

**СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ  
ДУБНА**

**P10-86-773**

**Г.А.Ососков, Н.И.Чернов**

**СТАТИСТИЧЕСКИЙ ПОДХОД  
В ЗАДАЧЕ РАСПОЗНАВАНИЯ ОБРАЗОВ  
В ФИЗИКЕ ВЫСОКИХ ЭНЕРГИЙ  
Реконструкция треков по сжатым данным**

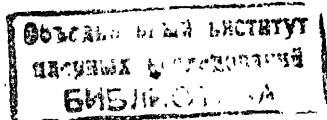
**1986**

1. Настоящая работа примыкает к исследованию<sup>/1/</sup>, обобщая предложенный там статистический подход на случай предварительного сжатия данных сканирования в линейные трек-элементы (ЛТЭ). Основой подхода являлось применение принципа максимума правдоподобия для организации множества точек изображения. Благодаря включению в функцию правдоподобия шумовых точек, пропусков между точками одного и того же трека и использованию локально-циркулярной аппроксимации треков, были построены алгоритмы распознавания трековых изображений, эффективно работающие как при наличии целеуказаний (алгоритм АПК1), так и при их отсутствии (АПК2).

Развитие АПК2 на случай предварительного сжатия данных в ЛТЭ потребовало создания новой вероятностной модели изображения и, соответственно, иного определения полной функции правдоподобия. Наиболее значительные усилия были приложены для ускорения работы алгоритма распознавания, ориентированного на работу без каких-либо целеуказаний. В результате формализации процесса перестройки кластеров, состоящих из ЛТЭ и нескольких численных приемов, резко сокращающих объем перебора данных и ускоряющих вычисление функции правдоподобия, была достигнута скорость обработки в 3-6 с на реальных кадрах, содержащих до 300 произвольно расположенных ЛТЭ случайной длины.

2. Для построения модели набора данных, состоящего из ЛТЭ, необходимо обобщить предложенную в<sup>/1/</sup> точечную модель данных сканирования с учетом способа сжатия их в линейные трек-элементы. Предварительное сжатие трековой информации в ЛТЭ проводится с помощью специальной аппаратуры и допускает несколько вариантов сжатия<sup>/2-5/</sup>. Все ЛТЭ могут иметь одинаковую длину (как в<sup>/2-4/</sup>) или она может зависеть от кривизны треков (как в<sup>/5/</sup>), но во всех алгоритмах сжатия ЛТЭ объединяют отсчеты, лежащие внутри полосы шириной  $2\delta$  вокруг ЛТЭ с  $\delta$ , определяемой заданной точностью алгоритма сжатия.

Для построения общей модели, в которой наблюдаемыми данными будут сами ЛТЭ безотносительно от исходных данных и алгоритма сжатия, нам потребуется сделать три вполне естественных предложения:



( i ) ЛТЭ задается парой своих концов и числом реальных отсчетов, которые он объединяет,

( ii ) если ЛТЭ объединяет отсчеты, принадлежащие только одному треку, то в заданный ЛТЭ попадают все отсчеты этого трека на участке между концами ЛТЭ,

( iii ), если отсчеты, объединяющие ЛТЭ, принадлежат более, чем одному треку или включают шумовые отсчеты, то все отсчеты этого ЛТЭ считаются шумовыми ("шумовой" ЛТЭ).

Выполнение этих условий должно обеспечиваться алгоритмом сжатия. Условие ( iii ) фактически следует из замечания к формуле (4) в работе /1/.

Логарифмическая функция правдоподобия для выборки из полного набора ЛТЭ записывается в виде

$$L(J_1, \dots, J_k) = L(J_{i_1}^{(1)}, \dots, J_{i_k}^{(1)}) + \dots + L(J_{i_1}^{(m)}, \dots, J_{i_{k_m}}^{(m)}) + L(J_{i_1}^{(0)}, \dots, J_{i_{k_0}}^{(0)}) \quad (I)$$

где  $J_1, \dots, J_k$  - набор всех ЛТЭ,  $i_1, \dots, i_{k_m}$  - номера ЛТЭ, объединяющих отсчеты  $m$ -го трека ( $1 \leq m \leq M$ ),  $i_1, \dots, i_{k_0}$  - номера шумовых ЛТЭ, а  $L$  обозначает отрицательный логарифм плотности распределения.

**Замечание.** В силу ( ii ) трек-элементы одного трека не перекрывают друг друга ( не пересекают одну и ту же скан-линию), и мы можем занумеровать их по порядку слева направо.

В соответствии с точечной моделью изображения, предложенной в /1/, последний член в (I), соответствующий шумовым ЛТЭ, может быть представлен в виде:

$$L(J_{i_1}, \dots, J_{i_{k_0}}) = -\ln n^{|J_{i_1}| + \dots + |J_{i_{k_0}}|} = -N_0 \ln n \quad (2)$$

где  $|J_{i_1}|$  - число реальных отсчетов, объединенных в  $i$ -м ЛТЭ, а  $n$  - как и в /1/, постоянная плотность равномерного распределения одного шумового отсчета.

Для определения  $m$  первых членов в (I) рассмотрим ЛТЭ, объединяющий точки одного из треков (рис. I). Пусть этот ЛТЭ пересекает  $s$  скан-линий и содержит  $r$  отсчетов ( $r \leq s$ ). По способу получения ЛТЭ в соответствии с моделью работы /1/ вероятность  $\chi_i$  того, что точки трека, лежащие на  $i$ -й скан-линии ( $2 \leq i \leq s-1$ ), будут принадлежать этому ЛТЭ, т.е. попадут в полосу ширины  $2\delta$ , равна

$$\chi_i = \Phi(a_i + \delta) - \Phi(a_i - \delta) \quad (3)$$

где  $a_i$  - расстояние от трека до оси ЛТЭ вдоль  $i$ -й скан-линии,  $\Phi(x)$  - нормальная функция распределения  $N(0, \sigma^2)$  (величина  $\sigma$  предполагается известной согласно модели /1/).

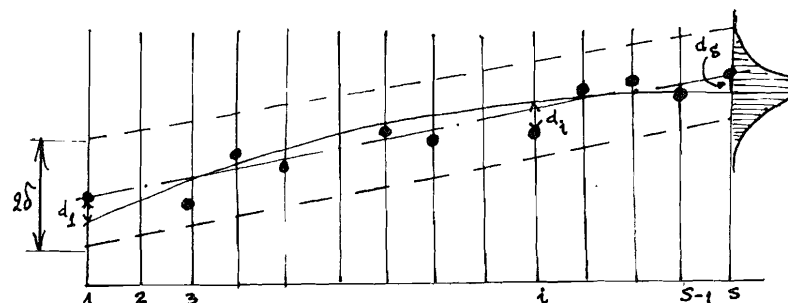


Рис. I.

Обозначая  $p$  вероятность получить отсчет от трека на скан-линии ( $p$  считается известным или подлежит подбору), получаем плотность вероятности для отсчетов данного ЛТЭ:

$$f(J) = \frac{1}{2\pi\sigma^2} e^{-\frac{d_1^2 + d_2^2}{2\sigma^2}} \sum_{i_1, \dots, i_{r-2}}^* \chi_{i_1} \dots \chi_{i_{r-2}} (1-p)^{s-r} p^{r-2} \quad (4)$$

где  $\sum^*$  - сумма по всем наборам ( $i_1, \dots, i_{r-2}$ ) из чисел (2, 3, ..., s-1);  $d_1$  и  $d_2$  - расстояния от концов ЛТЭ до оси трека.

Если обозначить  $\bar{\chi} = (\chi_2 + \dots + \chi_{s-1}) / (s-2)$ , то приближенное выражение для (4) будет иметь вид

$$f(J) \approx \frac{1}{2\pi\sigma^2} e^{-\frac{d_1^2 + d_2^2}{2\sigma^2}} c_{s-2}^{r-2} \bar{\chi}^{r-2} (1-p)^{s-r} p^{r-2} \quad (5)$$

Обозначим  $i$  число скан-линий в промежутке между  $i$ -м и ( $i+1$ )-м ЛТЭ в данном треке ( $1 \leq i \leq k_m - 1$ ), а  $l_0$  и  $l_{k_m}$  - число скан-линий от концов, крайних на треке (соответственно, слева и справа) ЛТЭ до реальных концов трека. Поскольку  $l_0$  и  $l_{k_m+1}$  неизвестны, мы, как и в /1/, положим  $l_0 + l_{k_m} = l$  - константа, подбираемая экспериментально. Тогда первые  $M$  членов в (I), согласно модели /1/, вычисляются по формуле

$$L(J_{i_1}^{(m)}, \dots, J_{i_{k_m}}^{(m)}) = -\sum_{j=1}^{k_m} \ln f(J_{i_j}^{(m)}) - \sum_{j=1}^{k_m-1} l_j \ln(1-p) - l \ln(1-p) \quad (6)$$

в которой слагаемые первой суммы соответствуют трек-элементам, второй -  $(k_m - 1)$  промежуткам между ЛТЭ, а последний член, как и в работе<sup>/1/</sup>, выполняет роль штрафа.

3. Формулы (1-6), в принципе, завершают построение вероятностной модели трековых данных, сжатых в ЛТЭ. Однако для получения отсюда вычислительного алгоритма минимизации величины (1) необходимо решить целый ряд вопросов, связанных с аппроксимацией осевой линии трека, организацией вычислений и уточнением понятия кластерной перестройки на случай ЛТЭ.

Аппроксимация линии трека, нужная для вычислений величин (5), выполняется локально с помощью МНК-подгонки окружности по нескольким ЛТЭ, соседних с данным. Их число зависит от минимальной кривизны трека, дисторсий сканатора, таких параметров, как  $\delta$ , длина слайса и т.д. Рассмотрение изображений реальных событий стримерной камеры РИСК, сжатых с помощью алгоритма<sup>/5/</sup>, например, показало, что обычно достаточно двух-трех ЛТЭ (т.е. подгонки окружности по четырем (шести) их конечным точкам). В дальнейшем для определенности мы будем говорить о тройках ЛТЭ (хотя все выводы справедливы и для подгонки по парам). Если в треке ровно два ЛТЭ, то он фитируется по четырем конечным точкам, если только один ЛТЭ, то линия трека определяется осью этого ЛТЭ.

Приемом, позволяющим значительно ускорить вычисление  $\bar{\delta}$  при фиксированных  $\delta$  и  $\sigma$ , известных заранее, является предварительная табуляция функции (3), зависящей только от  $a_i$ . Вклад каждого трек-элемента  $J$  в плотность (5) зависит только от него самого и от двух соседних с ним (иногда одного). Поэтому можно вычислить этот вклад заранее в виде таблицы, позволяющей для каждого  $J$  найти величину  $\ln f(J)$ , вычисленную по окружности, подогнанной по тройке ЛТЭ, включающей  $J$ . В соответствии с местом, которое занимает  $J$  в тройке, для нее должны быть вычислены и занесены в таблицу три величины

$$R_i(J_1, J_2, J_3) = -\ln f(J_i), \quad i=1,2,3. \quad (7)$$

Аналогично вычисляются величины

$$R_i(J_1, J_2) = -\ln f(J_i), \quad i=1,2 \quad (8)$$

при подгонке окружности по конечным точкам пары ЛТЭ  $J_1, J_2$ .

Очевидно, что в таблицу должны быть внесены не все тройки, а только "допустимые", т.е. те, которые могут войти в один и тот же трек в непосредственном соседстве друг с другом.

Для быстрого отсева неподходящих можно предложить ряд пороговых критериев. Работа с вышеупомянутыми сжатыми данными показала, что после такого отсева из  $k$  ЛТЭ остается от  $k/2$  до  $3k/2$  пар и троек. Составление таблицы величин (7) и (8) требует 1-2 с работы ЭВМ CDC-6500 при  $k \leq 300$ . Такая таблица позволяет производить пересчет функции (1) после каждой перестройки кластеров с помощью нескольких операций сложения и вычитания, что свидетельствует о богатых скоростных возможностях алгоритма.

4. Общая схема кластерной перестройки в случае ЛТЭ полностью аналогична алгоритму АПК2 работы<sup>/1/</sup>. Сходимость алгоритма гарантирована в силу монотонного убывания величины (1) после каждой перестройки.

Понятие "перестройка", однако, требует уточнения, так как и при табличном пересчете функции (1) алгоритм не может за разумное время просматривать и оценивать качество всех возможных перестроек. Для выхода из этого положения необходимо ограничиться только такими типами перестроек, которые фактически исчерпывали все реально встречавшиеся случаи при анализе сжатых данных с установки РИСК. Таких перестроек оказалось четыре:

- А. Слияние двух кластеров.
- Б. Расщепление кластера на две части.
- В. Отщепление части кластера и присоединение ее к другому кластеру.

Г. Отщепление по одной части от двух разных кластеров и "обмен" этими частями между ними.

Пробные подсчеты на реальных данных со стримерной камеры установки РИСК показали, однако, что несмотря на быстроту вычислений с помощью таблицы величин (7) и (8), общее число вариантов при перестройке по правилам А-Г остается все же весьма значительным. Например, при  $k=100$  один шаг алгоритма перестройки, приводящий к уменьшению (1), может занимать до 2 с работы ЭВМ CDC-6500. В этой ситуации эффективность алгоритма зависит исключительно от удачной начальной кластеризации.

5. Используя наше знание специфики материала, подлежащего кластеризации, сформулируем алгоритм начальной кластеризации (АНК), выделив его в самостоятельный этап.

Предлагаемые в литературе способы начальной кластеризации основаны на общепринятом в кластерном анализе понятии "функции расстояния". Существует много вариантов функции расстояния (близости) между ЛТЭ (см., например, обзор в<sup>/6/</sup>), но в нашем случае удобнее всего построить функции расстояния на основе таблицы троек или пар, т.е. величин (7), (8) как функций вида

$$d_3(J_1, J_2, J_3) = \frac{1}{s_1 + s_2 + s_3 + \ell_1 + \ell_2} \sum_{i=1}^3 R_i(J_1, J_2, J_3) - (\ell_1 + \ell_2) \ln(1-p) \quad (9)$$

и

$$d_2(J_1, J_2) = \frac{1}{s_1 + s_2 + \ell_1} \sum_{i=1}^2 R_i(J_1, J_2) - \ell_1 \ln(1-p) \quad (10)$$

где  $s_i$  - число скан-линий, пересекающих  $J_i$ ; а  $\ell_i$  - число скан-линий в промежутке между  $J_i$  и  $J_{i+1}$ . В случае отсутствия тройки или пары ЛТЭ в таблице полагаем  $d_3 = \infty$  (соответственно  $d_2 = \infty$ ).

Алгоритм начальной кластеризации, использующий функцию  $d_3$ , имеет следующий вид:

Шаг 0. Все ЛТЭ считаем свободными - не связанными в кластеры.

Шаг 1. Упорядочиваем множество троек ЛТЭ по возрастанию функции  $d_3$ .

Шаг 2. Выбираем  $i$ -ю тройку ЛТЭ ( $i = 1, 2, \dots$ ).

Если она состоит только из свободных ЛТЭ, то создаем из нее новый кластер. Если она содержит ЛТЭ из разных кластеров, то пытаемся объединить их в один, проверяя выполнение условия (ii) (см. также замечание после формулы (I)), и чтобы значения  $d_3$  для всех соседних троек ЛТЭ в объединенном кластере не превосходили  $d_3$  данной  $i$ -ой тройки ЛТЭ. Если  $i$ -ая тройка содержит один свободный ЛТЭ, а оставшиеся входят в один из созданных кластеров, то пытаемся включить в него этот свободный ЛТЭ, с помощью проверки тех же двух условий, что и при слиянии кластеров.

Шаг 3. Увеличиваем  $i$  на 1, и если для  $i$ -й тройки  $d_3 < \infty$ , то возвращаемся к шагу 2. В противном случае - заканчиваем работу.

После завершения начальной кластеризации все ЛТЭ, оставшиеся свободными, объединяются в один шумовой кластер.

Аналогично определяется алгоритм начальной кластеризации, использующий пары ЛТЭ и функцию  $d_2$ .

Следует заметить, что применение функции расстояния  $d_3$ , вычисляемой по тройкам элементов, не является характерным для кластерной перестройки, где обычным является расстояние  $d_2$  [6,7], позволяющее использовать хорошо разработанные алгоритмы ближайшего соседа или наименьшего связующего дерева (см, например, [8-10]).

6. Проверка работы АНК на нескольких десятках кадров, полученных при сжатии реальных данных с мониторной ТВ-системы [11] (в каждом кадре по 200-300 ЛТЭ и от 5 до 15 треков), выявила сравнительно невысокий процент ошибок (менее 20% треков были опознаны иначе, чем человеком, что неплохо для начальной кластеризации). Фортранная

программа АНК состояла из ~600 операторов, занимала в памяти ЭВМ ~20К слов и обрабатывала кадр в среднем за 3-5 с.

Вариант программы, реализующей АНК на основе  $d_2$ , оказался на 16% короче и быстрее (4 с/кадр), но делал несколько больший процент ошибок.

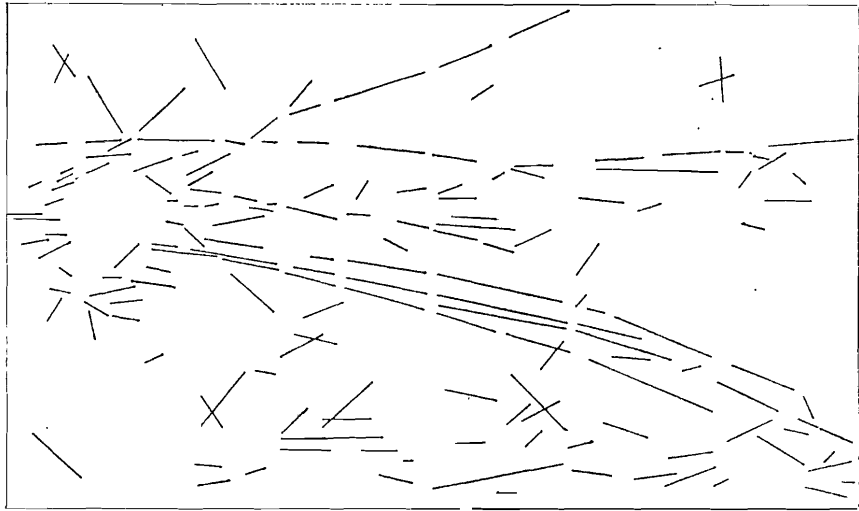
Применение к полученной начальной кластеризации метода кластерной перестройки для минимизации (I) с помощью таблиц троек на статистике из двух десятков 5-10-лучевых событий показало практически полное совпадение с распознаванием, выполненным человеком (около 5% расхождений объяснялись ошибками сжатия, главным образом, из-за нарушения предположения (ii)).

На рис.2 приведены типичные изображения после предварительного сжатия данных сканирования в ЛТЭ. Результат реконструкции треков показан на рис.3. Визуальная проверка подтверждает правильность работы алгоритма. После начальной кластеризации с использованием функции  $d_3$  метод кластерной перестройки завершал реконструкцию за 3-5 итераций. Время работы программы - 3-6 с на ЭВМ CDC-6500. Объем - около 900 операторов фортрана.

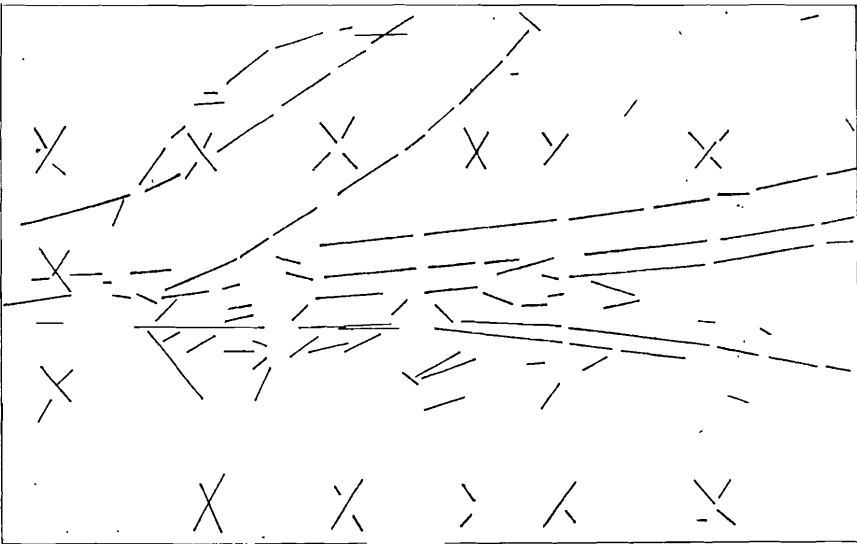
Созданные алгоритмы для применения их в массовой обработке, конечно, нуждаются в доводке на большей статистике. Следует также отметить высокую чувствительность алгоритма к нарушениям предположений (ii), (iii) - включение в ЛТЭ посторонней концевой точки или пропуск нескольких точек могут привести к неправильному восстановлению всего трека. Поэтому применение предложенного алгоритма кластерной перестройки требует высокой надежности работы алгоритмов, заложенных в аппаратуру сжатия данных в ЛТЭ. Исследования, проведенные на имеющемся материале, позволяют констатировать, что при выполнении этого условия алгоритм позволяет восстанавливать треки на достаточно сложных кадрах, содержащих до 300 произвольно расположенных ЛТЭ, случайной длины, из которых более половины - шумовых.

В заключение сделаем замечание относительно применимости предлагаемого алгоритма к данным сжатия с помощью спецпроцессоров типа SHP [2-4], которые отличаются от рассматриваемых в настоящей работе тем, что число ЛТЭ в одном изображении на порядок больше, но сами ЛТЭ расположены упорядоченно в узких слайсах, что определяет не произвольную, а равную длину всех ЛТЭ в 5-10 скан-линий.

Очевидно, что хотя такой тип сжатых данных укладывается в вероятностную модель, предлагаемую в разделе 2, и алгоритмы принципиально применимы, однако без существенных изменений для учета упорядоченности ЛТЭ и сокращения объема таблиц его непосредственное использование вряд ли рационально. Один из наиболее простых путей - двухэтапное применение алгоритма, сначала к укрупненным слайсам, объединяющим 5-10 ЛТЭ, а потом повторно для сшивки полученных дуг окружностей.

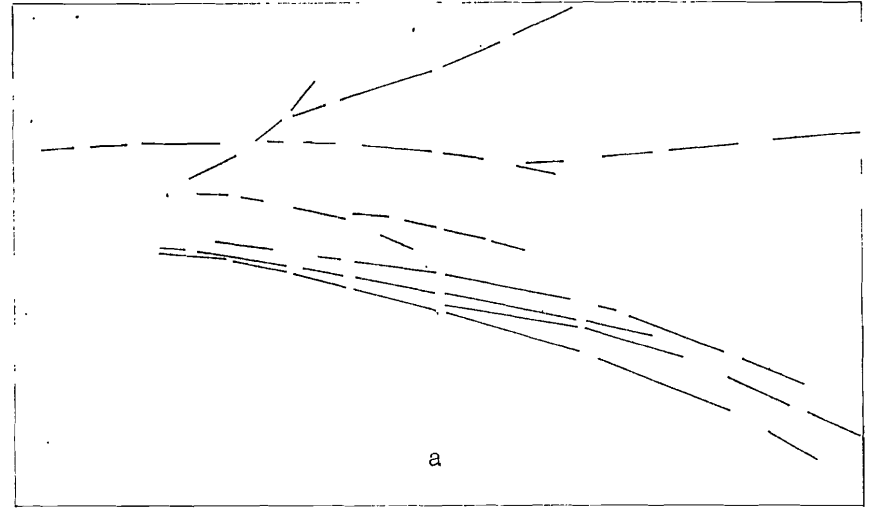


а

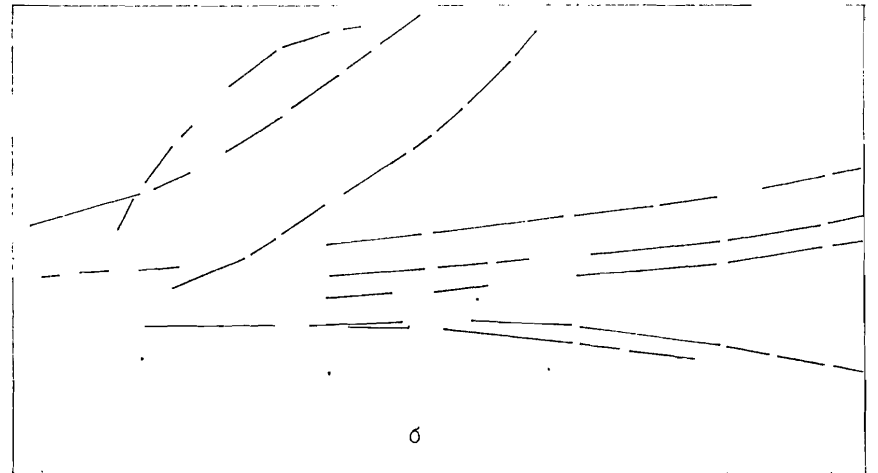


б

Рис. 2.



а



б

Рис. 3.

## НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги,  
если они не были заказаны ранее.

- Литература
- I. Иоселиани Ц.И. и др. ОИЯИ, Р10-86-666, Дубна, 1986.
  2. Flavell A.J. et al. В кн.: Труды международного симпозиума по обработке данных пузырьковых камер. ОИЯИ, Д10-6142, Дубна, 1971, с.248-272.
  3. Крупнов В.Е., Федотов О.П. В сб.: Материалы II Всесоюзного семинара по обработке физической информации. Изд-во ЕрФИ, Ереван, 1978, с.314-321.
  4. Bacilieri P et al. Nucl. Instr. Meth., 135, 1976, p.427-433.
  5. Крушинский Д. и др. В кн.: Труды У международного совещания по проблемам математического моделирования, программирования и математическим методам решения физических задач. ОИЯИ, Д10, II-84-818, Дубна, 1984, с.343-345.
  6. Дюран Б., Оделл П., Кластерный анализ. Статистика, М., 1977.
  7. Классификация и кластер (под ред. Д.Ван Рейзена), Мир, М., 1980.
  8. Zahn C. T. IEEE Trans. on Comp., C-20, 1971, p.68-86.
  9. Ососков Г.А., Пахомов В.Л. В кн.: Совещание по программированию и математическим методам решения физических задач. ОИЯИ, Д10, II-11264, Дубна, 1978, с.288.
  10. Ососков Г.А. и др. В кн.: Труды У международного совещания по проблемам математического моделирования, программирования и математическим методам решения физических задач. ОИЯИ, Д10, II-84-818, Дубна, 1984, с.340-342.
  11. Бечер Д. и др. В кн.: Материалы II Всесоюзного семинара по обработке физической информации. Изд. ЕрФИ, Ереван, 1978, с.428-434.

Д2-82-568	Труды совещания по исследованиям в области релятивистской ядерной физики. Дубна, 1982.	1 р. 75 к.
Д9-82-664	Труды совещания по коллективным методам ускорения. Дубна, 1982.	3 р. 30 к.
Д3,4-82-704	Труды IV Международной школы по нейтронной физике. Дубна, 1982.	5 р. 00 к.
Д11-83-511	Труды совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике. Дубна, 1982.	2 р. 50 к.
Д7-83-644	Труды Международной школы-семинара по физике тяжёлых ионов. Алушта, 1983.	6 р. 55 к.
Д2,13-83-689	Труды рабочего совещания по проблемам излучения и детектирования гравитационных волн. Дубна, 1983.	2 р. 00 к.
Д13-84-63	Труды XI Международного симпозиума по ядерной электронике. Братислава, Чехословакия, 1983.	4 р. 50 к.
Д2-84-366	Труды 7 Международного совещания по проблемам квантовой теории поля. Алушта, 1984.	4 р. 30 к.
Д1,2-84-599	Труды VII Международного семинара по проблемам физики высоких энергий. Дубна, 1984.	5 р. 50 к.
Д17-84-850	Труды III Международного симпозиума по избранным проблемам статистической механики. Дубна, 1984. /2 тома/	7 р. 75 к.
Д10,11-84-818	Труды V Международного совещания по проблемам математического моделирования, программированию и математическим методам решения физических задач. Дубна, 1983	3 р. 50 к.
	Труды IX Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1984 /2 тома/	13 р. 50 к.
Д4-85-851	Труды Международной школы по структуре ядра, Алушта, 1985.	3 р. 75 к.
Д11-85-791	Труды Международного совещания по аналитическим вычислениям на ЭВМ и их применению в теоретической физике. Дубна, 1985.	4 р.
Д13-85-793	Труды XII Международного симпозиума по ядерной электронике. Дубна 1985.	4 р. 80 к.

Рукопись поступила в издательский отдел  
1 декабря 1986 года.

Заказы на упомянутые книги могут быть направлены по адресу:  
101000 Москва, Главпочтамт, п/я 79  
Издательский отдел Объединённого института ядерных исследований

**ТЕМАТИЧЕСКИЕ КАТЕГОРИИ ПУБЛИКАЦИЙ  
ОБЪЕДИНЕННОГО ИНСТИТУТА ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ**

Индекс	Тематика
1.	Экспериментальная физика высоких энергий
2.	Теоретическая физика высоких энергий
3.	Экспериментальная нейтронная физика
4.	Теоретическая физика низких энергий
5.	Математика
6.	Ядерная спектроскопия и радиохимия
7.	Физика тяжелых ионов
8.	Криогеника
9.	Ускорители
10.	Автоматизация обработки экспериментальных данных
11.	Вычислительная математика и техника
12.	Химия
13.	Техника физического эксперимента
14.	Исследования твердых тел и жидкостей ядерными методами
15.	Экспериментальная физика ядерных реакций при низких энергиях
16.	Дозиметрия и физика защиты
17.	Теория конденсированного состояния
18.	Использование результатов и методов фундаментальных физических исследований в смежных областях науки и техники
19.	Биофизика

Осоков Г.А., Чернов Н.И.  
Статистический подход в задаче распознавания образов  
в физике высоких энергий. Реконструкция треков по сжатым  
данным

P10-86-773

Дается обобщение работы /1/ на случай, когда данные сканирования изображения события в трековой камере предварительно сжаты в виде линейных трекоэлементов /ЛТЭ/. Задача состоит в разбиении множества ЛТЭ на группы - кластеры так, чтобы первые "M" кластеров соответствовали трекам, имеющимся на изображении /"M" может быть заранее неизвестно/, а остальные ЛТЭ остались в "шумовом" кластере. На основе вероятностной модели такого изображения построена полная функция правдоподобия, позволяющая организовать процесс распознавания треков методом динамической кластерной перестройки, направленной на максимизацию функции правдоподобия. Основное внимание при алгоритмической реализации направлено на ускорение вычислений. В результате формализации процесса перестройки и численных приемов, резко сокращающих объем перебора данных и ускоряющих вычисление функции правдоподобия, была достигнута скорость обработки информации 3-6 с на реальных кадрах, содержащих до 300 произвольно расположенных ЛТЭ случайной длины.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ. . Сообщение Объединенного института ядерных исследований. Дубна 1986

Перевод О.С.Виноградовой

Ososkov G.A., Chernov N.I.  
Statistical Approach in a Problem of Pattern Recognition  
in High Energy Physics. Track Reconstruction for Compressed  
Data

P10-86-773

The work generalizes the results of /1/ to the case of data previously compressed into linear track elements (LTE). The set of LTE is to be clustered into M clusters corresponding to the tracks on the image (M is possibly unknown) leaving the rest of LTE in the "noise" cluster. A statistical model of compressed data is suggested, which makes it possible to organize the recognition process by means of the dynamic clustering method. intended to maximize the likelihood function. The main goal of the algorithm implementation was to accelerate calculations. The cluster reconstructing process was formalized and some tricks were applied to reduce considerably the data sorting and to speed the likelihood function calculations. As a result, the processing speed of 3-6 sec/frame has been obtained for real frame containing up to 300 arbitrary situated LTE having arbitrary lengths.

The investigation has been performed at the Laboratory of Computing Technique and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1986