

21/VI-71

Ц8406

B-689

СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

2086/2-71

P 10 - 5762 e



5762

Н.Г. Волченков, Б.М. Головин,
А.Б. Преображенский

ЛАБОРАТОРИЯ ЯДЕРНЫХ ПРОБЛЕМ

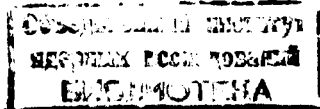
ОБ АВТОМАТИЧЕСКОМ
АНАЛИЗЕ ИНФОРМАЦИИ,
СНИМАЕМОЙ С ИСКРОВЫХ КАМЕР

1971

P 10 - 5762

**Н.Г. Волченков, Б.М. Головин,
А.Б. Преображенский**

**ОБ АВТОМАТИЧЕСКОМ
АНАЛИЗЕ ИНФОРМАЦИИ,
СНИМАЕМОЙ С ИСКРОВЫХ КАМЕР**



В настоящее время в самых различных областях науки и техники разрабатываются и находят применение методы математической лингвистики. В частности, многие концепции этого направления нашли свое отражение в решении проблемы распознавания образов, носящих ярко выраженный структурный характер/1/.

В начале 60-х годов американскими специалистами была высказана мысль о возможности применения лингвистических методов для анализа трековых изображений, в частности фотографий, получаемых при работе с пузырьковыми камерами/1/.

В настоящей работе рассматривается задача, связанная с предварительным анализом событий в проволочной искровой камере. Такой анализ сводится в основном к следующему: для указанных групп искровых промежутков необходимо установить, существуют ли в них треки каких-либо заданных типов, и если они существуют, то описать эти треки в виде (в простейшем случае) пространственных прямых или дуг окружностей. Затем необходимо ответить на вопросы, связанные с конкретными экспериментами, относительно углов между треками, точек пересечения треков, радиусов кривизны дуг и т.п.

Следующим этапом анализа информации, снимаемой с искровых камер, является выделение таких комбинаций треков, которые отвечают заданным требованиям и могут принадлежать исследуемому каналу реакции. Эту задачу мы будем называть выделением положительной реализации.

В данном случае это означает создание так называемой семиотической модели управления большой системой, в роли которой выступают события, происходящие в искровой камере. Основным свойством больших

систем является наличие, вообще говоря, счётного числа состояний, которые не могут быть заданы аналитически. Для такого рода систем модель управления состоит из модели восприятия, модели укрупнения информации и модели принятия решений.

Рассмотрим основные этапы разработки системы автоматического анализа, предназначенной для выполнения траекторных измерений, а также для выбора из всего множества комбинаций треков в искровых камерах только положительных реализаций.

1. Построение языка описания реализаций экспериментов. Такой язык должен обеспечить : возможность проведения траекторных измерений, возможность формального описания реализаций экспериментов, возможность применения методов структурного распознавания, простоту использования для неспециалистов по вычислительной математике.

2. Построение процедуры автоматического получения описаний реализаций экспериментов в виде текстов построенного языка. Процедура использует первичные описания, задаваемые массивами координат искровых пробоев.

3. Построение процедур автоматической обработки текстов с целью выделения тех из них, которые являются описаниями положительных реализаций.

Построение любого языка заключается, в первую очередь, в выделении словаря терминальных (простейших) понятий, из которых строится любой текст данного языка.

Были выделены следующие простейшие понятия.

1. Наличие линейного трека в заданной группе искровых промежутков.
2. Наличие дугового трека в заданной группе искровых промежутков.
3. Пересечение треков в некоторой области пространства.
4. Попадание угла между треками в некоторый интервал.
5. Попадание угла между заданным направлением и треком в некоторый интервал.
6. Попадание в некоторый интервал радиуса дугового трека.
7. Попадание в некоторый интервал длины трека, если трек заканчивается в заданной группе искровых промежутков.

На рис. 1 приведены примеры реализаций гипотетических экспериментов и совокупности первичных понятий, соответствующих этим реализациям. При этом используются следующие обозначения: $ЛТ_i$ - линейный трек, $ДТ_i$ - дуговой трек, N_j - заданное направление, D_k^l - области пространства, содержащие точки пересечения треков, интервалы для углов, радиусов и точек окончания треков, $A(T_i, N_j)$ - угол между треком и заданным направлением, $A(T_i, T_j)$ - угол между треками в точке пересечения, $C(T_i, T_j)$ - точка пересечения треков, $R(T_i)$ - радиус кривизны трека, $E(T_i)$ - точка окончания трека. Символ T_i может принимать значения $ЛТ_i$ или $ДТ_i$.

Основной задачей при интерпретации реализаций эксперимента в терминах выбранных понятий является выделение (распознавание) линейных и дуговых проекций треков в каждой из рассматриваемых групп искровых промежутков. При этом необходимо решить ряд вопросов, связанных с тем, что в искровых промежутках, с одной стороны, возможно появление случайных искровых пробоев, не связанных с пролетом исследуемых частиц, с другой - эффективность искровых промежутков может отличаться от единицы и поэтому в некоторых из них пробой при прохождении частицы может отсутствовать.

После выделения проекций треков путем простого расчёта нетрудно получить описание реализаций экспериментов в терминах простейших понятий. Для этого, конечно, должны быть заданы описания геометрии экспериментальной установки и указаны интервалы углов, длин треков и т.д., соответствующие рассматриваемому каналу реакции. Выделение линейных и дуговых проекций треков разбивается на два этапа. На первом этапе выделяются проекции тех пробоев, которые группируются вблизи предполагаемых треков. На втором этапе через все эти точки (пробои) проводится линия регрессии, дающая наименьшее среднеквадратичное отклонение предполагаемой проекции траектории от истинной.

Для выделения проекций треков строятся специальные распознающие грамматики. Полагая, что в экспериментах используется однородное магнитное поле, направленное вдоль какой-либо из осей используемой системы декартовых координат, в одной проекции мы получим треки в виде отрезков прямой, а в другой - в виде дуг окружностей. Поэтому вводят-

ся две грамматики: грамматика Γ_x для выделения линейных проекций и Γ_y - для выделения дуговых проекций. Структура грамматики предполагает наличие

- словаря элементарных (терминальных) понятий,
- словаря вспомогательных (нетерминальных) понятий,
- правил вывода (P), раскрывающих вспомогательные понятия через элементарные. Грамматика Γ_x имеет терминальный словарь вида

$$V_T = \{ \Pi, \phi \},$$

где символ Π означает наличие пробоя, а ϕ - отсутствие пробоя.

Вспомогательные понятия образуют множество:

$$V_H = \{ \text{ЭТ, ЛКТ, ЛКТС, ЛТ} \},$$

где ЭТ - элементарный трек - наличие двух пробоев в соседних искровых промежутках, координаты которых удовлетворяют ограничениям на угол влета частиц; ЛКТ - линейный квазитрек - совокупность n ($n = 3, 4, \dots, K-1$) точек, лежащих в ϵ -окрестностях точек, которые принадлежат линейной проекции трека; ЛКТС - понятие, учитывающее возможность отсутствия пробоя в ϵ -окрестности точки линейной экстраполяции ЭТ или ЛКТ; ЛТ - начальный символ грамматики - совокупность точек, которые лежат в ϵ -окрестностях точек, образующих линейную проекцию трека, причем число этих точек больше некоторого наперед заданного числа K .

Символы грамматики приведены здесь без характеристик, в которых содержится информация о координатах точек, составляющих понятия, о номерах искровых промежутков и т.д.

Правила вывода грамматики Γ_x имеют вид:

$$P = \{ \text{ЛТ} \rightarrow \text{ЛТ} \overset{\wedge}{\Pi} \mid \text{ЛКТ} \overset{\wedge}{\Pi} \mid \text{ЛКТС} \overset{\wedge}{\Pi}, \\ \text{ЛКТ} \rightarrow \text{ЛКТ} \overset{\wedge}{\Pi} \mid \text{ЛКТС} \overset{\wedge}{\Pi} \mid \text{ЭТ} \overset{\wedge}{\Pi}, \\ \text{ЛКТС} \rightarrow \text{ЛКТ} \overset{\wedge}{\phi} \mid \text{ЭТ} \overset{\wedge}{\phi}, \\ \text{ЭТ} \rightarrow \Pi \overset{\wedge}{\Pi} \} .$$

Грамматика Γ_y , предназначенная для выделения дуговых проекций треков, аналогична Γ_x , поэтому особо здесь не рассматривается. Специфика заключается в вычислении величины и центра ϵ -окрестности продолжения трека в следующем искровом промежутке, а также в количестве

необходимой числовой информации, записанной в характеристиках понятий. Понятия ЛКТ, ЛКТС, ЛТ заменяются в этом случае на понятия ДКТ, ДКТС, ДТ соответственно.

После выделения совокупности проекций искровых пробоев, лежащих вдоль соответствующих проекций треков, необходимо провести линии регрессии, имеющие минимальные среднеквадратические отклонения от выделенных проекций искровых пробоев.

Таким образом, модель восприятия для системы автоматического анализа ядерных реакций в экспериментальной системе, содержащей искровые камеры, в данном случае состоит из грамматик Γ_x и Γ_y , алгоритмов построения линий регрессии, а также алгоритмов вычисления углов между треками, нахождения точек пересечения и т.д.

Следующим этапом анализа является создание модели укрупнения информации и принятия решений. Эта модель должна содержать средства описания событий в искровой камере, средства задания целей управления и алгоритмы принятия решений. Для построения модели укрупнения информации и принятия решений в данном случае используется лингвистическая модель А. Шоу^{/2/}, содержащая удобное средство для описания широкого класса образов - язык PDL. На этом языке могут быть формально описаны всевозможные реализации рассматриваемых классов физических экспериментов. При этом положительная реализация эксперимента задается контекстно-свободной грамматикой^{/3/}, и на множестве описаний реализаций эксперимента на языке PDL ищутся описания, содержащие тексты, которые выводимы в данной грамматике. Правила этой грамматики строятся в соответствии с физическими явлениями, которые лежат в основе положительных реализаций данного эксперимента. Рассмотрим конкретный пример построения грамматики для выделения положительных реализаций некоторого эксперимента.

Пучок частиц падает на мишень, возбуждая в ней всевозможные ядерные реакции. Необходимо выделить такие реализации эксперимента, в которых обнаруживается искривленный трек отрицательно заряженной частицы (предполагается, что в данном эксперименте используется однородное магнитное поле) либо совокупность треков, получаемая в результате одной из перечисленных ниже ядерных реакций.

Обозначим начальные и вторичные частицы через A_m , A_p , A_n (соответственно отрицательные, положительные и нейтральные частицы), а частицы мишени через P_m , P_p , P_n . Пусть нас интересуют реакции двух типов:

1. Взаимодействие (ВЗМ) отрицательно заряженной частицы с мишенью

$$\text{ВЗМ}_m: A_m + P_p \rightarrow A_n,$$

$$\text{ВЗМ}_m: A_m + P_p \rightarrow A'_m + A_p,$$

$$\text{ВЗМ}_m: A_m + P_p \rightarrow A'_m + A_p + A_n;$$

2. Реакции распада (РСП) начальной или одной из вторичных частиц:

$$\text{РСП}_m: A_m \rightarrow A'_m + A_n,$$

$$\text{РСП}_p: A_p \rightarrow A'_p + A_n,$$

$$\text{РСП}_n: A_n \rightarrow A_m + A_p.$$

Правила грамматики, определяющие положительные реализации эксперимента, должны соответствовать процессам, связанным с реакциями распада и взаимодействия, приведенными выше. При этом используются следующие простейшие (элементарные) изображения, из которых строятся тексты языка, порождаемого данной грамматикой

$$V_T = \{ DT_m, DT_p, e, \lambda \}$$

Каждое простейшее изображение характеризуется двумя характерными точками: началом (h) и концом (t), в которых осуществляется соединение этих простейших изображений между собой. При этом можно осуществить линейную запись достаточно сложных изображений с помощью скобок и специальных операторов соединения. В данном случае используются два таких оператора: $+$ – оператор и \times – оператор. Они обозначают соответственно соединение конца одного изображения с началом другого и соединение двух изображений в их началах.

Символы DT_m и DT_p обозначают дуговые треки отрицательных и положительных частиц соответственно. Символ e соответствует "невидимому" простейшему изображению, обозначающему траекторию нейтральной частицы. Символ λ обозначает точечное простейшее изображение (рис. 2).

Вспомогательные символы грамматики образуют следующее множество:

$$V_n = \{ A_m, A_p, A_n, P_p, ВЗМ_m, РСП_m, РСП_p, РСП_n, N \} .$$

Правила грамматики для выделения (порождения) изображений, соответствующих "положительным" реализациям, имеют в данном случае вид

$$A_m \rightarrow (DT_m + РСП_m) | (DT_m + ВЗМ_m) | DT_m$$

$$A_p \rightarrow (DT_p + РСП_p) | DT_p$$

$$A_n \rightarrow (e + РСП_n)$$

$$РСП_m \rightarrow (A_m \times A_n) | A_m$$

$$РСП_p \rightarrow (A_p \times A_n) | A_p$$

$$РСП_n \rightarrow (N + (A_m \times A_p))$$

$$ВЗМ_m \rightarrow (P_p + ((A_m \times A_p) \times A_n)) | (P_p + (A_m \times A_p)) | (P_p + A_n)$$

$$P_p \rightarrow \lambda$$

$$N \rightarrow \lambda$$

Рассмотрим одну из возможных реализаций эксперимента, изображенную на рис. 2. На рисунке приведена структурная запись этой реализации на языке PDL. Такая запись представляет собой так называемое простейшее структурное описание $T_s(a)$ объекта a (в данном случае объектом является реализация физического эксперимента). Следует заметить, что в случае применения лингвистической модели Шоу для решения задачи выделения положительных реализаций эксперимента необходимо автоматически получать их структурные описания на языке PDL. Это заставляет производить определенную перекодировку информации, получаемой в процессе работы модели восприятия. Модель восприятия выдает описание некоего мультиграфа, отражающего данную реализацию эксперимента, в виде, по существу, матрицы различного рода связей между отдельными треками. Существует алгоритм^{/2/} получения описаний таким образом заданных мультиграфов в виде линейных цепочек языка PDL. Несложные преобразования цепочек в соответствии с алгебраическими свойствами этого языка позволяют получить единственное для данной реализации так называемое "правильное описание", на котором производится дальнейший грамматический анализ в соответствии с правилами грамматики, порождающей всевозможные тексты, соответствующие "положительным" реализациям эксперимента.

Кроме структурной части $T_s(a)$, простейшее описание изображения содержит и так называемую семантическую часть $T_v(a)$, которую составляет числовая информация, характеризующая в данном случае координаты начал и концов дуговых треков, а также их радиусы кривизны.

Применяя правила грамматики к структурному описанию $T_s(a)$, можно получить так называемое иерархическое описание $H_s(a)$ (или грамматический анализ) данного изображения, что и соответствует процессу укрупнения информации.

Иерархическое описание рассмотренной реализации эксперимента представлено на рис. 3. Если описание $T_s(a)$ не соответствует положительной реализации эксперимента, то на некотором шаге применения правил грамматики для распознавания мы не найдем подходящего правила и не получим начального символа грамматики A_m . В таком случае

реализация не является "положительной" с точки зрения рассмотренной грамматики.

При построении иерархического описания необходимо наряду со структурной иерархией $H_s(\alpha)$ (правилами грамматики) рассматривать и семантическую иерархию $H_v(\alpha)$, т.е. правила преобразования числовых параметров символов грамматики. Эти правила преобразования вводятся в соответствии с тем, какая числовая информация необходима для дальнейшей статистической обработки, что определяется смыслом конкретных экспериментов.

Для экспериментальной проверки рассматриваемых в данной работе лингвистических методов анализа трековых изображений был использован алгоритмический язык рекурсивных функций (РЕФАЛ)/4,5/. Основной причиной использования РЕФАЛа для решения большинства проблем структурного распознавания послужил тот факт, что положенное в его основу понятие рекурсии находит широкое применение в лингвистических методах. Будучи языком, предназначенным в основном для всевозможных символьных преобразований, РЕФАЛ может быть эффективно использован для машинной реализации порождающих и распознающих грамматик, в которых, как и в РЕФАЛе, применяются рекурсивные процедуры.

В настоящее время модель восприятия системы автоматического анализа интерпретирована на языке РЕФАЛ и реализована на ЭВМ БЭСМ-6. Были просчитаны примеры на получение описаний в терминах простейших понятий ряда реализаций экспериментов.

Л и т е р а т у р а

1. Автоматический анализ сложных изображений. Сб. статей, М., "Мир", 1969.
2. А. Shaw. *Inf. and Control*, 14, N1, 9 (1969).
3. С. Гинзбург. Математическая теория контекстно-свободных языков, М., "Мир", 1970.
4. В.Ф. Турчин. Алгоритмический язык рекурсивных функций (РЕФАЛ). Препринт ИПМ АН СССР, 1968.

5. С.А. Романенко, В.Ф. Турчин. РЕФАЛ-компилятор. Труды 2-ой Всесоюзной конференции по программированию. Заседание Б., стр. 31. Новосибирск, 1970.

Рукопись поступила в издательский отдел

21 апреля 1971 года.

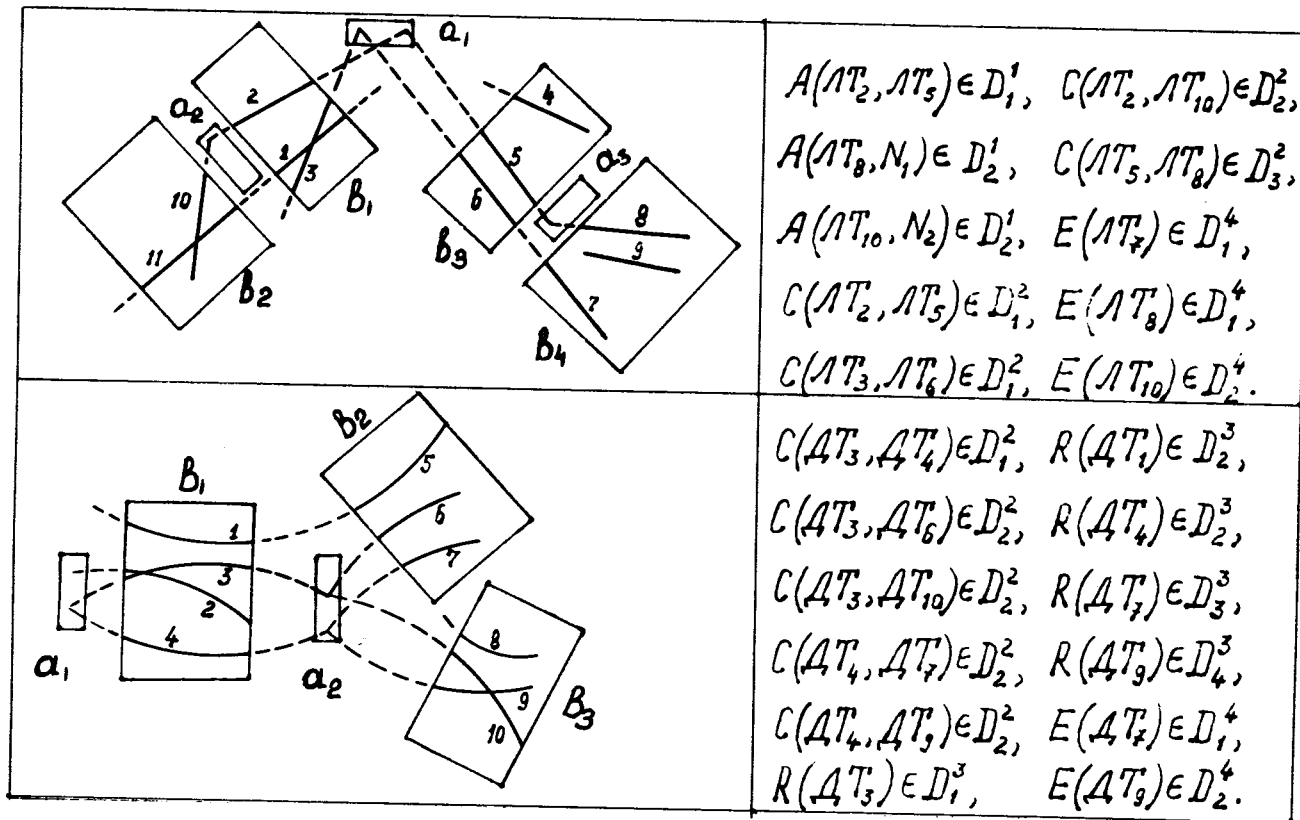


Рис. 1. Описание реализаций экспериментов в терминах простейших понятий. a_1, a_2, a_3 - рассеиватели, B_1, B_2, B_3 - группы искровых промежутков.

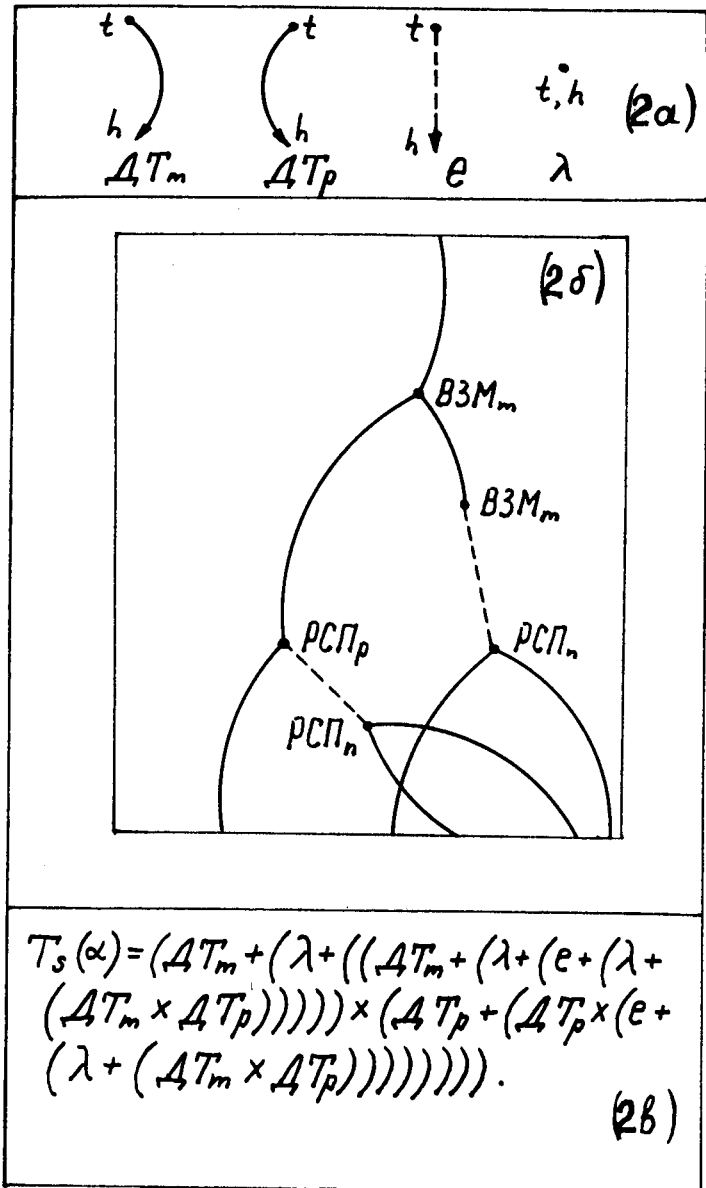


Рис. 2. Простейшее структурное описание трекового изображения:
 а - используемые простейшие изображения, б - пример системы треков
 (трековые изображения), в - структурное описание изображения.

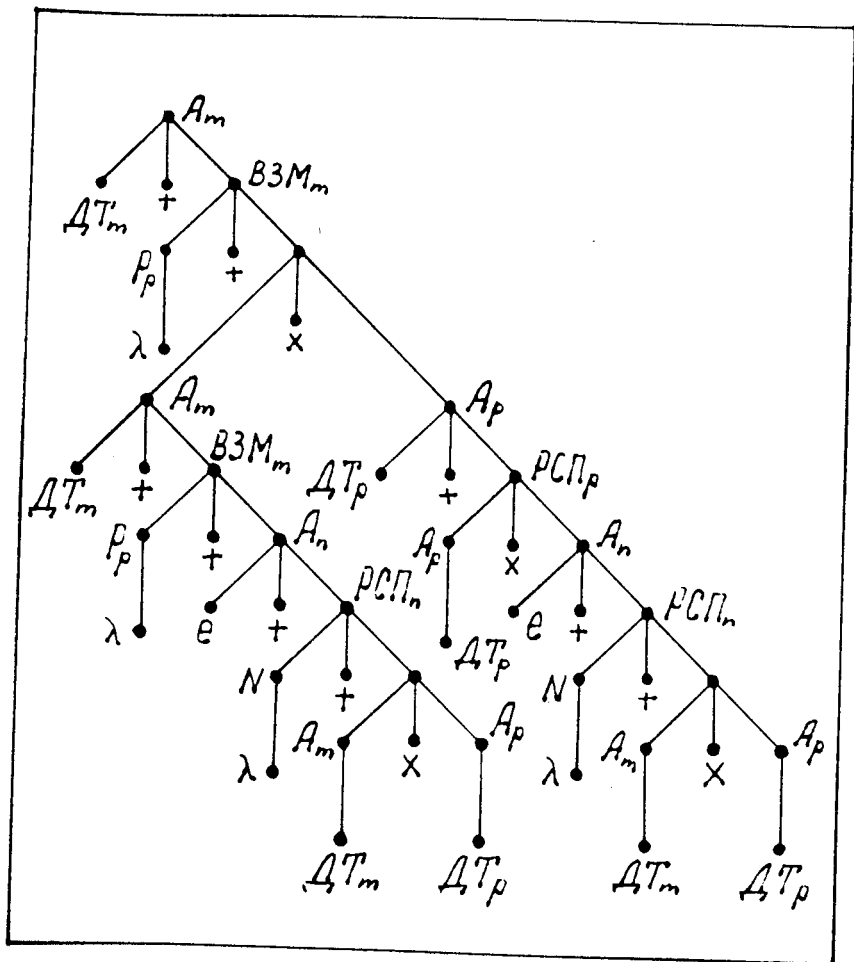


Рис. 3. Иерархическое описание трекового изображения.