를 나타내며, 검출 픽셀 간의 거리가 T_d 보다 작은 픽셀들은 하나의 그룹으로 그룹화한다. 본 논문에서는 주어진 데이터베이스 환경에서 실험을 통해 얻은 최적의 T_d 값으로 1.7이 사용되었다.

식 (14)에 의해 형성된 각 픽셀들의 그룹은 그 크기가 일정 영역 보다 작거나 클 경우 제거된다. 이 영역은 사용하는 데이터의 해상도와 미세석회화의 크기 및 분포를 기반으로 계산하는 과정으로, 본 논문에서는 5-400 pixels에 해당하며, 이 조건에 적합한 의심 픽셀의 그룹을 ROI로 정의한다.

표 1. Fine 검출에 사용한 특징

특징 벡터	설명
\mathbf{F}_{num}^{2nd}	ROI를 구성하는 픽셀의 개수
$\mathbf{F}_{density}^{2nd}$	주변 ROI 분포 정도
\mathbf{F}_{var}^{2nd}	ROI 내의 분산
\mathbf{F}_{grad}^{2nd}	ROI 내 모든 픽셀의 Sobel 마스크[16] 결과 평균값
$\mathbf{F}_{direct}^{2nd}$	ROI의 기울기 벡터의 x축 수평 방향과의 각도
$\mathbf{F}_{edge}^{2nd} = \{ \mathbf{F}_{edge,1}^{2nd}, \mathbf{F}_{edge,2}^{2nd}, \mathbf{F}_{edge,3}^{2nd}, \mathbf{F}_{edge,4}^{2nd} \}$	ROI 내 모든 픽셀의 2×2 방향성 에지 정보 (그림 7(b),(c),(d),(e) 참조)
$\mathbf{F}_{edge,5}^{2nd}$	ROI 내 모든 픽셀의 2×2 비방향성 에지 정보 (그림 7(f) 참조)
$\mathbf{F}_{high_info}^{2nd}$	ROI를 포함하는 블록과 주변의 여덟 방향성 블록들과의 블록단위 고대역 성분인 에지 정보(그림 8 참조)

B. Fine 검출 단계

그림 6은 Fine 검출 단계의 블록도로서, Coarse 검출 단계에서 형성된 ROI의 크기에 따른 적응적 특징 추출을 통하여 최종 미세석회화를 결정하는 단계를 나타낸다.

가. Fine 검출을 위한 특징 추출

Fine 검출 단계에서는 미세석회화의 다양한 모양 및 크기 특성을 기반으로 Coarse 검출에서 형성된 ROI로부터 미세석회화로 의심되는 영역의 검출을 위한 특징을 추출한다. 사용하는 특징 벡터는 $F^{2nd} = \{F_{num}^{2nd}, F_{density}^{2nd}, F_{var}^{2nd}, F_{grad}^{2nd}, F_{direct}^{2nd}, F_{edge_1}^{2nd}, F_{edge_2}^{2nd}, F_{edge_3}^{2nd}, F_{edge_4}^{2nd}, F_{edge_5}^{2nd}, F_{high_info}^{2nd}\}$ 의 총 11차원으로, 표 1과 같다.

$F_{density}^{2nd}$ 는 균집하여 존재하는 미세석회화의 분포 특성을 기반으로 주변 ROI의 분포 정도를 의미한다. 해당 ROI의 중심으로부터 100×100pixel 이내 면적에 몇 개의 ROI가 존재하는지를 측정한다.

각 ROI의 수평, 수직, 2개의 대각선 방향 및 비방향성 등 총 5개의 에지 정보의 특징 벡터인 F_{edge}^{2nd} 는 미세석회화의 모양 특성과 MPEG-7 Edge Histogram Descriptor [17]를 기반으로 추출한다. n 번째 ROI에서 이 특징들은,

$$F_{edge}(ROI_n) = \sum_{n=1}^M \left\{ z_{edge,TL} \cdot \sum_{(i,j) \in TL_n} p_n(i,j) + z_{edge,TR} \cdot \sum_{(i,j) \in TR_n} p_n(i,j) + z_{edge,BL} \cdot \sum_{(i,j) \in BL_n} p_n(i,j) + z_{edge,BR} \cdot \sum_{(i,j) \in BR_n} p_n(i,j) \right\} \quad (15)$$

와 같이 추출된다. 이때 M 은 각 ROI 내에 존재하는 에지 정보를 추출하기 위해 사용되는 이미지 블록의 개수를 의미한다. Edge Histogram Descriptor는 주어진 영상의 다양한 에지뿐만 아니라 에지가 존재하지 않는 경우도 고려해야 하기 때문에 픽셀 단위가 아닌 블록 단위로 추출한다. ROI를 구성하는 모든 픽셀들은 4×4의 이미지 블록으로 나눈 후, 각 이미지 블록을 그림 7(a)와 같은 4개의 하위 블록으로 나누며 각 하위 블록에 포함되는 픽셀들의 밝기 합에 그림 7의 5개 마스크를 적용한다. 이때, TL (Top Left), TR (Top Right), BL (Bottom Left), BR (Bottom Right)는 4×4 pixel로 구성된 이미지 블록을 구성하는 4개의 하위 블록을 뜻하며, $p(i,j)$ 는 하위 블록 내에서 (i,j) 에 위치하는 픽셀의 밝기를 뜻한다. 미세석회화의 경우 비방향성 에지가 강조가 되는 반면 정상 세포의 경우 방향성 에지가 강조된다. 이는 미세석회화와 유방의 유선 모양의 차이를 기반으로 얻은 실험적 결과이다.

$F_{high_info}^{2nd}$ 는 ROI단위로 고대역 성분을 추출하기 위해 그림 8과 같이 ROI를 포함할 수 있는 가장 작은 크기의 블록을 기준으로 이 블록의 크기와 같은 8개의 인접 블록을 정하고, 정의된 블록 단위로 3×3 라플라시안 필터 [16]를 사용하여 공간적 고대역 성분인

에지 정보를 추출한다. 이렇게 추출된 특징 벡터는 마모그램 상에 작고 밝은 점의 형태로 분포하는 미세석회화의 밝기 및 모양 특성에 의해 정상 영역에 비해 미세석회화 영역에서 비교적 높은 고대역 정보 분포를 보인다.

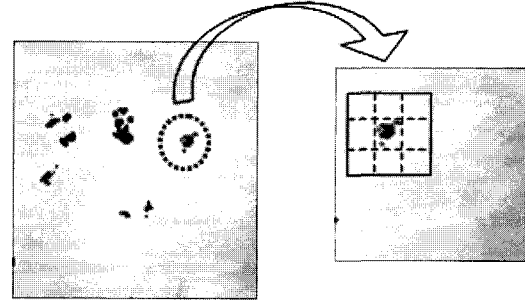


그림 8. ROI와 8개의 인접 블록.
Fig. 8. The block enclosing ROI and 8-adjacent blocks.

나. 의심 영역 검출

의심 영역 검출을 위한 분류기로는 SVM을 사용하였고 의심 영역 검출을 위한 선형 판별식은 다음과 같다.

$$f(F^{2nd}) = \sum_{i=1}^N \hat{\alpha}_i y_i K(s_i, F^{2nd}) + b \quad (16)$$

다. 미세석회화 결정

위의 과정을 통해서 검출된 의심 영역들은 주변의 미세석회화들과 밀접해 분포하는 미세석회화 특성을 기반으로 일정 거리 이내에 존재하는 의심 영역들을 하나의 미세석회화 지역으로 무리짓는 과정을 통해서 최종적으로 미세석회화와 그 위치를 결정하게 된다.

$$Dis(R_{potential_i}, R_{potential_j}) < T_{d_cluster} \quad (17)$$

균집화가 이뤄지는 경계는 위의 식 (17)과 같으며, 이 식에서 $R_{potential}$ 은 검출된 의심 영역을 의미하고, $T_{d_cluster}$ 는 한 덩어리로 간주될 수 있는 적절한 거리 경계 값으로서 미세석회화의 분포 특성과 전문 의사들의 의견을 토대로 인간의 시각 특성에 기반을 두어 실험에 사용한 데이터베이스에서 검출 영역 간의 거리가 70pixel보다 작을 경우 하나의 동일 영역으로 간주하고, 이들을 각각의 미세석회화 분포 영역으로 최종 결정하게 된다.

V. 실험 및 결과

제안하는 미세석회화 검출 CAD 시스템이 한국형 마모그래피에 최적화되도록 분류기 훈련을 위한 데이터베이스로 한국형 디지털 마모그래피 영상을 사용하였다. 이 영상은 삼성의료원에서 제공받은 것으로 2001년부터 2004년까지 수집된 4,835장의 마모그래피 데이터이다. 사용된 데이터의 해상도는 100μm/pixel이며

12bit/pixel로 구성되어 있다.

분류기 학습을 위해 미세석회화가 포함된 유방 영상 30개와 정상 유방 영상 30개로 구성된 총 60개의 데이터가 사용되었다. 데이터 집합은 총 92개로 41개의 미세석회화 영상과 51개의 정상 영상으로 구성되었다. 모든 데이터는 전처리 과정을 거치지 않은 원본 마모그래피 영상이다. 미세석회화 영역에 대한 결정은 의사의 진단을 기반으로 이루어졌다. CAD 시스템의 처리 시간을 줄이기 위해 분할된 마모그래피 영상이 실험에 사용되었다.

미세석회화 검출의 각 단계별 결과는 그림 9에서 확인할 수 있다. 그림 9(a)는 CAD 시스템에 입력된 원본 마모그래피 영상이다. (b)는 Coarse 검출 단계 결과로써 의심 픽셀들이 표시되어 있다. (c)는 이전 단계에서 검출된 의심 픽셀들을 미세석회화 크기 특성을 기반으로 그룹화하여 ROI를 추출한 결과의 중심점을 표시하고 있다. 마지막으로 (d)는 형성된 ROI로부터 Fine 검출한 결과로써 하얀 점들은 의심 영역의 검출 결과이고, 주변의 미세석회화들과 밀집되어 분포하는 미세석회화의 분포 특성을 기반으로 군집한 영역의 결과를 세모 형태로 나타낸 것이다.

본 논문에서는 제안하는 CAD 시스템의 성능을 기존의 시스템 [9]과 비교하기 위해 Free-Response Operation Characteristics (FROC) 커브를 사용하였다. 이는 한 개 영상을 검출할 때 발생하는 오검출(False-Positive:FP) 개수를 기준으로 시스템의 정확도(True-Positive:TP)를 측정하는 방법이다.

그림 10는 제안하는 방법의 성능을 나타내는 FROC 곡선이다. case1은 제안한 방법인 SVM을 사용하고 학습 데이터 60개를 사용하여 훈련한 시스템이고, case3은 기존의 방법인 ANN을 분류기로 사용하여 제안한 방법과 동일한 환경에서 실험한 결과이다. 4개와 6개의 노드로 구성되어 있는 2단계 은닉층 (2-layer hidden

layer)로 구성된 ANN이 실험에 사용되었다. 한 개의 마모그래피 영상 당 1.25개의 오검출이 있을 때 case1은 98.04%, case3은 88.24%의 검출 성능을 보인다. 실험 결과를 통해 미세석회화 검출을 위한 CAD 시스템에서 SVM이 ANN에 비해 우수한 성능을 보임을 확인할 수 있다. case2는 앞에서 사용한 학습 데이터 중 29장의 마모그래피 영상으로 분류기를 훈련한 것으로 1.25개의 오검출 시 90.91%의 검출률을 보인다. case2와 case3을 비교함으로써 SVM이 적은 훈련 데이터의 환경에서도 좋은 성능을 보임을 알 수 있다.

분류기로 ANN을 사용한 경우, 학습을 위해 5시간이 소요된 반면 SVM을 사용한 경우 30초 정도의 시간이 소요되어 분류기 훈련 시 ANN과 SVM의 현저한 연산량의 차이를 확인할 수 있었다.

VI. 결 론

본 논문에서는 한국형 마모그래피의 환경에 최적화되어 있는 SVM 분류기를 사용한 계층적 미세석회화 검출 방법을 제안한다. 제안한 방법은 2개의 계층적 구조의 검출 단계로 구성되어 있으며, 각 단계에서는 미세석회화의 크기, 모양 및 분포 등의 다양한 특성을 기반으로 추출한 여러 개의 특징들을 사용하여 미세석회화를 검출하였다.

의료보조시스템인 CAD 시스템의 중요 관점이 분류 및 발견이라는 점을 감안할 때, 적절한 분류기의 사용은 시스템의 성능에 직접적인 영향을 끼친다. SVM은 기존의 다양한 패턴분류방법에 비해 분류기의 조작 및 세팅이 간단하여 최적화에 용이하여 최근 다양한 분야에서 각광받고 있는 패턴분류방법이다. 더욱이, 적은 양의 학습 데이터 환경과 고차원 입력 벡터에서도 성능이 우수하다.

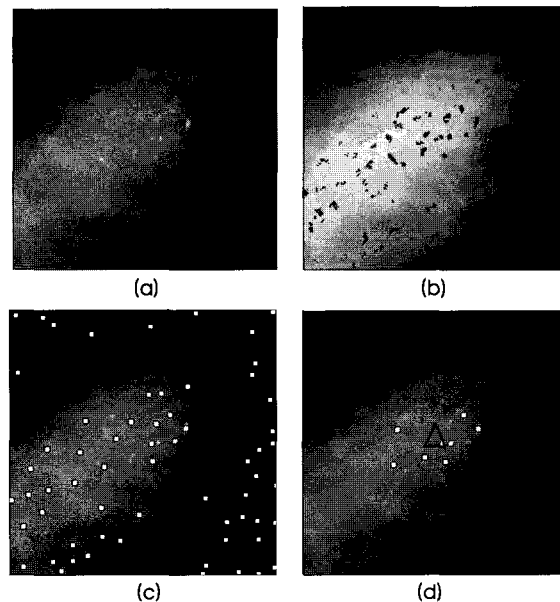


그림 9. 각 단계별 검출 결과, (a) 원본 마모그램, (b) Coarse 검출 결과, (c) ROI 형성 결과, (d) 검출된 미세석회화

Fig. 9. Microcalcification detection result at each step, (a) original mammogram, (b) Coarse detection result, (c) ROI generation, (d) detected microcalcification.

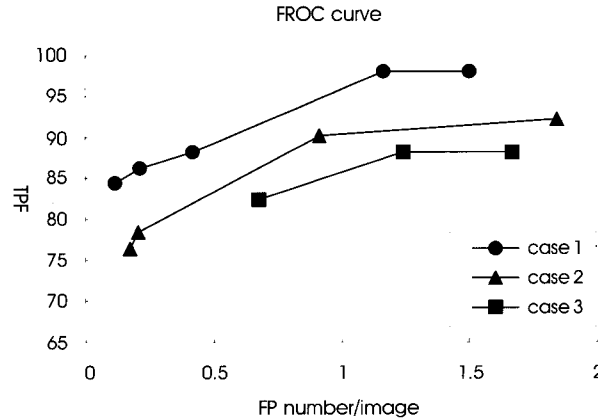


그림 10. 제안하는 방법의 FROC 곡선.

Fig. 10. The FROC curve of the proposed method.

따라서 미세석회화 검출처럼 학습 데이터 수집이 상대적으로 어렵고, 다양한 특성 때문에 이들의 상호 연관성에 의해 분류해야하는 경우, SVM의 사용은 효과적이다. 본 논문에서는 미세석회화 검출의 효율성을 위해 각 검출 단계별로 SVM을 사용하였으며, 제안한 방법의 우수성과 유효성은 실험 결과를 통해 확인할 수 있었다.

본 논문은 분할 영상을 사용하여 실험하였다. 마모그래피의 큰 해상도를 감안할 때 시스템의 연산량과 이에 따른 연산 속도를 위해 이미지 전처리 과정은 필수적이라 할 수 있겠다. 시스템의 실용화를 위해서는 향후 이 부분에 대한 연구가 수행되어야 할 것이다.

참고문헌

- [1] Y. Choi, Y.J. Kim, H.R. Shin, and K.Y. Yoo, "Long-term prediction of female breast cancer mortality in Korea," *Asian Pac. J. Cancer Prev.*, vol. 6, no. 1, pp. 16-21, 2005.
- [2] P.C. Johns and M.J. Yaffe, "X-ray characterization of normal and neoplastic breast tissues," *Phys. Med. Biol.*, vol. 32, no. 6, pp. 675-695, 1987.
- [3] "Full-Field Digital Mammography," *Technology Evaluation Center Assessment Program*, vol. 17, no. 7, pp. 1-22, 2002.
- [4] <http://www.komen.org>.
- [5] <http://www.acr.org>.
- [6] K.Thangavel and M.Karnan, "Computer Aided Diagnosis in Digital Mammograms: Detection of Microcalcifications by Meta Heuristic Algorithms," *GVIP Journal*, vol. 5, no. 7, pp. 41-55, 2005.
- [7] G.V. Serio and A.C. Novello, "A report by the superintendent of insurance in consultation with the commissioner of health to the Governor and the legislature on the advisability of the adoption of a law that would expand the definition of mammography screening to include the review of X-Ray examinations by use of a computer aided detection device," Insurance Department of New York State, USA, March, 2003.
- [8] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani, "An SVM classifier to separate false signals from microcalcifications in digital mammograms," *Phys. Med. Biol.*, vol. 46, no. 6, pp. 1651-1663, 2001.
- [9] H.K. Kang, Y.M. Ro, and S.M. Kim, "A Microcalcification Detection Using Adaptive Contrast Enhancement on Wavelet Transform and Neural Network," *IEICE Trans. on Information & Systems*, vol. E89-D, no.3, pp.1280-1287, 2005.
- [10] J.W. Kwon, W.K. Moon, Y.J. Lee, M.H. Moon, J.S. Yeo, J.K. Chung, D.Y. Noh, and J.G. Im., "Differentiation of Benign and Malignant Breast Masses: Comparison of Ultrasonography with Power Doppler Imaging and 99mTc-Scintimammography," *Journal of the Korean Radiological Society*, vol. 45, no.2, pp.215-220, 2001.
- [11] S. Haykin., *Neural Networks*, New Jersey, Prentice Hall, 1999.
- [12] C.J.C. BURGESS, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.
- [13] I. El-Naqa., Y. Yang, M.N. Wernick, N.P. Galatsanos, and R.M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imaging*, vol. 21, no. 12, pp.1552-1563, 2000.
- [14] C.W. Hsu, C.C. Chang, and C.J. Lin, *A Practical Guide to Support Vector Classification*, Technical report, National Taiwan University, 2003.
- [15] L. Wei, Y. Yang, R.M. Nishikawa, and Y. Jiang, "A study on several Machine-learning methods for classification of Malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imaging*, Vol. 24, No. 3, pp. 371-380, 2005.
- [16] R.C. Gonzalez and R.E. Woods, *Digital image processing*, New Jersey, Prentice Hall, 2002.
- [17] B.S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, LTD, 2002.