

錐制約部分空間法における正常データの 確率密度推定に基づく異常検出

山崎 優大^{1,a)} 野里 博和² 岩田 昌也² 高橋 栄一² 何森 亜由美³ 岩瀬 拓士⁴ 坂無 英徳²

受付日 2014年5月28日, 再受付日 2014年7月18日,
採録日 2014年9月5日

概要: 錐制約部分空間法は, 非負の特徴ベクトルに対して錐形状の空間を形成することで学習パターンを精度良く表現し, 錐との角度を基にパターン認識を行う. しかし, 錐形状の空間内では表面付近と中心付近の特徴ベクトルの区別ができないため, 錐形状の空間の広がりが多い場合は, 認識性能が低下するという問題がある. そこで本論文では, 錐の表面付近の異常を検出するため, 錐形状の空間における確率密度を基にした異常検出手法を提案する. 提案手法では, 錐形状の部分空間の広がり方向を表す空間上において学習パターンの確率密度関数を作成し, 確率密度が低い位置に存在する特徴ベクトルを異常として検出する. 実験では, 乳腺超音波画像の実データを用いて病変検出精度の検証を行い, 提案手法の有効性を確認した.

キーワード: 錐制約部分空間法, 確率密度推定, 異常検出, パターン認識

Anomaly Detection Based on Density Estimation of Normal Data in Cone-restricted Subspace

YUDAI YAMAZAKI^{1,a)} HIROKAZU NOSATO² MASAYA IWATA² EIICHI TAKAHASHI² AYUMI IZUMORI³
TAKUJI IWASE⁴ HIDENORI SAKANASHI²

Received: May 28, 2014, Revised: July 18, 2014,
Accepted: September 5, 2014

Abstract: A cone-restricted subspace method can express learning patterns accurately by generating a convex cone for non-negative feature vectors. Classification of conventional method is performed based on the angle between the input vector and the cone. However, recognition performance is reduced if the spread of the convex cone is large, because it is impossible to distinguish between vectors near the surface and those around the center of the cone. This paper proposes an anomaly detection method based on probability density of normal data in cone-restricted subspace. Classification by the proposed method is based on the probability contained in the convex cone. We demonstrate anomaly detection from breast ultrasound images using proposed method, and confirmed effectiveness of the method.

Keywords: cone-restricted subspace, density estimation, anomaly detection, pattern recognition

¹ 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
² 産業技術総合研究所情報技術研究部門
Information Technology Research Institute, National Institute of Advanced Industrial Science and Technology,
Tsukuba, Ibaraki 305-8568, Japan
³ 高松平和病院外科
Department of Surgery, Takamatsu Heiwa Hospital,
Takamatsu, Kagawa 760-8530, Japan
⁴ がん研有明病院乳腺センター外科
Breast Oncology Center, The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Koto,
Tokyo 135-8550, Japan

1. はじめに

画像認識技術に基づく異常検出は, 監視カメラ映像からの不審者検出や医用画像からの病変検出など, 様々な分野において重要性を増している. 異常検出では, 正常クラスの分布を事前に定義し, その分布から外れるデータを異常として検出する手法が一般的である. 多種多様な異常データのすべてを事前に定義することは困難であるのに対し, 正常データの分布はとらえやすく, 正常を定義することが

a) yamazaki-yudai@aist.go.jp

比較的容易なためである。

正常データの分布から外れるデータを異常として検出する手法として、部分空間法を用いた異常検出手法が提案されている [1]。この手法は、正常クラスの特徴ベクトル群を近似する部分空間を算出し、検査対象から抽出した特徴ベクトルと部分空間との距離や角度を基に異常を検出する。

部分空間法の認識性能を高めるための改良手法として、錐制約部分空間法が提案されている [2]。この手法では、画像から抽出される特徴量が非負値ベクトルであることに着目し、特徴空間上の非負領域に偏在する分布を錐形状の空間により近似する。この手法を用いた異常検出手法では、検査対象から抽出した特徴ベクトルと正常クラスの錐との角度を算出し、事前に設定した閾値より大きいものを異常として検出する。

しかし、錐制約部分空間法において、正常クラスを表す錐形状の空間の広がり大きい場合は、異常が見落とされる恐れがある。錐の中心付近と比較して、表面付近の特徴ベクトルは、錐に属する度合いが小さいものの、錐制約部分空間法では、どちらの特徴ベクトルも錐との角度が 0 と算出され、区別できないためである。

そこで本論文では、この問題を解決するため、錐形状の空間内における正常データの確率密度に基づく異常検出手法を提案する。錐制約部分空間法に正常データの確率密度を導入することで、錐の表面付近と中心付近にある特徴ベクトルが区別され、表面付近の異常を見落とすことなく検出できる。提案手法ではまず、正常クラスの錐の広がり方向を表す空間において、正常データの確率密度関数を作成する。次に、作成した確率密度関数を基に、密度が小さい位置に存在する特徴ベクトルを異常として検出する。

実験では、特徴空間上で錐形状に分布する人工データと、実データとしての乳腺超音波画像を用いて提案手法による異常検出精度の検証を行った。その結果、従来の錐制約部分空間法と比較して提案手法の有効性を確認した。

2. 錐制約部分空間法

2.1 画像認識における部分空間法

画像認識において、部分空間法が多く用いられており、静止画像や動画からの異常検出に適用されている [1], [3]。部分空間法を用いた異常検出では、まず、正常クラスの特徴ベクトル群に対して主成分分析を適用し、主成分ベクトルを算出する。次に、検査対象のデータから抽出した特徴ベクトルと主成分ベクトルが張る空間（部分空間）との距離や角度が大きい場合に異常として検出する。この手法は、画像の明るさの変化や画像内に描出される対象物が複数存在する場合においてもロバスト性を有し、高い認識性能を持つ。画像の明るさの変化や対象物の数の変化は特徴ベクトルのスカラー倍や加法として影響し、主成分ベクトルの線形結合として算出される部分空間は、これらの演算

に関して閉じているためである。

画像から抽出される特徴ベクトルとして用いられる輝度値などの統計量やヒストグラムにより表現される HLAC 特徴 [4], SIFT 特徴 [5], LBP 特徴 [6] などは、特徴空間上の非負領域にのみ存在し、分布に偏りがある。部分空間法は、このような偏りのある分布を広い空間で近似してしまうため、誤認識が発生する。そこで、非負値の特徴ベクトルに対し、事前にその分布を考慮した空間を定義すれば、学習対象となる特徴ベクトル群の近似精度は向上すると考えられる。

非負値に着目したパターン認識の代表的な手法として、小林らが提案する錐制約部分空間法があげられる [2]。この手法では、学習する特徴ベクトル群が非負領域に偏在することに着目し、その分布を錐形状の空間で厳密に近似する。これにより、通常の部分空間法より高い認識性能を得ることができる。

2.2 錐の定義

D 次元の特徴空間上において、錐 C は次のように定義される。

$$C = \{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i = \Xi \boldsymbol{\alpha}, \alpha_i \geq 0 \} \quad (1)$$

$$\text{where } \Xi = \{ \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_N \}$$

$$\boldsymbol{\alpha} = \{ \alpha_1, \alpha_2, \dots, \alpha_N \}^T$$

N は錐を構成する基底ベクトル $\boldsymbol{\xi}_i$ ($\in \mathbf{R}^D$) の数、 α_i は非負の結合係数である。なお、従来の部分空間法により算出される部分空間 S は次式で表現される。

$$S = \{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i = \Xi \boldsymbol{\alpha} \} \quad (2)$$

式 (1) と (2) の比較より明らかなように、錐 C は従来の部分空間 S における結合係数 α_i に非負の制約を与えたものである。そのため、部分空間法と同様に錐制約部分空間法は、画像の明るさや対象物の数の変化に対してロバスト性を持ち、高い認識性能を得ることができる。

2.3 特徴ベクトルと錐のなす角度

錐制約部分空間法を用いた異常検出では、正常クラスに属する特徴ベクトル群を学習することで正常クラスを表す錐を構成し、検査対象から抽出された特徴ベクトル \mathbf{y} と正常クラスを表す錐 C とのなす角度 θ が、事前に設定した閾値より大きいものを異常として検出する。角度 θ は、図 1 に示すように特徴ベクトル \mathbf{y} とその錐 C への正射影ベクトル \mathbf{x} とのなす角度により定義される。

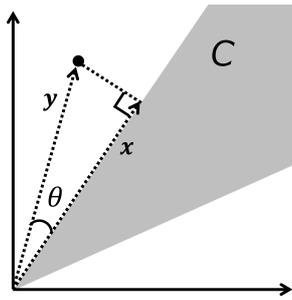


図 1 錐とのなす角度の例

Fig. 1 Angle between the input vector and the cone.

$$\theta = \arcsin\left(\frac{\min_{\mathbf{x} \in C} \|\mathbf{y} - \mathbf{x}\|}{\|\mathbf{y}\|}\right)$$

$$= \arcsin\left(\frac{\sqrt{\min_{\alpha_i \geq 0} \|\mathbf{y} - \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i\|^2}}{\|\mathbf{y}\|}\right) \quad (3)$$

式 (3) における $\min_{\alpha_i \geq 0} \|\mathbf{y} - \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i\|^2$ は、非負最小二乗法 [7] を適用することで計算できる。なお、特徴ベクトル \mathbf{y} が錐 C に含まれる場合、 \mathbf{y} は錐を構成する基底ベクトル $\boldsymbol{\xi}_i$ の線形結合で表現できる。そのため、式 (3) において、 $\mathbf{y} = \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i$ となり、角度 θ は 0 となる。また、非負値の特徴ベクトルは、特徴空間上の非負領域に分布するため、特徴ベクトルと錐との角度 θ の範囲は $0 \leq \theta \leq \frac{\pi}{2}$ となる。

2.4 錐の構成方法

文献 [2] で提案されている 3 つの錐（厳密凸錐，包括凸錐，円錐）の構成方法のうち、学習対象の特徴ベクトル群を最も精度良く近似できる厳密凸錐の構成方法について述べる。

厳密凸錐は、学習対象の特徴ベクトル群 X が構成する錐から、その錐に從属する特徴ベクトルを Algorithm 1 の処理により除去することで錐の基底ベクトル C を算出する。錐に從属する特徴ベクトルを判定する際、式 (3) を用いて錐と特徴ベクトルの角度を算出し、その角度が事前に設定した閾値 θ_{th} (≈ 0) より小さいものを錐に從属すると見なす。

2.5 錐制約部分空間法の問題点

実世界での異常検出問題において正常データと異常データの分布が完全に分離することは非常に少なく、正常クラスの錐が広がる場合は、錐の内側に異常が存在する危険性が高まる。一般に、正常クラスを表す錐の内側において異常が存在する場合、正常の度合いは錐の中心付近ほど大きく、表面付近ほど小さい。そこで、錐の中心付近と表面付近を区別できる連続値の算出方法が必要である。

しかし、錐制約部分空間法における式 (3) の角度算出では、図 2 に示すような、中心付近の特徴ベクトル \mathbf{y}_1 と表面付近の特徴ベクトル \mathbf{y}_2 の錐 C との角度がともに 0 とな

Algorithm 1 厳密凸錐を張る基底ベクトル C の算出

```

input
  特徴ベクトル群:  $X = [\mathbf{x}_1, \dots, \mathbf{x}_M]$  ( $M$  はサンプル数)
  閾値:  $\theta_{th} (\approx 0)$ 
initialize
  繰り返しカウンタ:  $t \leftarrow 0$ 
  錐の基底ベクトル  $C$  を初期化:  $C \leftarrow X$ 
  錐の基底ベクトル  $C$  の数  $N^t$  を初期化:  $N^0 \leftarrow M$ 
repeat
  step1
     $X$  の要素をランダムに選択し、部分集合  $X_S$  とその補集合  $X_{\bar{S}}$  を作成。
  step2
    各  $\mathbf{x}_i \in X_S$  に対し、 $X_S$  から  $\mathbf{x}_i$  を除いた凸錐  $C_{S_i}$  との角度  $\theta_i$  を算出し、 $\theta_i < \theta_{th}$  ならば、 $X$  と  $X_S$  から  $\mathbf{x}_i$  を除去
  step3
     $\mathbf{x}_j \in X_{\bar{S}}$  に対し、 $X_S$  から構成される凸錐  $C_S$  との角度  $\theta_j$  を算出し、 $\theta_j < \theta_{th}$  ならば、 $X$  から  $\mathbf{x}_j$  を除去
  step4
    繰り返しカウンタを更新:  $t \leftarrow t + 1$ 
    錐の基底ベクトル  $C$  の更新:  $C \leftarrow X$ 
    錐の基底ベクトル  $C$  の数  $N$  を更新:  $N^t \leftarrow [X \text{ のサンプル数}]$ 
until ( $N^t = N^{t-1}$ )
output
  錐の基底ベクトル:  $C$ 
  
```

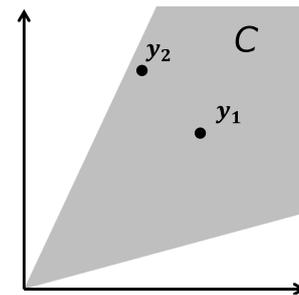


図 2 錐に含まれる特徴ベクトル
Fig. 2 Feature vectors in cone.

るため、両者を区別できない。そのため、正常データと異常データの分布が完全に分離されず、正常クラスの錐に含まれる異常を検出することは、本質的に不可能であるという問題がある。

また、錐制約部分空間法では単峰性で凸形状の分布を仮定し、1 つの凸錐で学習対象の特徴ベクトル群を近似するため、多峰性や非凸形状などの複雑な分布について精度良く近似できない。そこで、クラスタリング手法を用いて、複数の錐を構成することで複雑な分布への対応を試みる手法が提案されている [2]。しかし、複数の錐に対して錐制約部分空間法での式 (3) を適用するため、錐の中心付近と表面付近に存在する特徴ベクトルを区別できないという問題は解決されない。

3. 提案手法

前節で述べた問題を解決するため、錐制約部分空間法に確率密度推定を導入し、錐形状の空間内における正常デー

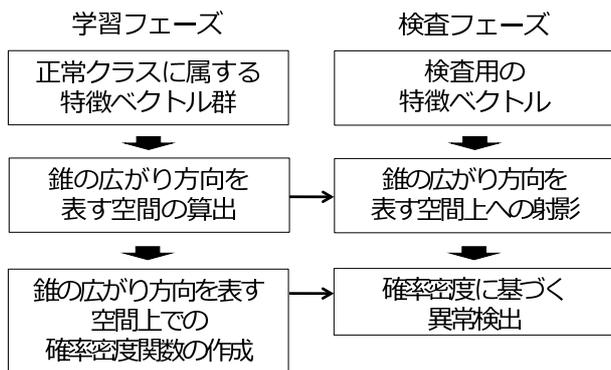


図 3 提案手法のフローチャート

Fig. 3 Flowchart of proposed method.

タの確率密度に基づく異常検出手法を提案する。

提案手法は、学習フェーズと検査フェーズにより構成されている (図 3)。学習フェーズでは、正常クラスの特徴ベクトル群から構成される錐の広がり方向を表す空間を算出し、この空間において正常クラスの分布を推定して確率密度関数を作成する。検査フェーズでは、学習フェーズで算出した錐の広がり方向を表す空間に検査用の特徴ベクトルを射影し、その空間での確率密度が小さい場合に異常として判定する。以下、図 3 に示すフローチャートに沿って各処理について説明する。

3.1 錐の広がり方向を表す空間の算出

錐の広がり方向を表す空間の算出方法には、錐制約部分空間法における包括凸錐の構成方法を利用する。サンプル数 N の正常クラスに属する特徴ベクトル群 $X = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ (\mathbf{x}_i は D 次元の列ベクトル) に対し、次式に示すようにベクトルのノルムを 1 とする正規化を施して、特徴ベクトル群を単位超球面上に射影する。

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \quad (i = 1, \dots, N)$$

単位超球面上の特徴ベクトル群 $\hat{X} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N]$ に対して自己相関行列 $R_{\hat{X}}$ を算出する。

$$R_{\hat{X}} = \hat{X} \hat{X}^T$$

自己相関行列 $R_{\hat{X}}$ に対して固有値問題を解くことで固有値・固有ベクトルが算出される。

$$R_{\hat{X}} U = U \Lambda$$

ここで、 $U = \{\mathbf{u}_1, \dots, \mathbf{u}_D\}$ は、固有ベクトル \mathbf{u}_i を列とする行列であり、 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_D)$ は、固有値 λ_i を対角要素とする対角行列である。

自己相関行列に対する固有値問題では、元の特徴空間における正常クラスに属する特徴ベクトル群を低次元空間に正射影したときの写像の誤差が最小となるベクトルが算出される。このとき固有ベクトルを固有値の大きさに従って

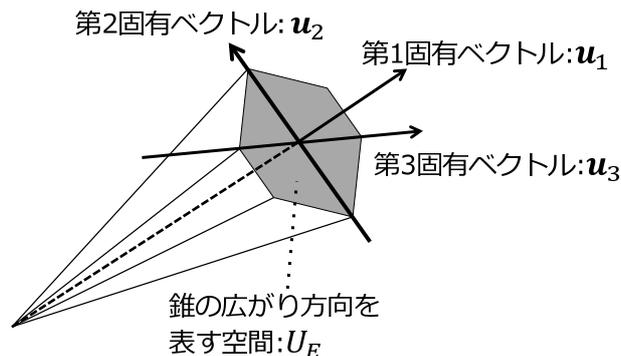


図 4 錐の広がり方向を表す空間の例

Fig. 4 Space that represents the spread direction of cone.

降順に並べると、第 1 固有ベクトル \mathbf{u}_1 は特徴空間における原点からの錐の方向ベクトルとなり、第 2 以降の固有ベクトルが張る空間は正常クラスに属する特徴ベクトル群の単位超球面上での分布の広がり方向を表す。この空間上への射影行列 U_E は次式により定義される。

$$U_E = \{\mathbf{u}_2, \dots, \mathbf{u}_{M+1}\} \quad (1 \leq M \leq D - 1)$$

ここで、 U_E は錐の広がり方向を表す空間を張る基底ベクトルを列とする行列である。また、 M は錐の広がり方向を表す空間の次元を表すパラメータである。 M が大きくなるほど錐の広がり方向を表す空間が高次元となり、正常クラスの確率密度の推定が困難となる。

図 4 に 3 次元 ($D = 3$) の特徴空間の場合の錐の方向ベクトルと、錐の広がり方向を表す 2 次元 ($M = 2$) の空間の例を示す。

3.2 錐の広がり方向を表す空間上での確率密度関数の作成

正常クラスの特徴ベクトル群の分布に基づいて、錐の広がり方向を表す空間上での確率密度関数を作成する。確率密度関数を作成する手法としては、モデルを仮定するパラメトリック手法とモデルを仮定しないノンパラメトリック手法がある。錐の広がり方向を表す空間上における正常クラスの様々な分布に対応するため、本論文ではノンパラメトリック手法の 1 つであるカーネル密度推定法 [8] を採用する。

錐の広がり方向を表す空間上において正常クラスの確率密度関数の作成手順は以下のとおりである。まず、単位超球面上に射影された特徴ベクトル $\hat{\mathbf{x}}_i$ を次式により錐の広がり方向を表す空間上に射影して、特徴ベクトル \mathbf{x}_i^E を算出する。

$$\mathbf{x}_i^E = U_E^T \hat{\mathbf{x}}_i \quad (i = 1, \dots, N)$$

次に、錐の広がり方向を表す空間上での特徴ベクトル群を基に、カーネル密度推定法を用いて確率密度関数 $p(\mathbf{t})$ を作成する。

$$p(\mathbf{t}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h^M} k\left(\frac{\mathbf{t} - \mathbf{x}_i^E}{h}\right) \quad (4)$$

$$\text{where } k(\mathbf{x}) = \frac{1}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{x}\right)$$

\mathbf{t} は確率変数であり，検査対象となる特徴ベクトルである． M は錐の広がり方向を表す空間の次元， h はバンド幅を表すパラメータである．また， $k(\mathbf{x})$ は，ガウスカーネル関数である．

カーネル密度推定法では，カーネル関数 $k(\mathbf{x})$ の選び方には影響を受けないが，バンド幅 h には影響を受け，次元 M が大きくなると確率密度の推定が困難となる．そのため，正常データの確率密度を精度良く推定するためには，適切な次元 M とバンド幅 h を決定する必要がある．本論文における次元 M とバンド幅 h は，4.2.2 項と 4.3.2 項に述べる予備実験により決定される．

3.3 錐の広がり方向を表す空間上での確率密度に基づく異常検出

検査フェーズでは，式 (4) で定義された確率密度関数を用いて，検査対象の特徴ベクトルの確率密度を算出し，それを基に異常を検出する．

まず，単位超球面上へ射影された検査対象となる特徴ベクトル \mathbf{y} を学習フェーズにおいて算出した錐の広がり方向を表す空間へ射影して，特徴ベクトル \mathbf{y}^E を算出する．

$$\mathbf{y}^E = U_E^T \hat{\mathbf{y}}$$

$$\text{where } \hat{\mathbf{y}} = \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

次に，錐の広がり方向を表す空間において，学習フェーズで作成した確率密度関数を基に異常を検出する．学習フェーズで作成した確率密度関数は，正常クラスの錐に含まれる度合いを表す．そこで，検査用の特徴ベクトル \mathbf{y}^E での確率密度 $p(\mathbf{y}^E)$ が，事前に設定した閾値 p_{th} より小さいものを異常として判定する．

4. 計算機実験

本章では，提案手法の有効性を検証するために行った 2 つの計算機実験，(1) 3 次元の特徴空間上において 2 つの錐を形成する人工データに対する異常検出実験と，(2) 乳腺超音波画像を使った実データに対する異常検出実験について述べる．

4.1 評価方法

本実験における評価方法としては，ROC (Receiver Operating Characteristic) 解析を用いた．ROC 曲線は，正常と異常とを分類する閾値を変化させたときの偽陽性率 (False Positive Rate) を横軸に，真陽性率 (True Positive Rate) を縦軸にプロットしてできる軌跡である．ここで，

偽陽性率は正常データを異常として検出した割合で，真陽性率は異常データを正しく異常として検出した割合を表す．ROC 曲線が左上方に近づくほど，真陽性率が 100%，偽陽性率が 0% に近づいていること，すなわち検出性能が向上していることを意味している．ROC 曲線が左上に近づくということを定量的に表す指標が AUC (Area Under the Curve) 値である．AUC 値は，ROC 曲線の下側面積であり，異常検出が完全な場合は 1，ランダムな場合は 0.5 となる．

4.2 人工データを用いた実験

提案手法の有効性を検証するため，3 次元の特徴空間上で凸形状と非凸形状の錐の分布をとる 2 種類の人工データを用いて，部分空間法 (比較手法 1)，錐制約部分空間法における厳密凸錐 (比較手法 2)，提案手法における確率密度関数を正規分布を仮定した最尤法で推定した手法 (比較手法 3) の 3 手法と提案手法とを比較した．

4.2.1 人工データの生成方法

本実験で使用した 2 種類の人工データは，それぞれ凸形状および非凸形状の分布をとる．図 5(a) に示す人工データ 1 では，正常データと異常データの錐の軸方向をそれぞれ， $(x, y, z) = (1, 1, 1), (2, 1, 2)$ として設定した．また，錐の広がり方向を表す空間上で，平均 $\boldsymbol{\mu} = (0, 0)$ ，分散 $\sigma^2 = 0.5$ として円形の正規分布をとるよう作成した (図 5(b))．

図 6(a) に示す人工データ 2 では，錐の軸を表す方向ベクトルを人工データ 1 と同様に設定した．非凸形状の錐となるように，錐の広がり方向を表す空間上で，半円の円周上にノイズを付加して三日月状の分布をとり (図 6(b))，正常および異常データの分布が重なるように生成した．

実験では，学習フェーズで使用する正常データとして 1000 サンプル，検査フェーズで使用する正常データとして 1000 サンプル，異常データとして 1000 サンプル用意した．

4.2.2 人工データに対する実験結果

人工データ 1 と 2 に対する ROC 曲線をそれぞれ図 7(a) と図 7(b) に示す．実験で使用した提案手法のパラメータは，人工データ 1 に対しては次元 $M = 2$ ，バンド幅 $h = 0.25$ とし，人工データ 2 に対しては次元 $M = 2$ ，バンド幅 $h = 0.2$ とした．これらは，予備実験において錐の広がり方向を表す空間の次元 $M = 2$ とし，カーネル密度推定法におけるバンド幅 h を 0.01 から 20 ままで 0.01 の間隔で変化させたときに AUC 値が最大となったパラメータである．

人工データ 1 を用いたときの結果について考察する．まず，提案手法と比較手法 1 および 3 では，偽陽性率が約 0.5 以上のとき真陽性率が 1 となり，異常を見落とすことなく検出できることが分かる．いずれの手法でも錐の表面付近と中心付近を区別できたためであると考えられる．一方，比較手法 2 は，他の手法と比べて AUC 値が低い．従来の

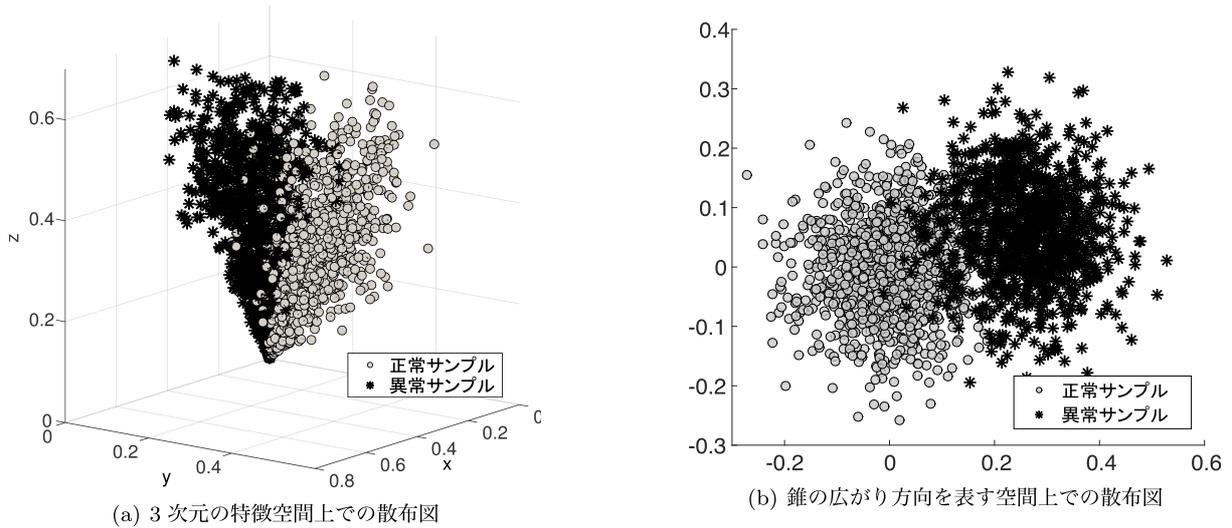


図 5 人工データ 1: 凸形状の錐
Fig. 5 Synthetic dataset1: Convex cone.

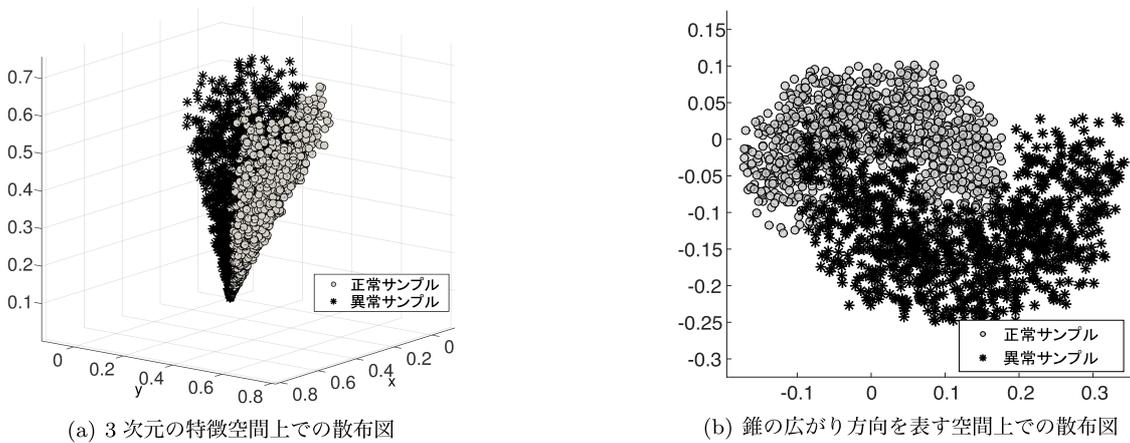


図 6 人工データ 2: 非凸形状の錐
Fig. 6 Synthetic dataset2: None convex cone.

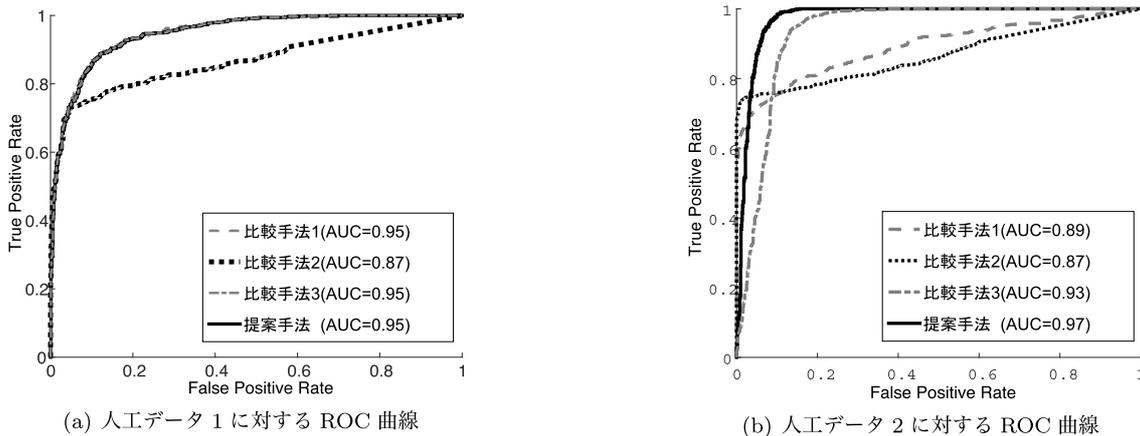


図 7 人工データに対する ROC 曲線
Fig. 7 ROC curve of synthetic dataset.

錐制約部分空間法では、錐の中に含まれる特徴ベクトルと錐とのなす角度がすべて 0 となり、錐の中にある異常データをすべて正常として判定し、異常を見落としてしまったことが原因である。

次に、人工データ 2 を用いたときの結果について述べる。比較手法 1 と 2 では真陽性率が 1 に達することはなく、異常の見落としが発生している。両手法ともに錐の表面付近と中心付近を区別できなかったためであると推察される。

一方で、比較手法3では錐の表面付近と中心付近を区別できたため、比較手法1と2と比べてAUC値が高くなる。さらに、提案手法では、カーネル密度推定法により複雑な分布に対しても確率密度関数を精度良く推定できるため、他の比較手法よりも高いAUC値となる。

以上の結果より、比較手法2は正常データと異常データの分布が重なる場合に異常の見落としが発生し、比較手法1と3は正常データの分布が複雑な場合にAUC値が低下する。その一方で、提案手法は、複雑な分布に対しても錐の表面付近と中心付近を区別できたため、人工データ1と2ともに高いAUC値を得ることができる。

4.3 乳腺超音波画像からの腫瘍像検出実験

本節では、実世界のデータに対する提案手法の有効性を検証するために行った乳がん検診で用いられる乳腺超音波画像から病変を検出する実験について述べる。

わが国のがん患者数は年々増加している。女性に最も多いのは乳がんで、16人に1人が経験するといわれている [9], [10]。その検査方法として近年、乳腺超音波検査が注目されているが、超音波プローブの操作と正確な診断を同時に行うための非常に高度な技量が必要なため、熟練した医師の養成に時間がかかるという課題がある。そこで、画像認識技術を用いて病変を自動検出し、医師の診断を支援するシステムの開発が急務となっている [11], [12]。

医用画像からの病変の自動検出手法では、以下に示す2つの理由のため、正常データをモデル化した異常検出が有効である。第1に、乳腺超音波画像などの医用画像では、多くの正常データを収集できるのに対して病変（異常データ）を収集することは困難で、サンプル数に偏りが生じるためである。第2に、病変の形状は多種多様であり、正常データと比較して異常データをモデル化することが困難なためである。

本論文では、乳がんの疑いのある病変の1つである腫瘍像を検出することを目的に、次の2つの実験を行った。1つ目は、乳腺超音波画像から腫瘍像を異常として検出する能力を検証するための実験である。2つ目は、異なる被験者に対する提案手法の汎化能力を検証するための実験である。この2つの実験において、ROC解析を用いて4.2節で述べた3つの手法と提案手法の比較を行った。

4.3.1 乳腺超音波画像

本実験で使用した乳腺超音波画像は、グレースケール256階調、画素数800×600、フレームレートが30fpsの動画画像である。図8に例を示す。実験では、腫瘍像ありと診断された被験者7人から撮影された動画画像A~Gを用意した。これらの動画画像の各フレームにおいて、正常な乳腺領域を50×50画素のサイズでランダムに切り出して正常な乳腺画像を複数作成した。異常データについては動画画像のフレームや位置を変更することで、1つの腫瘍像から複数

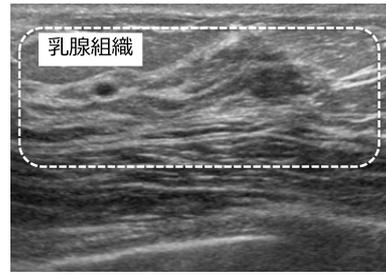


図8 乳腺超音波画像の例

Fig. 8 Example of the breast ultrasound image.

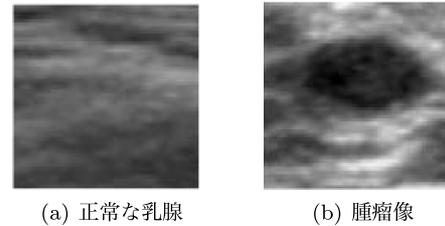


図9 実験で使用した画像の例

Fig. 9 Examples of experiment data.

表1 各被験者におけるサンプル数

Table 1 Number of samples in each patients.

	正常な乳腺画像	腫瘍像画像 (症例数)
被験者 A	4167	976 (2)
被験者 B	3180	50 (1)
被験者 C	3098	70 (1)
被験者 D	3811	79 (1)
被験者 E	2411	63 (1)
被験者 F	662	69 (1)
被験者 G	4712	56 (1)
合計	22041	1363 (9)

の腫瘍像画像を作成した。作成した正常な乳腺画像の例を図9(a)に、腫瘍像画像の例を図9(b)に示す。各被験者から作成した正常な乳腺画像と腫瘍像画像のサンプル数、腫瘍像の症例数を表1に示す。

画像からの特徴抽出手法には、対象画像の直線成分や曲率成分といった幾何学的情報をとらえる高次局所自己相関 (Higher-order Local Autocorrelation; HLAC) 特徴 [4] を用いる。この特徴抽出手法を採用した理由は2つある。1つ目は、自己相関関数により定義されていることからノイズに頑健であり、超音波画像に多く含まれるノイズの影響を低減させる効果が期待されるためである。2つ目は、HLAC特徴抽出では、検出対象の輪郭を切り出す処理が不要であり、腫瘍像の輪郭が不鮮明であることが多い超音波画像に適していると考えられるためである。

4.3.2 乳腺超音波画像に対する有効性の検証

本実験では、学習フェーズにおいて、被験者7人の正常な乳腺画像からランダムに10000枚の画像を取得してHLAC特徴を抽出することで、正常クラスの特徴ベクトル

群を作成する。検査フェーズにおいては、学習フェーズで使用した画像とは異なる正常な乳腺画像 1000 枚と腫瘍像画像 1363 枚を使用した。なお、提案手法のパラメータは錐の広がり方向を表す空間の次元 $M = 1$ 、バンド幅 $h = 0.25$ を採用した。これは、予備実験において錐の広がり方向を表す空間の次元 M を 1 から 34 まで変化させ、カーネル密度推定法におけるバンド幅 h を 0.01 から 20 まで、0.01 の間隔で変化させたときに AUC 値が最大となるパラメータである。

実験結果を図 10 に示す。比較手法 1 と 2 は、偽陽性率が 1 未満では真陽性率が 1 とならず、異常の見落としが発生する。一方、提案手法と比較手法 3 は、偽陽性率が 0.6 以上のときに真陽性率が 1 となり、正常と異常を分類する閾値を適切に設定することで見落とすことなく異常を検出できる。提案手法で算出された錐の広がり方向を表す空間における確率密度関数が正常データの分布を精度良くとらえ、厳密凸錐では見落とされていた腫瘍像画像を検出できたことが、性能向上の理由であると考えられる。

比較手法 2 と提案手法の両手法ともに検出できた腫瘍像

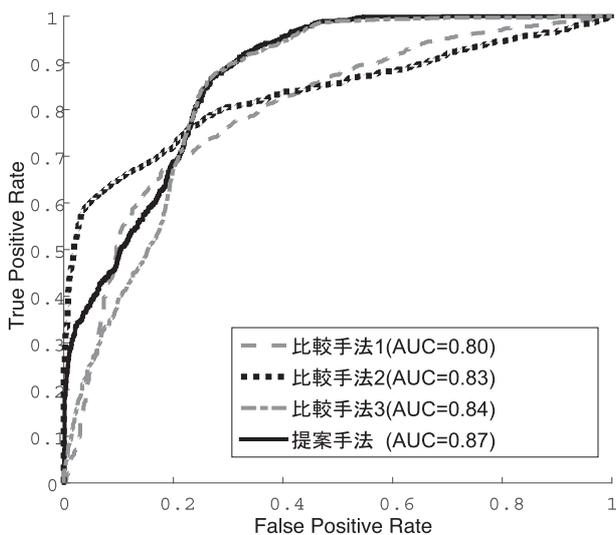


図 10 乳腺超音波画像に対する ROC 曲線

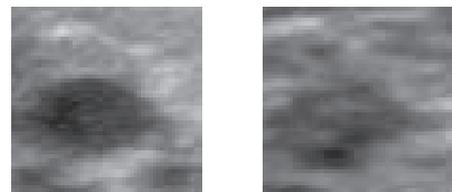
Fig. 10 ROC curves of achieved by the proposal method and comparison methods.

画像の例を図 11 (a) に示す。また、比較手法 2 では見落とされた一方で提案手法により検出された腫瘍像画像の例を図 11 (b) に示す。図 11 (b) は腫瘍像の輪郭が図 11 (a) の腫瘍像と比較して不鮮明であるため、比較手法 2 では正常な乳腺と誤って判定されたと考えられる。一方、提案手法では、錐の表面付近と中心付近を区別できたため、正常と異常を分類する閾値を調整することで、図 11 (b) のような腫瘍像を見落とすことなく検出できる。

4.3.3 異なる被験者に対する汎化能力の検証

乳腺超音波画像は被験者の年齢や体格などにより、乳腺組織や腫瘍像の写り方の様子が異なるが、実際の診断ではこれらの影響を受けずに腫瘍像を検出する必要がある。そこで、学習フェーズで使用する被験者とは異なる被験者に対する提案手法の汎化能力を検証するため、被験者ごとの交差検定を行った。7 人の被験者のうち 6 人の正常な乳腺画像を学習フェーズで使用し、残りの 1 人の被験者の正常な乳腺画像と腫瘍像画像を用いた異常検出の性能評価を、被験者の人数分繰り返した。

各手法の AUC 値を表 2 に示す。表 2 における条件 1 は検査対象の被験者を学習フェーズに使用したときの AUC 値であり、条件 2 は検査対象の被験者を学習フェーズに使用しない場合の AUC 値である。条件 1 と条件 2 を比較すると、どの手法も AUC 値に大きな差がない。つまり、検査対象の被験者を学習フェーズで使用した場合と使用していない場合で同等の異常検出性能であり、いずれの手法も高い汎化能力を有することが分かる。



(a) 両手法で検出された腫瘍像画像の例 (b) 提案手法でのみ検出された腫瘍像画像の例

図 11 検出された腫瘍像画像の例

Fig. 11 Examples of tumor images.

表 2 各被験者に対する AUC 値

Table 2 AUC values in each patients.

	比較手法 1 (部分空間法)		比較手法 2 (厳密凸錐)		比較手法 3 (最尤法)		提案手法	
	条件 1	条件 2	条件 1	条件 2	条件 1	条件 2	条件 1	条件 2
被験者 A	0.93	0.93	0.90	0.85	0.93	0.93	0.93	0.93
被験者 B	0.50	0.47	0.65	0.58	0.74	0.74	0.77	0.75
被験者 C	0.72	0.70	0.56	0.49	0.60	0.59	0.76	0.73
被験者 D	0.83	0.83	0.83	0.82	0.91	0.91	0.91	0.91
被験者 E	0.49	0.49	0.29	0.24	0.22	0.21	0.61	0.51
被験者 F	0.76	0.75	0.79	0.73	0.78	0.78	0.86	0.82
被験者 G	0.59	0.58	0.57	0.54	0.70	0.69	0.83	0.81

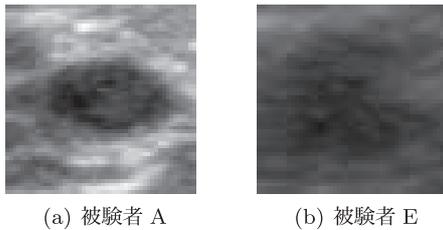


図 12 異なる被験者における腫瘍像画像の比較
Fig. 12 Tumor images of different patients.

一方で、各手法における結果を比較すると、提案手法の AUC 値が最も高い。このことから、提案手法では学習していない被験者に対しても高い異常検出性能を得られることが分かる。

また、すべての手法において、被験者 A と比較して被験者 E の AUC 値は低い。これは、図 12 (a) と図 12 (b) に示すように、被験者 A では腫瘍像の輪郭が鮮明に描出されているのに対し、被験者 E では輪郭が不鮮明であり、HLAC 特徴が被験者 E の腫瘍像の持つ幾何学的情報をとらえることができなかつたためであると考えられる。

5. おわりに

本論文では、錐制約部分空間法を正常データに基づく異常検出手法に適用した場合の問題点を明らかにし、その解決策として錐形状の空間での確率密度の算出方法とそれを基にした異常検出手法を提案した。

錐制約部分空間法では、錐の中心付近と表面付近の区別ができなため、正常クラスを表す錐の広がり大きい場合、錐の表面付近にある異常は正常として誤認識される危険性がある。そこで、提案手法では、正常クラスに属する特徴ベクトルに基づいて錐形状の空間内の確率密度分布を推定し、その空間内での正常データの確率密度を算出する。これにより、特徴ベクトルが錐形状の空間に属する度合いを算出でき、錐の中心付近と表面付近にある特徴ベクトルを区別できるようになるため、従来の錐制約部分空間法と比較して認識性能が向上する。実験では、提案手法の有効性の検証を行うため、2次元空間上で錐を形成する人工データを用いた異常検出実験と、実際の乳腺超音波画像から腫瘍像の検出実験を行った。その結果、提案手法では、通常の錐制約部分空間法と比較して異常検出性能が向上したことを確認した。

参考文献

- [1] Nanri, T. and Otsu, N.: Unsupervised abnormality detection in video surveillance, *MVA*, pp.574-577 (2005).
- [2] 小林 匠, 大津展之: パターン識別のための錐制約部分空間法, 電子情報通信学会論文誌 (D), Vol.92, No.1, pp.104-111 (2009).
- [3] Nosato, H., Kurihara, T., Sakanashi, H., Murakawa, M., Kobayashi, T., Furuya, T., Higuchi, T., Otsu, N., Terai, K. and Hiruta, N.: An extended method of higher-order

local autocorrelation feature extraction for classification of histopathological images, *IP SJ Trans. CVA*, Vol.3, pp.211-221 (2011).

- [4] Otsu, N. and Kurita, T.: *A new scheme for practical flexible and intelligent vision systems*, pp.431-435 (1988).
- [5] Lowe, D.: Object recognition from local scale-invariant features, *Proc. 7th IEEE International Conference on Computer Vision*, Vol.2 (1999).
- [6] Ojala, T., Pietikäinen, M. and Harwood, D.: A comparative study of texture measures with classification based on featured distributions, *Pattern recognition*, Vol.29, pp.51-59 (1996).
- [7] Bro, R. and De Jong, S.: A fast non-negativity-constrained least squares algorithm, *Journal of chemometrics*, Vol.11, No.5, pp.393-401 (1997).
- [8] Silverman, B.W.: *Density estimation for statistics and data analysis*, Vol.26, CRC press (1986).
- [9] 国立がん研究センターがん対策情報センター: 人口動態統計によるがん死亡データ (1958年-2012年) (オンライン), 入手先 (<http://ganjoho.jp/professional/statistics/statistics.html>).
- [10] 国立がん研究センターがん対策情報センター: 地域がん登録全国推計によるがん罹患データ (1975年-2008年) (オンライン), 入手先 (<http://ganjoho.jp/professional/statistics/statistics.html>).
- [11] 藤田広志, 原 武史, 周 向栄, 福岡大輔, 村松千左子, 林 達郎: コンピュータ支援画像診断技術の最先端, 非破壊検査: *Journal of N.D.I.*, Vol.60, No.12, pp.686-693 (2011).
- [12] 福岡大輔, 藤田広志: 超音波工学の基礎乳腺超音波画像における CAD 研究の現状と展望, 乳腺甲状腺超音波医学 (JABTS) *Journal of breast and thyroid sonology*, Vol.2, No.3, pp.18-24 (2013).



山崎 優大

1989年生。2012年東邦大学理学部情報科学科卒業。2014年3月筑波大学大学院システム情報工学研究科知能機能システム専攻博士前期課程修了。同年4月同大学院同研究科同専攻博士後期課程入学。パターン認識技術、医用画像診断支援技術等に関する研究に従事。電子情報通信学会学生会員。



野里 博和 (正会員)

1975年生。2003年東邦大学大学院理学研究科情報科学専攻博士課程修了。理学博士。同年日本学術振興会特別研究員(産業技術総合研究所学振特別研究員)。2006年産業技術総合研究所特別研究員。2011年産業技術総合研究所入所。現在、情報技術研究部門主任研究員。東邦大学連携大学院准教授兼任。最適化アルゴリズム、パターン認識技術、医用画像診断支援技術等に関する研究に従事。2007年電気学会「産学連携による論文」最優秀論文賞受賞。



岩田 昌也 (正会員)

1965年生。1993年大阪大学大学院工学研究科応用物理学専攻後期課程修了。1993~1994年フランス国立航空宇宙研究所ツールーズ研究センター客員研究員。1994年電子技術総合研究所(現、産業技術総合研究所)入所。最適化アルゴリズム、パターン認識の研究に従事。電子情報通信学会会員。



高橋 栄一 (正会員)

1963年生。1993年3月東京大学大学院工学系研究科博士課程修了。同年4月電子技術総合研究所(現、産業技術総合研究所)入所。並列・実時間処理、適応型デジタル回路構成法、情報セキュリティ、医用画像情報処理の研究に従事。東邦大学連携大学院教授兼任。電子情報通信学会、IEEE各会員。博士(工学)。



何森 亜由美

1969年生。1995年香川大学医学部卒業。2005年たけべ乳腺クリニック常勤医。2010年がん研究会がん研究有明病院乳腺センター外科。2011年高松平和病院外科、がん研究会がん研究有明病院乳腺センター非常勤。現在に至る。日本外科学会、日本乳癌学会、日本乳癌検診学会、日本乳腺甲状腺超音波診断医学会、日本臨床細胞学会各会員。第13回日本乳癌学会総会乳癌学会会長賞、2011年度がん研 Best English Paper of the Year 賞。



岩瀬 拓士

1954年生。1981年岐阜大学医学部卒業後、名古屋大学医学部名古屋第一赤十字病院にて外科勤務。1987年癌研究会附属病院にて癌の外科研修および病理部にて乳腺の病理研修。1989年癌研究会附属病院にて外科医員。1996年愛知県がんセンターにて乳腺外科医長。2003年癌研究会附属病院にて乳腺外科副部長。2005年癌研有明病院・レディースセンターにて乳腺科部長。2011年公益財団法人がん研究会がん研有明病院乳腺センター長。現在に至る。日本乳癌学会、日本外科学会、日本癌学会、日本乳癌検診学会各会員。



坂無 英徳 (正会員)

1996年北海道大学大学院工学系研究科博士課程修了。同年日本学術振興会特別研究員。1998年通商産業省工業技術院電子技術総合研究所(現、(独)産業技術総合研究所)入所。現在、情報技術研究部門主任研究員。筑波大学大学院准教授(連携大学院)、東邦大学客員教授。博士(工学)。医用画像処理、コンピュータ診断支援、適応学習手法に関する研究に従事。日本熱帯医学会会員。