

Сороко Л.М. и Стриж Т.А.
Б1-10-8364.

+✓

С 344. 1.011
С-655



ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

189/75

Б1-10-8364

ДЕПОНИРОВАННАЯ ПУБЛИКАЦИЯ

Дубна 19

74

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ
Лаборатория ядерных проблем
Лаборатория вычислительной техники и автоматизации

Б1-10-8364

Л.М.Сороко, Т.А.Стриж

с344.1м

С-655

СЖАТИЕ ИНФОРМАЦИИ С ПОМОЩЬЮ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ
(преобразование Карунена-Лоева)

5 ноября 74

Объединенный институт
ядерных исследований
БИБЛИОТЕКА

- Дубна, 1974 год -

А Н Н О Т А Ц И Я

Изложен метод главных компонент (преобразование Карунена-Лоеве). Даны примеры использования этого метода, в частности, для предварительной сортировки данных в искровых спектрометрах.

Введение

Поиск эффективных способов хранения, передачи и анализа сложных изображений является одной из наиболее актуальных проблем многих разделов современной науки. В связи с расширением возможностей вычислительной техники широкое распространение получают цифровые методы обработки изображений, которые используются для кодирования изображений и передачи их по цифровым каналам связи, для реставрации искаженных фотографий, повышения их качества и быстрого извлечения полезной информации из двумерных массивов данных. Такие методы используют при картографировании с помощью искусственных спутников небесных тел, для визуализации объектов, заданных аналитически или в виде голограмм, в рентгеновской и гамма-лучевой дефектоскопии, в частности, элементов ядерных реакторов, в биофизике, медицине, геофизике и акустике.

Современные методы получения информации о взаимодействии элементарных частиц и ядерных реакциях характеризуются огромным объемом двумерной и многомерной информации, поступающей с крупных экспериментальных установок со все возрастающей скоростью. Очевидно, применение последних достижений техники цифровой обработки изображений дало бы существенный выигрыш во времени, если, например, использовать быстрое считывание трековой информации, производить сжатие информации без потери полезной информации, а также увеличить скорость обработки событий.

В ходе исследований, проведенных в последнее время методами цифровой обработки изображений, было обнаружено, что использование интегрального образа изображения вместо самого изображения существенно повышает эффективность обработки. При выборе определенного интегрального преобразования взаимная корреляция между компонентами образа оказываются очень близкой к нулю. Сейчас получены данные о свойствах систем дискретных ортогональных базисов, таких как Фурье, Уолш-Адамара и ряда других^{/1 + 4/}. Эти преобразования, однако, являются всего лишь квазиоптимальными, поскольку компоненты образа в этих базисах не являются полностью некоррелированными. Другое направление исследований связано с дискретным преобразованием Карунена-Лоеве^{/5/}. Компоненты преобразованного таким образом изображения не коррелируют между собой и обладают тем свойством, что часть компонент образа может быть вырошена без заметных искажений исходных данных.

Цель данной работы - подробно изложить метод главных компонент с использованием преобразования Карунена-Лоеве и продемонстрировать возможное применение этого метода для предварительной сортировки данных в искровых спектрометрах.

Дискретное преобразование Карунена-Лоеве

Начнем с истории вопроса. В 1933 г. Хоттелинг^{/6/} нашел преобразование, которое позволяет перейти от массива дискретных

компонент к переменным, которые являются взаимно некоррелированными. Такой подход получил название метода главных компонент. Вариант преобразований для непрерывных переменных открыли Карунен и Лоеве. Повторно это преобразование формулировалось также в работах других исследователей. Первоначально метод главных компонент использовали для представления случайных процессов рядами. При этом исходные данные сводились к массиву данных, размерность которого была меньше размерности исходного массива, но в котором содержалась практически та же информация, что в первоначальных данных.

Пусть \vec{X} — измеряемый случайный вектор, состоящий из n компонент. Каждая его компонента принимает некоторые преимущественные значения, заключенные в определенных пределах. Подвергаем вектор \vec{X} некоторому линейному преобразованию:

$$\vec{X} = \sum_{i=1}^n y_i \vec{W}_i = \hat{W} \vec{Y}$$

где преобразованный случайный вектор \vec{Y} имеет компоненты

$$\vec{Y} = [y_1 \ y_2 \ \dots \ y_n]^T$$

а детерминистская матрица преобразования \hat{W} имеет линейно независимые векторы-столбцы

$$\hat{W} = [\vec{W}_1, \vec{W}_2, \dots, \vec{W}_n], \quad \det |\hat{W}| \neq 0$$

Условие линейной независимости можно заменить требованием взаимной ортогональности векторов-столбцов:

$$\vec{W}_i^T \vec{W}_j = \begin{cases} 1, & i=j \\ 0 & i \neq j \end{cases}$$

Тогда компоненты вектора \vec{Y} равны:

$$y_i = W_i^T X \quad (i = 1, 2, \dots, n)$$

Векторы \vec{X} и \vec{Y} являются взаимными ортогональными образами друг друга.

Рассмотрим оценку случайного вектора \vec{X} , когда взято только \bar{m} компонент вектора \vec{Y} , и $\bar{m} < n$. Отброшенные $(n - \bar{m})$ компонент вектора \vec{Y} в общем случае заменим некоторыми постоянными величинами β_i ($i = m + 1, m + 2, \dots, n$).

). Запишем эту оценку в виде:

$$\vec{X}(m) = \sum_{i=1}^m y_i W_i + \sum_{i=m+1}^n \beta_i W_i$$

Вносимая при такой аппроксимации ошибка восстановления равна

$$\Delta \vec{X}(m) = \vec{X} - \vec{X}(m) = \sum_{i=m+1}^n (y_i - \beta_i) W_i$$

а соответствующая ей средне-квадратичная ошибка равна

$$\begin{aligned} \overline{\varepsilon^2(m)} &= E \left\{ \|\Delta \vec{X}(m)\|^2 \right\} = \\ &= E \left\{ \sum_{i=m+1}^n \sum_{j=m+1}^n (y_i - \beta_i)(y_i - \beta_i)^T W_j^T \right\} = \\ &= \sum_{i=m+1}^n E \left\{ (y_i - \beta_i)^2 \right\} \end{aligned}$$

По определению, оптимальным ортогональным базисом называют такой базис, который минимизирует средне-квадратичную ошибку, а постоянные величины b_i равны математическим ожиданиям компонент вектора \vec{Y} :

$$b_i = E\{y_i\} = W_i^T E\{\vec{X}\}$$

Средне-квадратичную ошибку можно записать в виде:

$$\begin{aligned} \overline{\varepsilon^2(m)} &= \sum_{i=m+1}^n E [y_i - [E\{y_i\}]]^2 \\ &= \sum_{i=m+1}^n W_i^T E [(\vec{X} - E\{\vec{X}\})(\vec{X} - E\{\vec{X}\})^T] W_i = \\ &= \sum_{i=m+1}^n W_i^T R_x W_i, \end{aligned}$$

где R_x - ковариационная матрица случайного процесса, описываемого вектором \vec{X} :

$$R_x = E [(\vec{X} - E\{\vec{X}\})(\vec{X} - E\{\vec{X}\})^T]$$

Теперь найдем оптимальный ортогональный базис векторов \vec{W}_i ($i = 1, 2, \dots, n$). Докажем, что собственные векторы матрицы R_x являются теми векторами, которые минимизируют ошибку $\overline{\varepsilon^2(m)}$. Повсюду мы находимся в рамках линейных ортогональных преобразований.

Поскольку матрицы Ψ и W являются ортогональными, то

$$(\Psi^T W)^T (\Psi^T W) = (\Psi^T W) (\Psi^T W)^T = I$$

где I - единичная матрица. Заново разобьем матрицу H на подматрицы по схеме

$$H = \begin{matrix} & \begin{matrix} \overbrace{}^m & \overbrace{}^{n-m} \end{matrix} \\ \begin{matrix} m \\ n-m \end{matrix} & \left[\begin{array}{cc} K_1 & K_2 \\ K_3 & K_4 \end{array} \right] \end{matrix} \begin{matrix} & \begin{matrix} \overbrace{}^m & \overbrace{}^{n-m} \end{matrix} \\ & \left[\begin{array}{cc} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{array} \right] \end{matrix} \begin{matrix} & \begin{matrix} \overbrace{}^m & \overbrace{}^{n-m} \end{matrix} \\ & \left[\begin{array}{cc} K_1^T & K_3^T \\ K_2^T & K_4^T \end{array} \right] \end{matrix}$$

Тогда

$$\overline{\varepsilon^2(m)} = \text{Sp } H_{22} = \text{Sp} (K_3 \Lambda_1 K_3^T + K_4 \Lambda_2 K_4^T)$$

Для того чтобы выбранный ортогональный базис, образованный соответственными векторами матрицы ковариации, был оптимальным, достаточно, чтобы

$$\text{Sp } \Lambda_2 \leq \text{Sp} (K_3 \Lambda_1 K_3^T + K_4 \Lambda_2 K_4^T)$$

Из условия ортогональности $K = \Psi^T W$ получаем:

$$\left. \begin{aligned} K_1 K_1^T + K_2 K_2^T &= I \\ K_1^T K_1 + K_3^T K_3 &= I \\ K_2^T K_2 + K_4^T K_4 &= I \end{aligned} \right\}$$

Далее доказательство ведем следующим образом:

$$\begin{aligned}
 \text{Sp} (K_3 \Lambda_1 K_3^T + K_4 \Lambda_2 K_4^T) &= \text{Sp} (\Lambda_1 K_3^T K_3 + \Lambda_2 K_4^T K_4) = \\
 &= \text{Sp} \Lambda_2 + \text{Sp} (\Lambda_1 K_3^T K_3 - \Lambda_2 K_2^T K_2) = \\
 &\geq \text{Sp} \Lambda_2 + \{ \lambda_m \text{Sp} (K_3^T K_3) - \lambda_{m+1} \text{Sp} (K_4^T K_4) \} = \\
 &= \text{Sp} \Lambda_2 + \{ \lambda_m \text{Sp} [I - K_1^T K_1] - \lambda_{m+1} \text{Sp} [I - K_1 K_1^T] \} = \\
 &= \text{Sp} \Lambda_2 + (\lambda_m - \lambda_{m+1}) \text{Sp} [I - K_1^T K_1] = \\
 &= \text{Sp} \Lambda_2 + (\lambda_m - \lambda_{m+1}) \text{Sp} (K_3^T K_3) \geq \text{Sp} \Lambda_2
 \end{aligned}$$

Первое неравенство следует из того факта, что λ_m - наименьшее собственное значение в диагональной матрице Λ_1 , а λ_{m+1} - наибольшее собственное значение в матрице Λ_2 . Если $\Psi = W$, то $K_1 = K_4 = I$, $K_2 = K_3 = 0$ и неравенство переходит в равенство.

Итак, оптимальный ортогональный базис векторов является тот, который удовлетворяет условию:

$$R_x W_i = \Lambda_i W_i$$

и состоит из собственных векторов матрицы ковариации R_x заданного случайного процесса. При этом минимальная средне-квадратичная ошибка равна

$$\overline{\varepsilon^2(m)}_{opt} = \sum_{i=m+1}^n \lambda_i$$

Разложение случайного вектора по собственным векторам матрицы ковариации этого случайного процесса называют дискретным преобразованием Карунена-Лоеве. Это преобразование оптимально не только по критерию минимума средне-квадратичной ошибки, но также и по другим критериям, например, меры рассеяния или энтропии заселенности /II/.

Значимость каждого информационного признака в представлении Карунена-Лоеве определяется собственным значением матрицы ковариации. Если некоторый признак y_i отбрасывается, то средне-квадратичная ошибка аппроксимации возрастает на величину λ_i . Упорядочению собственных значений

$$\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$$

соответствует упорядоченное расположение соответствующих информационных признаков. Информационные признаки y_i взаимно некоррелированы. Это означает, что матрица ковариации преобразованного вектора \vec{Y} является диагональной.

Примеры преобразования Карунена-Лоеве

Процедура кодирования случайной последовательности с помощью метода главных компонент была опробована на ряде примеров.

Пример I. Рассмотрим случайную последовательность $f_0(i)$ из 4 компонент, средние значения которых равны

$f_0(0)$	$f_0(1)$	$f_0(2)$	$f_0(3)$
+ 1.00	-1.20	-0.78	+0.98

Возьмем только четыре реализации этого случайного процесса:

Номер реализации \ i	0	1	2	3
1	0.9949	-1.0984	-0.6834	0.8428
2	0.944	-1.3892	-1.0239	0.8649
3	1.2047	-1.1897	-0.7822	0.7501
4	1.0104	-1.1249	-0.8236	1.1005

Матрица корреляции этого процесса равна

$$R_c = \begin{bmatrix} 1.0883 & -1.2435 & -0.8552 & 0.9177 \\ -1.2435 & 1.4543 & 1.0075 & -1.0644 \\ -0.8552 & 1.0075 & 0.7014 & -0.7387 \\ 0.9177 & -1.0644 & -0.7387 & 0.8081 \end{bmatrix}$$

Ее собственные векторы и собственные значения равны

$$W = \begin{bmatrix} -0.5175 & 0.06012 & 0.4162 & -0.4444 \\ 0.7259 & 0.3792 & 0.5454 & 0.1784 \\ -0.3900 & 0.2873 & 0.0334 & 0.8742 \\ -0.2305 & -0.6420 & 0.7268 & 0.805 \end{bmatrix}$$

и

$$\Lambda_i = (4.0088; 0.0218; 0.0206; 0.0008).$$

Вычисления собственных векторов и собственных значений проводились с помощью стандартной программы MATRIX из библиотеки программ на ЭВМ БЭСМ-6.

На рис.1 приведены четыре собственных вектора $w_0(i)$, $w_1(i)$, $w_2(i)$, $w_3(i)$ и четыре решетчатых функции Уолша $h_0(i)$, $h_1(i)$, $h_2(i)$, $h_3(i)$.

Разложим в качестве примера вторую реализацию процесса, $f_2(i)$, по собственным векторам матрицы ковариации R_c и для сравнения по функциям Уолша. Компоненты преобразованных векторов равны (рис.2)

k	0	1	2	3
$k-L$	2,13	0,25	0,05	0,00
Walsh	0,30	0,15	2,11	0,22

Видно, что компоненты разложения для преобразования Карунена-Лоеве монотонно спадают с ростом индекса k . В противоположность этому, компоненты разложения по системе функций Уолша в секвентивном представлении $/I2/$ изменяются с k нерегулярно. Чтобы получить монотонный спад компоненты Уолш-образа, их необходимо перетасовать. Тогда они будут расположены в следующем порядке: $k = 2; 0; 3; 1$. Ему соответствует массив функций Уолша вида

$k \downarrow$	+	-	-	+	\longrightarrow	i	$S = 2$
	+	+	+	+			$S = 0$
	+	-	+	-			$S = 3$
	+	+	-	-			$S = 1$

где S - секвента функций Уолша. Такую последовательность функций Уолша, специфическую для каждой конкретной задачи, называют последовательностью Карунена-Лоеве.

На рис.3 приведены значения средне-квадратичных ошибок восстановления исходной последовательности при использовании различного числа членов разложения:

- 1) для разложения по системе собственных векторов;
- 2) для разложения по системе функций Уолша в секвентивном представлении;
- 3) для разложения по системе функций Уолша, перетасованных в порядке Карунена-Лоеве.

Пример 2. Рассмотрим случайный процесс, который описывается моделью Маркова. Его матрица ковариации равна

$$R(x, x') = \exp[-\alpha |x - x'|]$$

В табл. I приведены общие выражения для средне-квадратичных ошибок усечения при восстановлении случайного процесса методом Карунена-Лоеве, Фурье, Уолша-Адамара и для растрового сканирования.

Таблица I

Преобразование	Ошибка $\epsilon^2 = 1 - \frac{1}{N} \sum_{j=1}^N \sigma_j^2$ для марковского процесса
Карунена-Лоеве	$\sigma_j^2 = \lambda_j, \quad j = 1, 2, \dots, N,$ λ_j - собственные значения матрицы ковариации R_x
Фурье	$\sigma_j^2 = \frac{2\alpha N}{(\alpha N)^2 + (2\pi j)^2} + \frac{2(e^{-\alpha N} - 1) [(\alpha N)^2 - (2\pi j)^2]}{[(\alpha N)^2 + (2\pi j)^2]^2}$
Уолша-Адамара	$\sigma_j^2 = \sum_{x=0}^{N-1} \sum_{x'=0}^{N-1} e^{-\alpha x-x' } [-1] \sum_{i=0}^{\log_2 N - 1} \psi_j^b(i) [x_i^b + x_{i'}^b]$ x^b - двоичное представление чисел x $\psi_j^b(i)$ - число j в коде Грея
Растровое сканирование	$\sigma_j^2 = \frac{2}{\alpha^2 N} [\alpha N + \sqrt{N} (e^{-\alpha N \sqrt{N}} - 1)]$

Средне-квадратичные ошибки усечения при различных значениях $\rho = e^{-\alpha}$ ($\rho_1 = 0,8$; $\rho_2 = 0,5$) для $N = 32$ приведены на рис.4. Для преобразований: Карунена-Лоеве, Фурье, Уолша и для секвентно-упорядоченных функций Уолша и для диадно-упорядоченных функций Уолша.

Пример 3. За исходный сигнал взято двухмерное решетчатое двухградационное изображение размером 8×4 (рис.5), зашумлен-

ное случайными адекватными шумами, описываемыми моделью Маркова. Средняя интенсивность помехи равна 10% от максимальной интенсивности исходного сигнала. На рис. 5 даны значения средне-квадратичных ошибок усечения для преобразований Карунена-Лоеве, Фурье, и Уолша с различным расположением базисных функций - секвентивно-упорядоченных и диадно-упорядоченных.

Рассмотренные примеры демонстрируют существенные преимущества метода главных компонент с использованием преобразования Карунена-Лоеве. Однако широкому применению преобразования Карунена-Лоеве препятствует то, что оно требует большого объема вычислений. Сначала необходимо диагонализировать матрицу большой размерности. Эта матрица оказывается близкой к сингулярной матрице, так как некоторые ее собственные значения очень малы. Процесс диагонализации оказывается трудной вычислительной процедурой. После нахождения собственных векторов, образующих оптимальный ортогональный базис, выполняют само преобразование Карунена-Лоеве, которое требует n^2 операций умножения и сложения.

Правда, существуют способы быстрой диагонализации матрицы ковариации ^{/7/}. Однако, они дают лишь небольшое сокращение вычислительных операций, а соответствующие алгоритмы не являются такими же универсальными, как быстрые алгоритмы преобразования Фурье или Уолша-Адамара.

Диагонализацию матрицы ковариации ведут с помощью метода Якоби, разбивая исходную матрицу ковариации на подматрицы. Число итераций, к которым сводятся вычисления собственных векторов по методу Якоби, уменьшается благодаря выбору более точного нулевого приближения.

Критерием для выбора числа отсчетов или компонент исходного вектора является величина

$$J_n = \frac{1}{2} \frac{\sum_{i=1}^n \{R(x_{2i}, x_{2i}) - R(x_{2i}, x_{2i-1}) - R(x_{2i-1}, x_{2i}) + R(x_{2i-1}, x_{2i-1})\}}{\sum_{i=1}^n [R(x_{2i}, x_{2i}) + R(x_{2i-1}, x_{2i-1})]}$$

где $R(x, x')$ - матрица ковариации. Если $J_n \ll 1$, то для представления данного случайного процесса достаточно взять n отсчетов. Для оценки числа реализаций необходимы данные о значениях статистик процесса. В случае гауссовского процесса границы дисперсий собственных значений и компонент собственных векторов матрицы ковариации, найденных по N реализациям, равны

$$\sigma_{\lambda_i}^2 \leq \frac{2}{N} \lambda_i^2$$
$$\sigma_{w_i}^2 \leq \frac{1}{N} \sum_{\substack{i=1 \\ j \neq i}}^n \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}$$

Преобразование Карунена-Лоеве в многомерном
анализе

При обработке экспериментальных данных, получаемых с больших спектрометров, возникает проблема быстрого считывания и быстрого отбора полезной информации в сжатой форме, а также быстрая обработка полученной информации. В задачах физики частиц высоких энергий могут быть полезными методы обработки и сжатия информации, разработанные в технике связи.

Информация о ядерном событии, регистрируемом большими спектрометрами, обычно содержится в координатах точек, распределенных вдоль трека частицы с определенными интервалами. Для определения параметров частицы ($\rho, \theta, \varphi, E, Z$) используют несколько проекций события, позволяющие восстановить координаты элементов трека в пространстве.

Трек можно представить точкой в некотором многомерном пространстве, а каждое измерение — соответствующей координатой в этом пространстве /8/. Определенный класс треков попадает внутрь некоторой области этого многомерного пространства, и идентификацию треков можно перенести в это пространство. Трудность такого информационного подхода заключается в том, что размерность пространства оказывается очень высокой, а объем памяти и требуемый объем вычислений — огромным. Было бы полезно существенно понизить размерность пространства без потери полезной информации. Для этого можно использовать метод главных компонент или преобразование Карунена-Лоеве.

Разберем пример использования этого метода для распознавания и анализа треков.

Пусть, например, трек определяется пятью точками. На одной проекции будет измерено 10 координат. Для задания трека, между тем, достаточно всего пять параметров. Если найти эти пять значимых параметров, то значения импульса частицы P выражалось бы функцией этих пяти параметров:

$$P = P(\xi_1, \xi_2, \dots, \xi_5)$$

Чтобы найти параметры $\xi_1, \xi_2, \dots, \xi_5$, воспользуемся методом главных компонент. Совершим преобразование Карунена-Лоеве над измеренным вектором

$$\vec{X}(x_1, x_2, \dots, x_5; y_1, y_2, \dots, y_5)$$

с десятью компонентами. Чтобы найти матрицу преобразования \hat{W} , необходимо знать статистические характеристики класса наблюдаемых событий. Их находят либо с помощью метода Монте-Карло, либо аппроксимируют процесс определенными статистическими моделями. Последний метод, позволяющий представить матрицу ковариации в явном виде, заслуживает особого внимания.

После того как найдены собственные векторы матрицы ковариации, их располагают в порядке убывания собственных значений, и те значения, которые оказываются малыми по определенному признаку, отбрасывают. Происходит сжатие информации. Первые пять компонент разложения $\xi_1, \xi_2, \dots, \xi_5$ содержат

искомую информацию в сжатом виде, а компоненты $\{6, \{7, \dots, \{10$ указывают меру сопоставления треков разных типов. Тот факт, что компоненты $\{6, \{7, \dots, \{10$ малы, используют как критерий проверки факта, что вектор \vec{X} принадлежит к данной траектории. Значение критерия тесно связано с ошибками измерения десяти координат трека.

В общем случае трек реальной частицы образует сгущения или кластер в пространстве, размерность которого L меньше размерности первоначального пространства N . Это значит, что остальные информационные признаки $\{L+1, \{L+2, \dots, \{N$ являются малыми величинами. В качестве критерия принадлежности трека к данному кластеру берут величину

$$d_L = \sqrt{\sum_{l=L+1}^N \{l^2}$$

которая должна быть мала по сравнению с эффективной ошибкой измерений.

В работе^{/9/} такому анализу были подвергнуты 1004 трека, вид которых дан на рис.6. Координаты частиц в модулях искровых камер получались путем разыгрывания событий методом Монте-Карло. Траектории группировались в события, содержащие от 2 до 8 частиц. После сортировки всех координат, проводилась идентификация треков. Вычислялись величины $\{6, \{7, \dots, \{10$ для различных комбинаций, от одной до N_m пар координат

точек трека. Совокупность точек приписывалась к одному треку, если $d_5 < (2 + 3) \varepsilon$, где ε - эффективная ошибка измерений. Такая логика идентификации и выбора реальных траекторий осуществлялась в аналоговом процессоре /9/.

На рис.7 представлено распределение кластеров по критерию $d^2 / 10$ для искрового спектрометра R 603 по схеме моделирования, которая была описана выше. Видно, что все кластеры четко распределились на две группы: группу истинных треков и группу ложных треков. Последняя группа характеризуется гипертрофированными значениями величины d^2 .

На рис.8 дан график времени счета в процессе идентификации событий в зависимости от числа треков на кадре. Приведены значения для традиционного метода "грубой силы" и для метода нахождения главных компонент с помощью дискретного преобразования Карунена-Лоеве. Из рис.8 видно, что в типичном случае 6 + 7 треков на события выигрыш во времени счета составляет $10^2 : 1$. Это позволяет сделать вывод о том, что переход к анализу трековых событий в искровых спектрометрах с помощью метода главных компонент является экономически обоснованным.

М. С. Смирнов
10.10.72

Литература

1. H. Andrews, Computer techniques in image processing. Academic Press, 1970.
2. A. Habibi, P. Wintz, Purdue University School on Electronical Engineering, TR-EE, 70-2, March, 1970.
3. P. Wintz, Transform Picture Coding, Proc. IEEE, 60, Nr. 7, 809 - 820, 1972. см. также "Обработка изображений при помощи цифровых вычислительных машин" МИР, 1973, 98-112.
1973, 98 - 112.
4. K. Fukunaga, W.L. Koontz, Information and Control, 16, 85 - 101, 1970.
5. S. Watanabe, in Proc. of the 4 Prague Conference on Information Theory, 1965, см. также "Автоматический анализ сложных изображений", МИР, 1969.
1969.
6. H. Hotelling, J. Educ, Psychology, 24, 417 - 441, 498 - 520, 1933.
7. W. Pratt, J. Kane, H. Andrews, Proc. IEEE, 57, 58, 1969.
8. H. Wind, "Function Parametrization", in CERN 72 - 21, Proc. 1972 CERN Computing and Data Processing School.
9. M. Hansroul, C. Verkurk, "Point and track finding processors for multiwire chambers". CERN, PP/73/4, August 1973.
10. M. Hansroul, P. Townsend, P. Zanella, CERN, PP/73/51, October 1973.
11. K. Fukunaga, "Introduction to statistical pattern recognition", Academic Press, 1972.
12. С.О. Кожевникова, Л.М. Сороко. Об алгоритмах считывания прямолинейных треков с помощью масок Уолша. Сообщение ОИАИ, РЮ-7757, 1974.

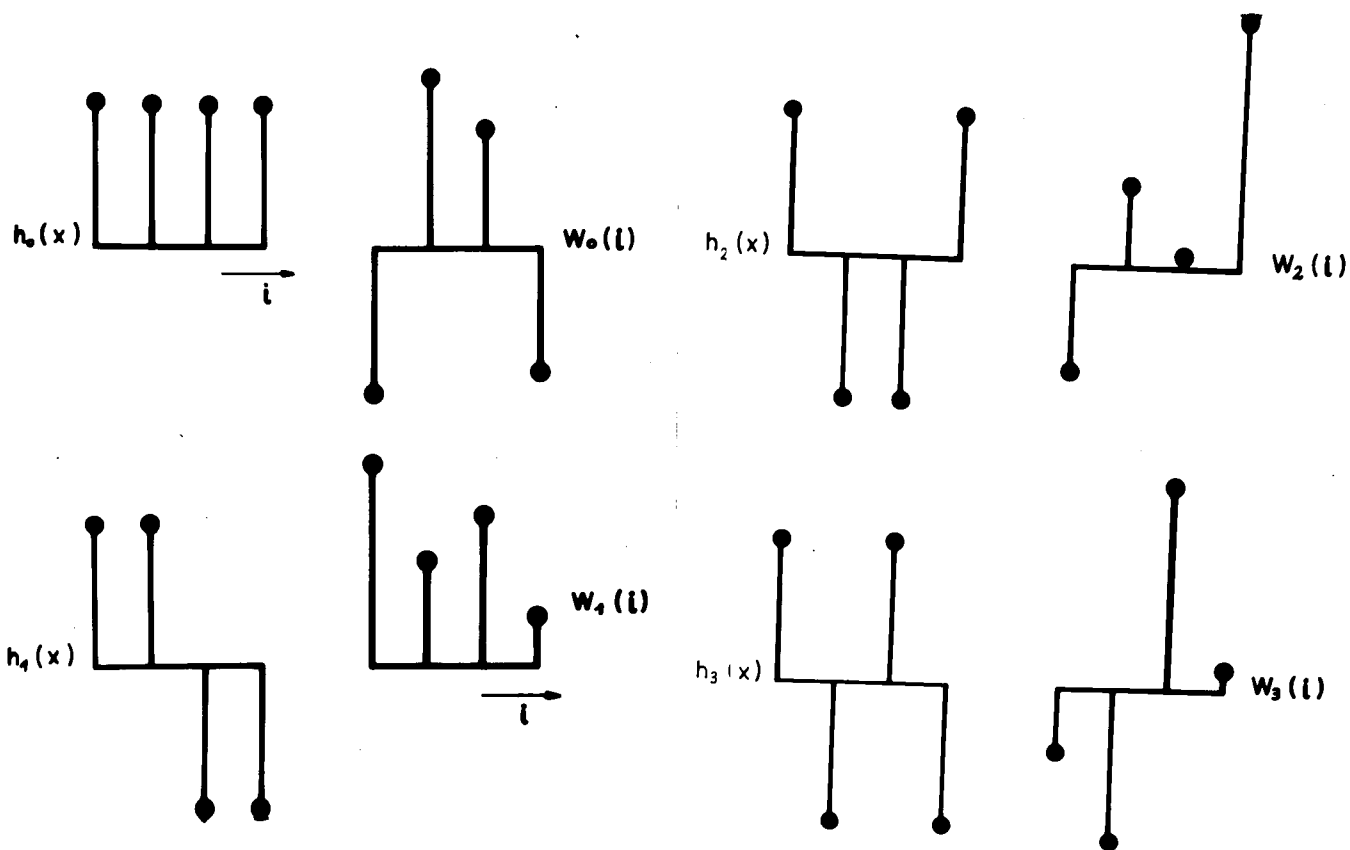


Рис. I Пример ортогональных базисов для $N = 4$.

$w_0(i), w_1(i), w_2(i), w_3(i)$,
 $(1 \leq i \leq 4)$ - собственные векторы корреляционной матрицы для последовательности $(1,00; -1,20; -0,78; 0,98)$. Они образуют базис для преобразования Карунена-Лоэва. $h_0(i), h_1(i), h_2(i), h_3(i)$,
 $(1 \leq i \leq 4)$ - базисные функции для преобразования Уолша.

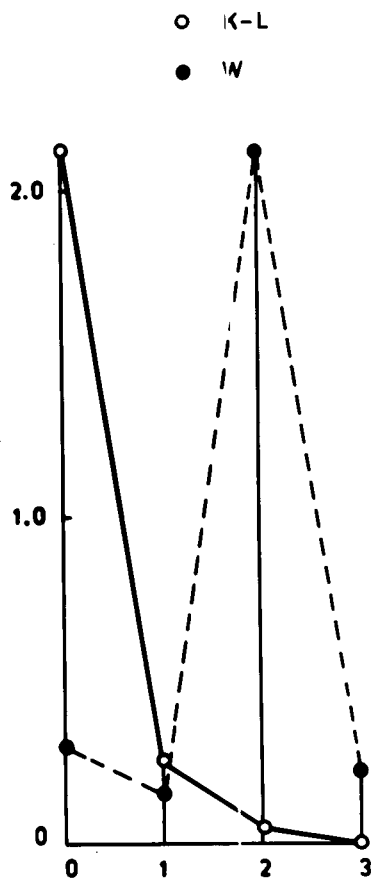


Рис.2. Пример разложения некоторой последовательности по собственным векторам корреляционной матрицы. Приведены значения главных компонент /сплошная кривая/ и компонент Уолш-образа /пунктирная кривая/.

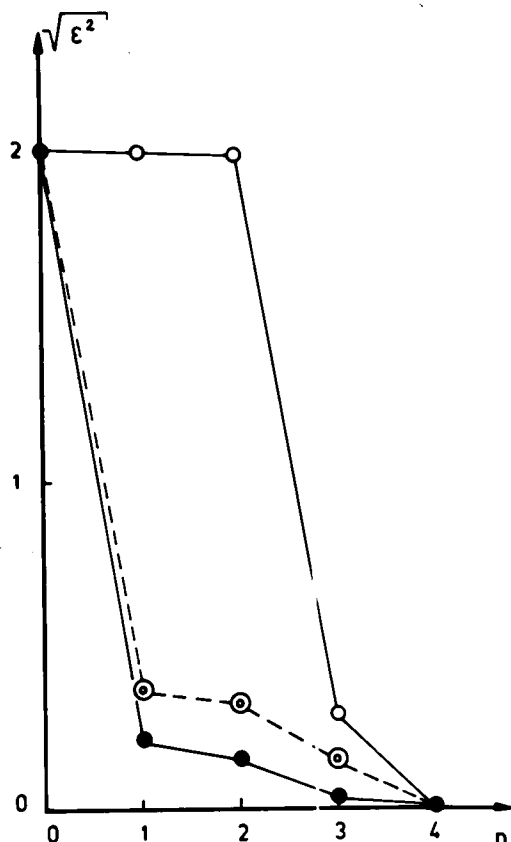


Рис.3. Среднеквадратичные ошибки восстановления исходной последовательности для различного числа членов разложения /информационных признаков/.
 Кривая 1 — для разложения Карунена-Лоева, кривая 2 — для разложения по секвентивно-упорядоченным функциям Уолша, кривая 3 — для разложения по системе функций Уолша, расположенным в порядке монотонного уменьшения компонент образа /упорядоченность по Карунену-Лоеву/.

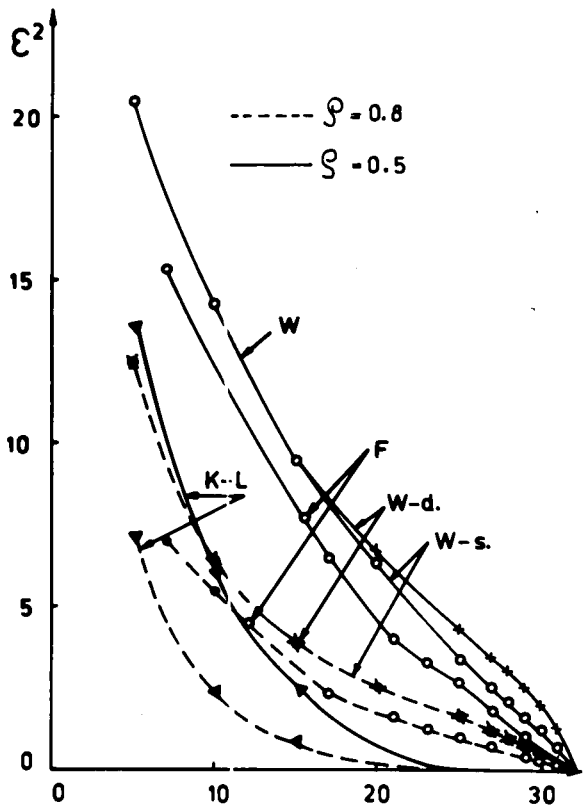


Рис.4. Зависимость среднеквадратичной ошибки восстановления \mathcal{E}^2 исходной последовательности от числа учтенных информационных признаков для случайного марковского процесса с коэффициентами корреляции $\rho = 0,8$ /пунктирные кривые/ и $\rho = 0,5$ /сплошные кривые/. "K-L" - преобразование Карунена-Лоева, "F" - преобразование Фурье, "Walsh-d" - для разложения по диадно-упорядоченным функциям Уолша, "Walsh-s" - для разложения по секвентивно-упорядоченным функциям Уолша.

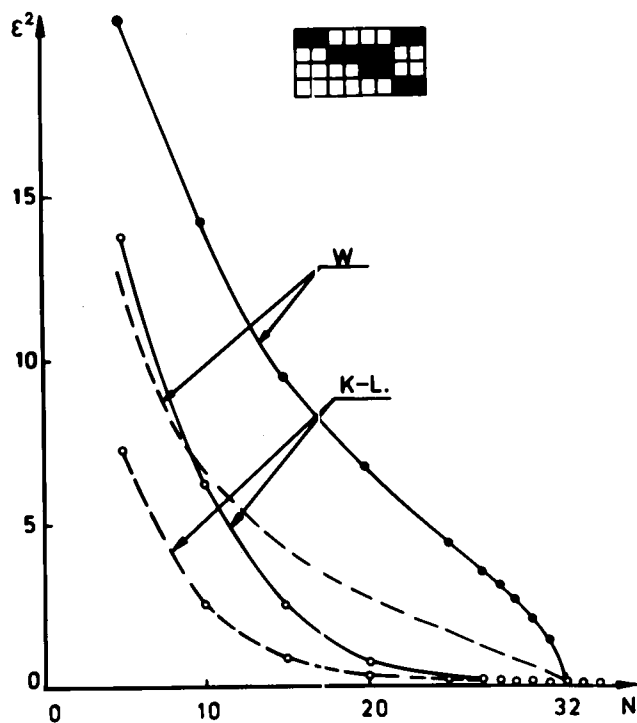


Рис. 5. Зависимость среднеквадратичной ошибки восстановления ϵ^2 двухградационного решетчатого изображения, помещенного в правом верхнем углу, от числа учтенных при восстановлении информационных признаков. Изображение искажено случайным аддитивным шумом, интенсивность которого составляет 10% от интенсивности изображения. Шум описывается случайным марковским процессом с коэффициентом корреляции $\rho = 0,8$ /пунктирные кривые/ и $\rho = 0,5$ /сплошные кривые/. "К-Л" - преобразование Карунена-Лоева, "Walsh-d" - разложение по диадно-упорядоченным функциям Уолша.

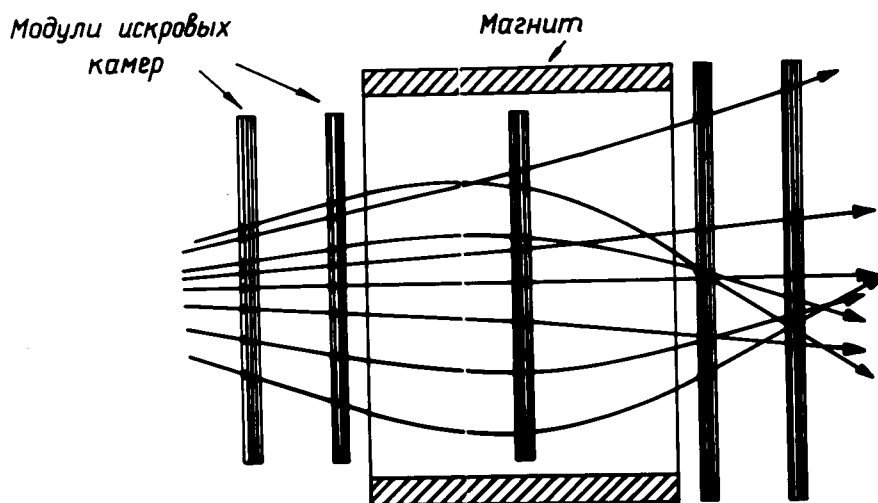


Рис.6. Схема искрового спектрометра R 603 и вид треков, которые были рассчитаны в работе /9/ с использованием моделирования Монте-Карло.

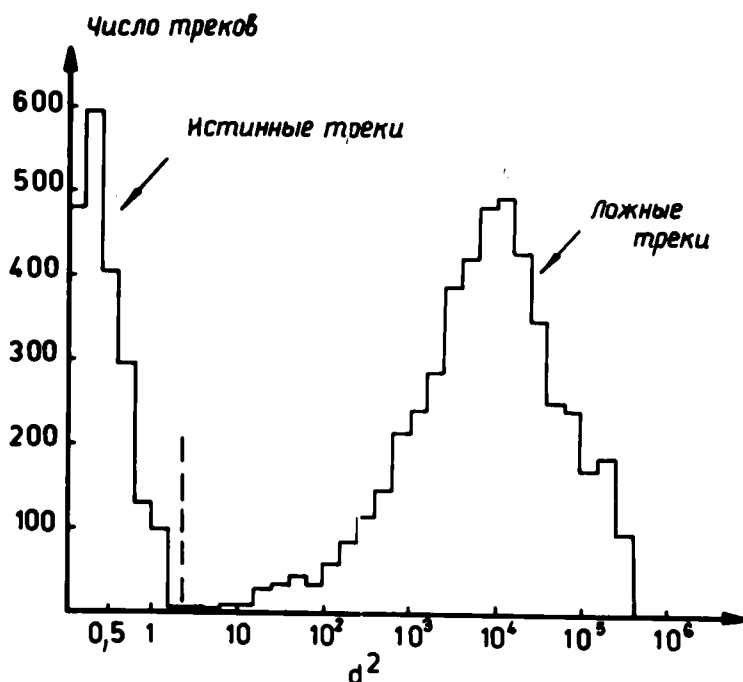


Рис.7. Рисунок, поясняющий отбор треков по критерию $d^2 = \sum_{i=6}^{10} \xi_i^2$, сделанный в работе /9/. Четко видно разделение событий на истинные и ложные треки.

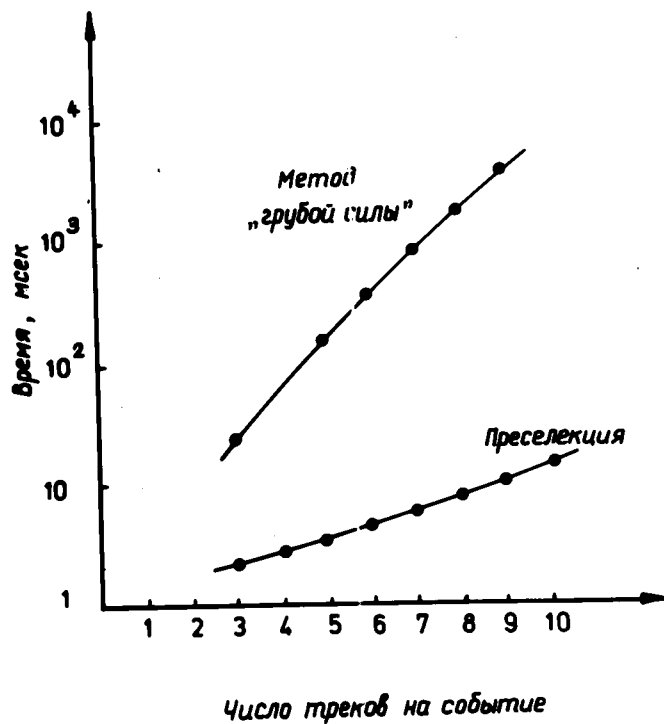


Рис.8. Время обработки, необходимое для идентификации треков многолучевого события с числом треков на событие от 3 до 8. Верхний график — для традиционного метода "грубой силы". Нижний график — для метода главных компонент с использованием ЭВМ СДС 7600 /9/.