

УДК 612.12: 004.93

Отображение и анализ многомерных медицинских данных в ортогональных информационных базисах

А. М. Ахметшин, К. А. Ахметшин

Днепропетровский национальный университет, Украина

Резюме

Рассмотрены информационные возможности нового метода качественного анализа многоканальных электрокардиограмм в фазовом пространстве, образованном ортонормированными векторами сингулярного разложения и многопараметровых медицинских изображений. Метод отличается чувствительностью к выделению визуально неразличимых морфологических особенностей анализируемых многопараметровых изображений.

Ключевые слова: многоканальная электрокардиограмма, фазовый портрет, сингулярное разложение, многопараметровое изображение, чувствительность.

Клин. информат. и Телемед.**2012. Т.8. Вып.9. с.30–34**

Вступление

Современные подходы к повышению достоверности медицинской диагностики базируются на использовании многомерных (многоканальных) диагностических систем (электрокардиографы – ЭКГ, ядерная магнитная томография, дерматология и др.). Основная проблема анализа многомерных данных базируется на двух особенностях:

- эти данные являются, как правило, коррелированными, что диктует необходимость использования процедур декорреляции, связанной с обеспечением сжатия данных в целях сокращения их избыточности;
- в любом случае, многомерные данные должны анализироваться как единое целое, вне зависимости от того используются ли процедуры сжатия данных или нет.

Впервые подобного рода задачи возникли в статистических методах прикладного анализа [1, 2], однако в задачах медицинской диагностики подобные подходы редко приводят к успеху из-за недостаточности объема данных применительно к одному пациенту. Более современный подход базируется на использовании нейросетевой диагностики [3], однако и здесь постоянно возникают трудности связанные с количеством и качеством баз данных, необходимых для предварительного обучения нейронной сети.

Более привлекательными по нашему мнению, по крайней мере с точки зрения простоты программной реализации и простоты восприятия, являются алгебраические методы, связанные с представлением многомерных данных в новых информационных базисах.

Впервые такой подход был использован в [4], однако с тех пор появилось много новых идей и методов.

Целью данной работы является демонстрация информационных возможностей представления и анализа многомерных данных (сигналы и изображения) в ортогональных информационных базисах на примерах 15-ти канальных ЭКГ и цветного дерматологического изображения.

Материалы, методы исследования, анализ результатов

1. Отображение и анализ многоканальных ЭКГ

Вопросам исследования и цифровой обработки ЭКГ посвящено множество работ, однако они в основном касаются анализа одноканальных ЭКГ [5, 6], тогда как в настоящее время основной интерес представляют многоканальные ЭКГ. Образцы 15-ти канальных ЭКГ с установленными диагнозами были получены из международной базы данных «Physionet».

На рис. 1а представлен «моментальный временной срез» многомерной ЭКГ (диагноз – инфаркт), а на рис. 1б

представлен график коэффициентов корреляции первого канала с четырнадцатью остальными. Из рассмотрения рис. 1 следует, что первый измерительный канал сильно коррелирован со 2, 3, 4, 5, 7, 14 и 15 каналами, т. е. данные являются избыточными, что затрудняет их интерпретацию.

Прямой путь решения задачи заключается в переходе к новому ортогональному информационному базису и проецировании многомерных данных в этот базис с последующим использованием только тех данных, проекции которых являются максимальными по отношению к другим проекциям.

Возможно использование множества различных методов построения такого базиса, однако наиболее рациональным представляется использование метода сингулярного разложения матрицы данных Z , размерностью $(M \times N)$ [7, 8],

$$Z = USV^T, \quad (1)$$

где $M \gg N$ (M – число отсчетов в одном канале ЭКГ, а N – число измерительных каналов). В (1) U – ортонормированная матрица сингулярных векторов, образующих основу нового информационного базиса, а нормированная матрица сингулярных чисел

$$S = \begin{bmatrix} \hat{s}_1 & & \\ & \hat{s}_2 & \\ & & \hat{s}_N \end{bmatrix}$$

$$\hat{s}_i = \frac{s_i}{\sum_i s_i} \times 100\%, i = 1, \dots, N \quad (2)$$

характеризует относительную величину проекций (ранг) матрицы данных Z в новом информационном базисе.

На рис. 2а представлен график спектра нормированных сингулярных чисел, а на рис. 2б – график сингулярного вектора, соответствующего максимальному сингулярному числу.

Из анализа рис. 2а следует, что первый сингулярный вектор содержит 99,92% в сей информации, содержащейся в 15-ти канальной ЭКГ, т. е. для диагностики достаточным является рассмотрение лишь первого (т. е. по сути дела одноканального) сингулярного вектора, представленного на рис. 2б.

Особый интерес в данном случае представляет синтез фазовых портретов, характеризующих многомерную измерительную систему. На рис. 3а представлен фазовый портрет в традиционной системе координат $U_1(t) \div \text{gradient}[U_1(t)]$, а на рис. 3б в системе координат $[U_1(t) \div U_2(t)]$.

Из рассмотрения рис. 3а следует, что использование операции вычисления первой производной приводит к увеличению влияния шумовых факторов,

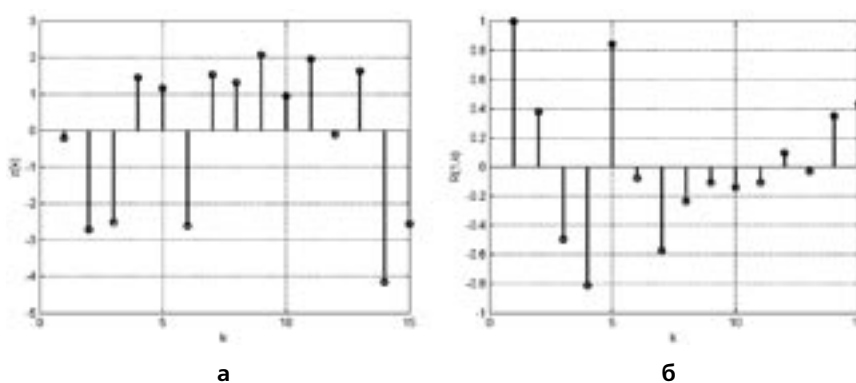


Рис. 1. «Моментальный» временной срез 15-ти канальной ЭКГ (а) и зависимость коэффициентов корреляции первого измерительного канала от четырнадцати остальных (б).

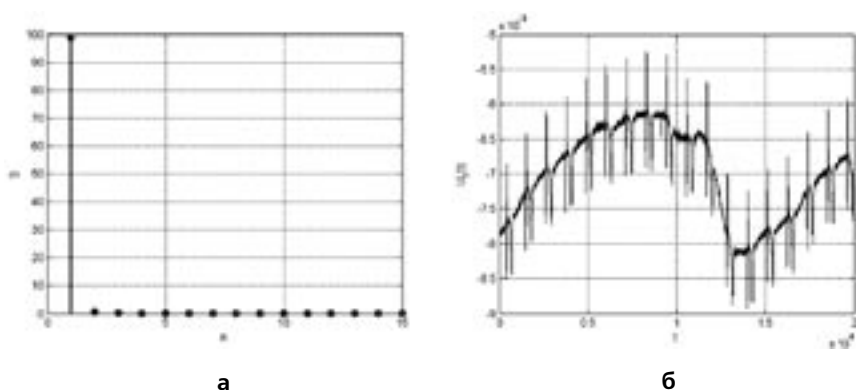


Рис. 2. Спектр нормированных сингулярных чисел (а) и первый сингулярный вектор $U_1(t)$ (б) соответствующий первому – наибольшему сингулярному числу.

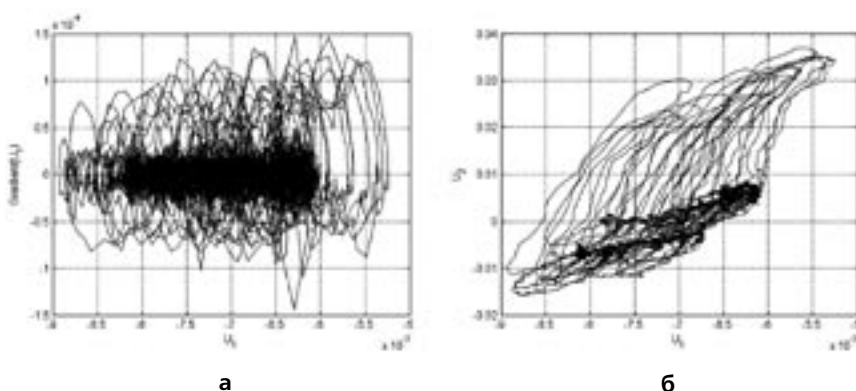


Рис. 3. Фазовый портрет многомерной ЭКГ построенный в системах координат: а – $U_1(t) \div \text{gradient}[U_1(t)]$; б – $[U_1(t) \div U_2(t)]$.

тогда как синтез фазового портрета в системе координат первых двух сингулярных векторов шумовой фактор в значительной степени нивелирует.

В целях практически полного устранения влияния шума возможно ис-

пользование методики синтеза фазовых портретов на основе [9] сингулярного разложения вложенных векторов, соответствующих первому сингулярному вектору $U_1(t)$, т.е. решается задача сингулярного разложения второго порядка,

при котором на основе $U_1(t)$ формируется новая матрица вида

$$U1 = \begin{bmatrix} u_1(t_1) & u_1(t_2) & \dots & u_1(t_L) \\ u_1(t_2) & u_1(t_3) & \dots & u_1(t_{L+1}) \\ \dots & \dots & \dots & \dots \\ u_1(t_{M-L+1}) & u_1(t_{M-L+2}) & \dots & u_1(t_M) \end{bmatrix}, \quad (3)$$

где L – глубина вложения (в экспериментах выбиралась равной $L = 30$), с последующим использованием вторичного сингулярного разложения матрицы $U1$ на основе выражения

$$U1 = QWP^T. \quad (4)$$

На рис. 4 представлены фазовые портреты ЭКГ синтезированные в системах координат $Q_1 \div Q_2$ и $Q_3 \div Q_4$.

Сопоставление рис. 3 и рис. 4 показывает, что использование метода сингулярного разложения второго порядка позволяет нейтрализовать влияние шума и более четко выделить типовые особенности фазовых портретов, связанных с «расслоением» фазовых траекторий, характерных для инфарктных состояний.

Конечно, не всегда ранг матрицы наблюдений Z равен единице. Так при анализе одних из данных соответствующих диагнозу «сердечная недостаточность», ранг оказался равным трем, т. е. при синтезе фазовых портретов первого порядка необходимо использовать систему координат $[U_1 \div U_2 \div U_3]$ (рис. 5а), однако и в данном случае, как показали результаты экспериментальных исследований, для более точного выделения топологических особенностей фазовых портретов многоканальных ЭКГ возможно использование метода вторичного сингулярного разложения первого сингулярного вектора $U_1(t)$ (рис. 5б), а в ряде случаев и второго сингулярного вектора $U_2(t)$.

Из рассмотрения рис. 5 следует, что как и в предыдущем случае, использование метода вторичного сингулярного разложения позволяет более четко выделить топологические особенности трехмерного фазового портрета многоканальной ЭКГ.

2. Анализ многопараметровых медицинских изображений

Отображение и анализ многопараметровых медицинских изображений в ортогональных базисах позволяет зачастую выявить визуально неразличимые особенности таких изображений.

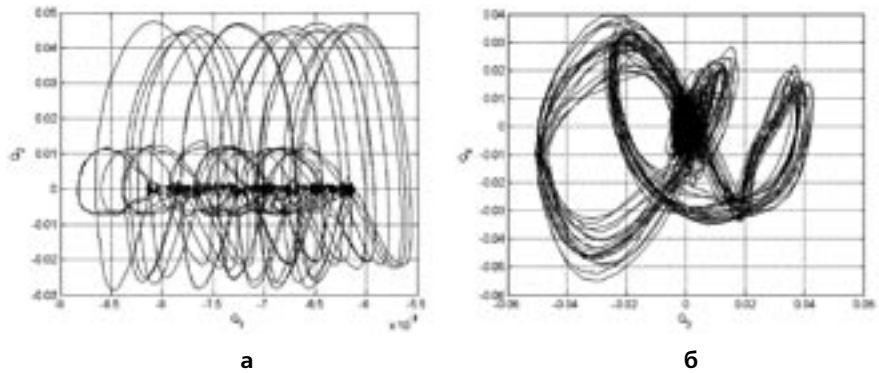


Рис. 4. Фазовые портреты синтезированные методом сингулярного разложения второго порядка вектора представленного на рис. 2б: а – система координат $Q_1 \div Q_2$; б – система координат $Q_3 \div Q_4$.

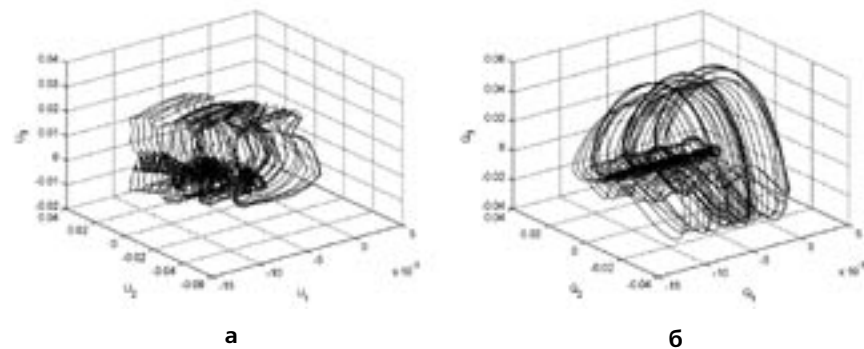


Рис. 5. Трехмерные фазовые портреты при ранге матрицы измерений Z равным трем: а – система координат $U_1 \div U_2 \div U_3$; б – система координат $Q_1 \div Q_2 \div Q_3$.

Так на рис. 6 представлена численная модель трех-параметрового изображения, на третьем из которых имеется семь скрытых участков.

Спектр нормированных сингулярных чисел моделируемого ансамбля равен $\hat{S}_1 = 85.3405\%$; $\hat{S}_2 = 14.6582\%$; $\hat{S}_3 = 0.0012\%$. В соответствии с вышеизложенным анализом информационных возможностей метода сингулярного разложения, третье «собственное изображение» [10], дающее лишь 0.0012% от общей информативности моделируемого ансамбля должно быть отброшено. Однако, как показывает анализ ансамбля «собственных изображений» на рис. 7, именно на этом третьем изображении и выявляются скрытые особенности моделируемого ансамбля.

Данный факт демонстрирует информационную значимость данных соответствующим минимальным сингулярным числам в задачах повышения чувствительности визуального анализа многопараметровых изображений.

На рис. 8 представлены три RGB канала цветного дерматологического изображения (диагноз – рак кожи;

изображение предоставлено кафедрой дерматологии Днепропетровской медицинской академии).

Анализ корреляционной матрицы C_{RGB} исходного изображения

$$C_{RGB} = \begin{bmatrix} 1.00 & 0.89 & 0.84 \\ 0.89 & 1.00 & 0.97 \\ 0.84 & 0.97 & 1.00 \end{bmatrix}$$

свидетельствует о том, что с информационной точки зрения изображения на рис. 8 очень похожи, т.е. избыточны. Спектр нормированных сингулярных чисел RGB ансамбля равен $\hat{S}_1 = 99.7157\%$; $\hat{S}_2 = 0.2638\%$; $\hat{S}_3 = 0.0012\%$. Следовательно, в первом собственном изображении (рис. 9.1) сосредоточено 99.7% все информации RGB ансамбля на рис. 8.

Однако, с точки зрения морфологического анализа, именно третье собственное изображение (обеспечивающее вклад лишь 0.0012%) позволяет выделить скрытые особенности формы патологии.

На рис. 10а представлено цветное изображение гемангиомы, приведенное

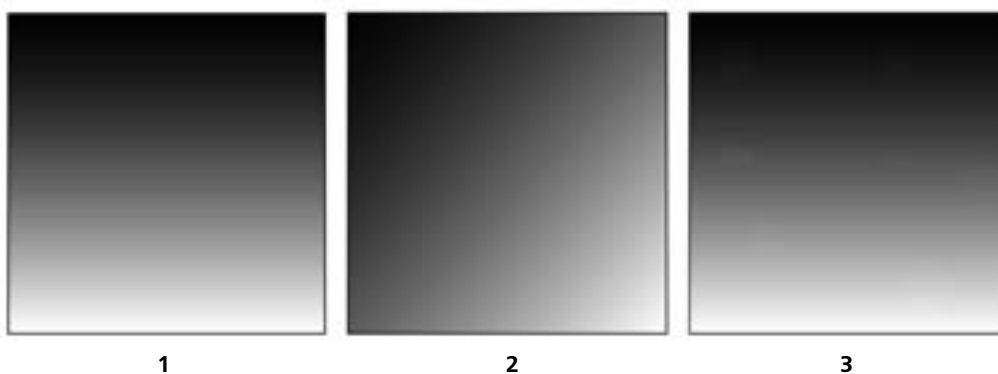


Рис. 6. Численная модель трех-параметрового изображения.

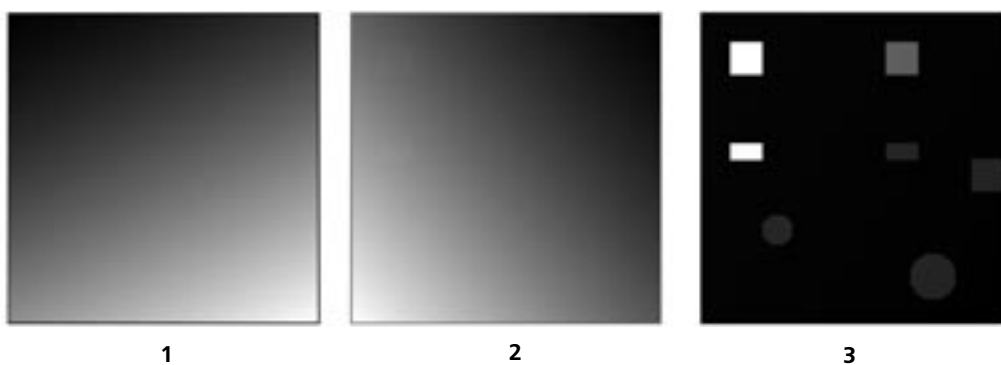


Рис. 7. Собственные изображения ансамбля на рис. 6.

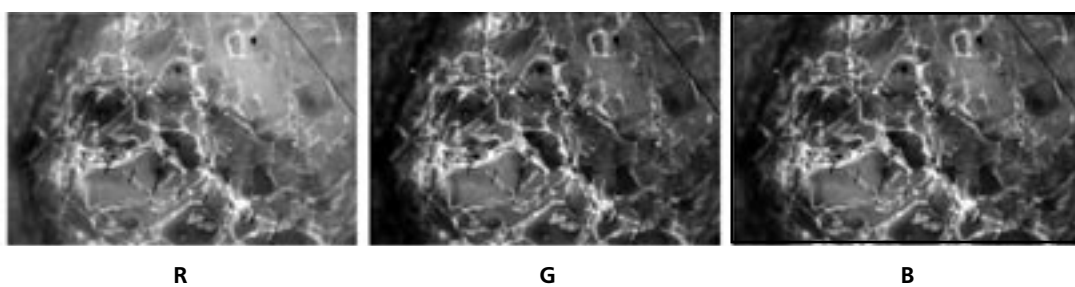


Рис. 8. RGB каналы цветного дерматологического изображения.

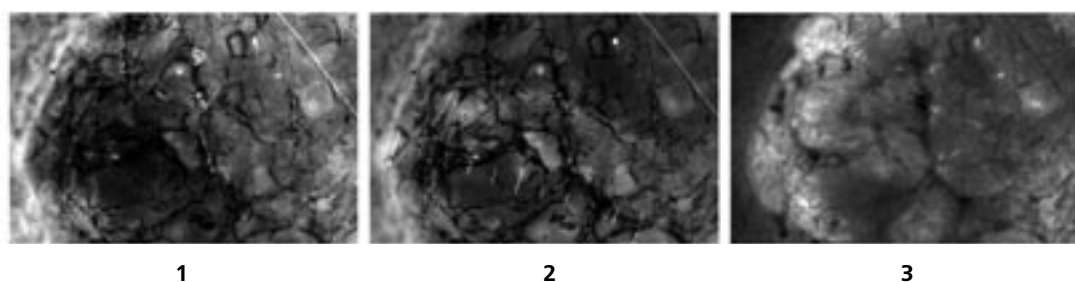


Рис. 9. Собственные изображения RGB ансамбля на рис. 8.

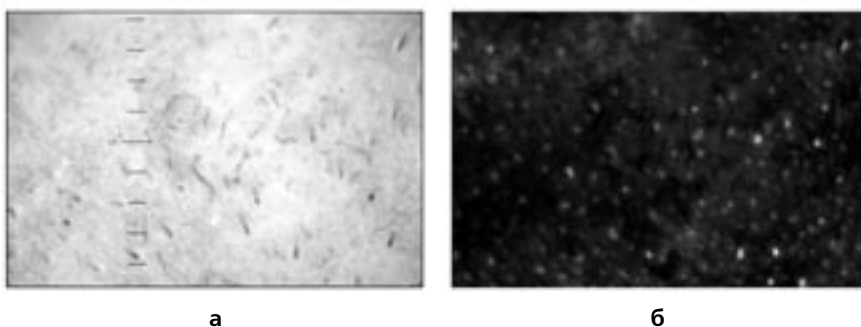


Рис. 10. Изображение гемангиомы: а — оригинал; б — третье собственное изображение оригинала.

к градиентам серого, а на рис. 10б — третье собственное изображение (вклад 0.0063%).

На исходном изображении практически невидимы характерные особенности заболевания, тогда как на третьем собственном изображении эти особенности (белые точки) четко проявились.

Однако следует подчеркнуть, что в процессе экспериментальных исследований (анализировались порядка 50 различных типов цветных дерматологических изображений) лишь приблизительно в 15% проявилась информационная значимость третьего собственного изображения. Данный факт подчеркивает необходимость сочетания различных методов цифровой обработки медицинских изображений в целях повышения достоверности диагностики.

Заключение

1. Применение метода сингулярного разложения многомерных данных, связанного с переходом к новому ортонормированному информационному базису, позволяет синтезировать обобщенные фазовые портреты многоканальных ЭКГ.

2. Использование метода вторичного сингулярного разложения матрицы вложенных векторов, соответствующих первому сингулярному вектору анализируемого ансамбля данных, позволяет более четко выявить характерные топологические особенности обобщенных фазовых портретов многоканальных ЭКГ.

3. Экспериментально установлено, что при анализе многомерных медицинских изображений в ряде случаев наиболее информативными, с точки зрения выявления характерных морфологических признаков участков

потенциального интереса, являются собственные изображения соответствующие наименьшим сингулярным числам, т.е. дающим наименьший вклад в общую изменчивость анализируемого ансамбля данных.

Литература

1. Айвазян С. А. Прикладная статистика. Классификация и снижение размерности / Айвазян С. А., Бухштабер В. М., Енюков Е. С. – М.: Финансы и статистика, 1989. – 607 с.
2. Фукунага К. Введение в статистическую теорию распознавания образов / Фукунага К.; [пер. И. Ш. Торговицкого]. – М.: Наука, 1979. – 367 с.
3. Кохонен Т. Самоорганизующиеся карты / Кохонен Т.; [пер. с англ. В. Н. Агеева] – М.: БИНОМ, 2008. 655 с.
4. Ахмед Н. Ортогональные преобразования при обработке цифровых сигналов / Ахмед Н., Рао К. Р.; [пер. с англ. Т. Э. Кренкеля]. – М.: Связь, 1980. – 248 с.
5. Рангайян Р. М. Анализ биомедицинских сигналов: практический подход / Рангайян Р. М.; [пер. с англ. А. П. Немирко]. – М.: ФИЗМАТЛИТ, 2007. – 440 с.
6. Файнзильберг Л. С. Информационные технологии обработки сигналов сложной формы: теория и практика / Файнзильберг Л. С.; – Киев, Наукова думка, 2008. – 333 с.
7. Беклемишев Д. В. Дополнительные главы линейной алгебры / Беклемишев Д. В. – М.: Наука, 1983. – 455 с.
8. Castells F. Principal Component Analysis in ECG Signal Processing / Castells F., Laguna P., Bollmann A. // EURASIP Journal on Advances in Signal Processing. – 2007. – Article ID 74580. – 21p.
9. Ахметшин А. М. Информационные возможности анализа и отображения электрокардиограмм в базисах сингулярного разложения вложенных векторов / Ахметшин А. М., Ахметшин К. А. // Клиническая информатика и телемедицина. – 2011. – Т. 7, – № 8. – С. 58-64.

10. Форсайт Д., Понс Ж. Компьютерное зрение: современный подход / Форсайт Д., Понс Ж.; [пер. с англ. А. В. Назаренко, И. Ю. Дорошенко]. – М.: Вильямс, 2004. – 926 с.

Representation and analysis of multidimensional medical data at orthogonal information bases

A. M. Akhmetshyn, K. A. Akhmetshyn
Dnepropetrovsk National University
Ukraine

Abstract

Information possibilities of a new method multidimensional electrocardiogram quality analysis in a phase space and multiparameter medical images are considered. The method provides by sensitivity to separation of visual invisible morphology features of analysed multiparameters images.

Key words: multichannel electrocardiogram, phase portrait, singular decomposition, multiparameters image, sensitivity.

Відображення та аналіз багатовимірних медичних даних у ортогональних інформаційних базисах

О. М. Ахметшин, К. О. Ахметшин
Дніпропетрівський національний університет, Україна

Резюме

Розглянуті інформаційні можливості нового методу якісного аналізу багатовимірних електрокардіограм у багатовимірному фазовому просторі утвореним ортонормованими векторами сингулярного розкладу та багатовимірних медичних зображень. Метод відрізняється чутливістю до виділення візуально нерозрізнених морфологічних особливостей аналізованих багатовимірних зображень.

Ключові слова: багатоканальна електрокардіограма, фазовий портрет, сингулярний розклад, багатовимірне зображення, чутливість.

Переписка

д. физ.-мат. наук
профессор **А. М. Ахметшин**
Днепропетровский национальный университет
пр. Гагарина, 72
Днепропетровск, 49000, Украина
эл. почта: akhmlu@mail.ru