

М. Н. Фаворская, А. М. Попов

ФОРМИРОВАНИЕ КОНТЕКСТНОЙ ГРАММАТИКИ ДЛЯ РАСПОЗНАВАНИЯ СЛОЖНЫХ СЦЕН С МНОГОУРОВНЕВЫМ ДВИЖЕНИЕМ ОБЪЕКТОВ

Рассматриваются вопросы формирования контекстных грамматик, описывающих как структурную информацию образа, так и взаимодействие образов в сложной сцене. Предложено использование трехуровневой грамматики, включающей задачу синтаксического анализа последовательности изображений (с расширенным содержимым основного и вспомогательного словарей) и задачу синтаксического анализа сцены с учетом многоуровневого движения объектов.

Ключевые слова: контекстная грамматика, синтаксическое распознавание, многоуровневое движение.

Изначально структурный, или лингвистический, подход был основан на использовании различных лингвистических конструкций, состоящих из словаря и правил конструирования фраз из заданного словаря. Такое структурное описание изображений позволяет проводить аналогию между структурой изображения и синтаксисом языка формальных грамматик. Отметим, что данное направление возникло в 60-х гг. XX в. как один из первых подходов к решению задач описания и распознавания изображений. Структурный подход позволяет не только отнести статический объект наблюдения к определенному образу, но и описать те свойства объекта, которые исключают его отнесение к другому образу.

Традиционные методы в рамках структурного подхода сводятся к синтаксическому описанию множеств сложных изображений с использованием небольшого набора производных элементов и грамматических правил. При этом считается, что изображения конструируются из соединенных различными способами элементов аналогично тому, как фразы и предложения естественных языков строятся путем соединения слов, а слова состояются из букв. Простейшие элементы, из которых строятся слова, а затем предложения, называются производными элементами. Правила конструирования композиций из производных элементов обычно задаются с помощью специальных грамматик описания изображений. Грамматическое правило (правило подстановки) может быть применено любое число раз, что позволяет достаточно компактно выражать основные структурные характеристики бесконечного множества предложений. Язык, который обеспечивает структурное описание изображений в терминах множества производных элементов и конструирования композиций этих элементов, называют языком описания изображений. В процессе распознавания осуществляется идентификация производных элементов и описание изображения в терминах заданного языка. Собственно распознавание состоит в синтаксическом анализе (или грамматическом разборе) «предложения», описывающего некоторое изображение. Процедура распознавания устанавливает синтаксическое соответствие анализируемого «предложения» или описания изображения заданной грамматике [1].

Система синтаксического распознавания образов состоит из трех основных модулей: модуля предварительной обработки, модуля описания и модуля синтаксического анализа. Модуль предварительной обработки осу-

ществляет кодирование, аппроксимацию, фильтрацию, восстановление и улучшение изображения. Модуль описания включает функции сегментации и выделения производных элементов на основе заранее заданных синтаксических операций. Каждая выделенная часть изображения идентифицируется относительно заданного множества производных элементов, а изображение в целом характеризуется набором цепочек производных элементов в виде структуры языкового типа. Модуль синтаксического анализа проверяет правильность составленных цепочек в рамках заранее заданных грамматик. Каждому образу соответствует своя заранее заданная грамматика, и если описание анализируемого изображения синтаксически верно в рамках данной грамматики, то изображение относится к тому образу, которому соответствует эта грамматика.

Формирование грамматики, описывающей как структурную информацию образа, так и взаимодействие образов, связано с необходимостью разработки алгоритма восстановления (или вывода) грамматики по заданному множеству динамических изображений, представляющих собой обучающую выборку. Такой алгоритм выполняет обучение распознающей системы. В результате обучения формируются структурные описания образов и описания их отношений, которые затем используются для синтаксического анализа событий и жанра сложной сцены. В простейшем случае процесс обучения не производится, а выбор грамматик и множеств производных элементов осуществляется учителем. Поскольку динамическая сцена с многоуровневым движением обладает очень сложной, постоянно изменяющейся во времени структурой, целесообразно применять контекстные грамматические правила, в связи с чем возникает понятие многоуровневой контекстной грамматики.

Рассмотрим некоторые основные положения, присутствующие в структурных методах описания сцены или распознавания.

Порождающая грамматика – это упорядоченный набор параметров $\mathbf{GR} = (V_T, V_N, P, S)$, где V_T – конечный алфавит, определяющий множество терминальных символов; V_N – конечный алфавит, определяющий множество нетерминальных символов; P – конечное множество правил вывода, т. е. множество пар следующего вида $u \rightarrow v$, где $u, v \in (V_T \cup V_N)^*$; S – начальный символ (аксиома грамматики), $S \in V_N$. Из терминальных символов состоят цепочки языка, порожденного грамматикой. Аксиомой

называется символ в левой части первого правила вывода грамматики. В грамматике \mathbf{GR} цепочка x непосредственно порождает цепочку y , если $x = \alpha u \beta$, $y = \alpha v \beta$ и $u \rightarrow v \in P$, т. е. цепочка y непосредственно выводится из цепочки x , что обозначается $x \Rightarrow y$. Языком, порождаемым грамматикой $\mathbf{GR} = (V_T, V_N, P, S)$, называется множество терминальных цепочек, выводимых в грамматике \mathbf{GR} из аксиомы: $L(\mathbf{GR}) = \{x \mid x \in V_T^*; S \Rightarrow^* x\}$, где символ \Rightarrow^* – выводимость.

Правила порождающих грамматик позволяют осуществлять преобразования строк. Ограничения же на виды правил позволяют выделить классы грамматик. Классификация, предложенная Н. Хомский, определяет четыре типа грамматик:

1) грамматики типа 0 – это грамматики, на правила вывода которых нет ограничений;

2) грамматики типа 1 (контекстные грамматики) – это грамматики, все правила которых имеют следующий вид: $xAy \rightarrow x\varphi y$, где $A \in V_N$, $x, y, \varphi \in (V_N \cup V_T)^+$;

3) грамматики типа 2 (бесконтекстные, или контекстно-свободные грамматики – КС-грамматики). Правила вывода для этих грамматик имеют следующий вид: $A \rightarrow \varphi$, где $A \in V_N$, $\varphi \in (V_N \cup V_T)^*$;

4) грамматики типа 3 – это автоматные грамматики, которые делятся на два типа:

– левосторонние (леворекурсивные), правила вывода для которых имеют следующий вид: $A \rightarrow aA \mid a$, где $A \in V_N$;

– правосторонние (праворекурсивные), правила вывода для которых имеют следующий вид: $A \rightarrow Aa \mid a$.

При этом язык L называется языком типа i , если существует грамматика типа i , порождающая язык L . Дерево вывода часто называется деревом грамматического разбора, или синтаксическим деревом, а процесс построения дерева вывода – грамматическим разбором (синтаксическим анализом). Одной цепочке языка может соответствовать больше одного дерева, так как эта цепочка может иметь разные выводы, порождающие разные деревья. Так, например, КС-грамматика $\mathbf{GR} = (V_T, V_N, P, S)$ называется неоднозначной (неопределенной), если существует цепочка $x \in L(\mathbf{GR})$, имеющая два или более дерева вывода. Однако дерево грамматического разбора не следует путать с представлением грамматики в виде графа. Граф грамматики в качестве вершин содержит синтаксические формы (любые цепочки, выводимые из аксиомы).

Основной недостаток рассмотренных грамматик состоит в том, что они в большей мере приспособлены для описания сцен, чем для распознавания. Указанный недостаток в значительной степени преодолен в работах, выполненных под руководством М. И. Шлезингера, с использованием метода двумерного программирования. Предложенная в работе [2] двумерная грамматика \mathbf{GR}_S представляет собой функцию следующих шести параметров:

$$\mathbf{GR}_S = \langle V_O, \mathbf{S}, T_V, T_S, R, \{Z, Z(t, t'); (t, t') \in \mathbf{N}\} \rangle.$$

Пусть на изображениях T_V заданы распознаваемые изображения объектов. Каждое из этих изображений представляет функцию, заданную на изображениях T_V и принимающую значения из алфавита объектов V_O , соответствующего основному алфавиту в одномерных грамматиках. Из элементарных изображений этого алфавита

составляются более сложные изображения. Кроме алфавита сигналов вводится алфавит \mathbf{S} структурных элементов, соответствующий алфавиту вспомогательных символов в одномерных грамматиках. Структурные элементы, с одной стороны, определяют допустимые значения соответствующих сигналов, а с другой стороны, позволяют выразить ограничения, накладываемые на структуру изображения в целом в виде локальных ограничений. Из структурных элементов составляется описание изображения, которое определяется как функция \bar{S} , заданная на конечном множестве T_S (описание) и принимающая значения из множества \mathbf{S} . В общем случае описания не обязательно изоморфны изображениям. Элемент множества $\mathbf{Z} = V_O \cup \mathbf{S}$ называется символом и обозначается z . Множество \mathbf{T} представляет собой объединение изображения и описания. Такой элемент называется клеткой и обозначается как t . Две клетки t и t' считаются соседними, если некоторый фиксированный для данной грамматики симметричный предикат $R(t, t')$ равен единице. При этом \mathbf{N} – множество пар соседних клеток.

Вариантом \bar{Z} называется пара (\bar{V}, \bar{S}) изображение – описание. Это означает, что вариант – это функция, заданная на множестве $\mathbf{T} = T_V \cup T_S$ и принимающая значения из множества \mathbf{Z} , причем такая, что $Z(t) \in V_O$, если $t \in T_V$ и $Z(t) \in \mathbf{S}$, если $t \in T_S$. Для каждой пары соседних клеток t и t' считаются заданными множества \mathbf{Z} , $Z(t, t')$ допустимых пар (Z, Z') символов $Z, Z' \in \mathbf{Z}$. Вариант \bar{Z} называется допустимым, если для любой пары $(t, t') \in \mathbf{N}$ выполняется соотношение $(Z(t), Z(t')) \in \mathbf{Z}, Z(t, t')$. Изображение \bar{V}^* называется допустимым, если существует допустимый вариант $\bar{Z}(\bar{V}^*, \mathbf{S})$. Если вариант $\bar{Z}(\bar{V}^*, \mathbf{S})$ является допустимым, то описание \bar{S} называется возможным описанием изображения V^* .

Зададим с помощью двумерной грамматики \mathbf{GR}_S не все множество изображений \mathbf{X}^* , относящихся к одному зрительному образу, а его незначительную часть, называемую множеством $V^*(\mathbf{GR}_S)$ идеальных или эталонных изображений. Любому эталонному изображению $\bar{V} \in V^*(\mathbf{GR}_S)$ соответствует некоторое множество реальных изображений, похожих на эталонное изображение \bar{V} . Функция принадлежности $f_V(\mathbf{X})$ распознаваемого изображения множеству $V^*(\mathbf{GR}_S)$, называемая сходством, принимает различные значения (не обязательно значения, равные 0 или 1). Задача синтаксического анализа изображения \mathbf{X} заключается в нахождении эталонного изображения, порождаемого грамматикой \mathbf{GR}_S и максимизирующей функцию сходства

$$V^*(\mathbf{X}) = \arg \max_{\bar{V} \in V^*(\mathbf{GR}_S)} f_V(\mathbf{X}). \quad (1)$$

В работе [2] предложен метод решения данной задачи, названный методом двумерного программирования. Метод позволяет одновременно с оптимальным изображением $V^*(\mathbf{X})$ получить также описание \mathbf{S}^* , соответствующее этому изображению, т. е. найти оптимальный допустимый вариант $B^* = (\bar{V}^*, \mathbf{S}^*)$. Наиболее существенными особенностями двумерных грамматик являются их универсальность (любое множество изображений можно задать с помощью соответствующей двумерной грамматики) и конструктивность (разработаны эффективные алгоритмы нахождения оптимального допустимого ва-

рианта B^*). Другое достоинство таких алгоритмов заключается в том, что они работают непосредственно со зрительным сигналом, что не исключает возможности использования различных методов предварительной обработки изображений. Отметим, что элементарные изображения грамматик, описывающих сложные эталонные изображения, могут иметь постоянные размеры (что сужает возможности двумерного программирования) и неодинаковые размеры. В последнем случае такие элементарные изображения относятся к так называемым блочным двумерным грамматикам.

Решение задачи (1) при большом уровне помех требует довольно значительных вычислительных затрат при решении практических задач. Однако, как правило, существует возможность распараллеливания вычислительного процесса при реализации алгоритмов двумерного программирования. При малых уровнях помех такие алгоритмы не являются более сложными, чем другие алгоритмы анализа изображений, но обеспечивают более высокую надежность результатов.

Однако двумерная грамматика М. И. Шлезингера предназначена для распознавания простейших бинарных графических примитивов в статических сценах. Для динамических сцен с многоуровневым движением подсистема синтаксического распознавания образов значительно усложняется. При этом появляются временные отношения между объектами, для описания которых требуется разработка подсистемы распознавания связей между образами. Именно на подсистему распознавания связей возлагается реализация основных четырех принципов распознавания динамических объектов: учет цели распознавания на начальных стадиях обработки видеопоследовательности; распознавание поведенческих ситуаций динамических объектов; оценка предыстории динамических объектов; переменное количество объектов наблюдения в сложных сценах.

Контекстная грамматика распознавания сложных сцен с многоуровневым движением объектов должна иметь средства для реализации следующих процедур:

- предварительной сегментации сцены;
- описания регионов с локальными признаками движения;
- группирования регионов с локальными признаками движения по принципу соседства;
- распознавания видеообъектов;
- группирования видеообъектов с глобальными признаками движения по уровням;
- описания многоуровневого движения в сцене;
- распознавания временных событий в сцене;
- распознавания жанра сцены (для цифровых видеотеки).

Анализ представленных процедур показывает, что в случае распознавания сложных динамических сцен с многоуровневым движением целесообразно использование трехуровневой грамматики следующего вида:

$$GR_D = \langle V_{O,E,G}, S_{S,LM,GM}, T_V, T_S, T_E, R_E \rangle$$

$$\{ \{E, E(a, a'); (a, a') \in M\}, R_O, \{R_R, \{Z, Z(t, t'); (t, t') \in N\}\} \},$$

где $V_{O,E,G}$ – основной словарь объектов, временных событий, жанров сцены; $S_{S,LM,GM}$ – вспомогательный сло-

варь структурных элементов, локальных признаков движения и глобальных признаков движения; R_R – предикат построения регионов; R_O – предикат построения объектов; R_E – предикат временных событий. Элемент множества $E = V_{O,E,G} \cup S_{S,LM,GM}$ называется событием. Множество T_E описывает набор событий. Множество $T = T_V \cup T_S \cup T_E$ в данном случае представляет собой объединение событий и описания.

Контекстная грамматика распознавания сложных сцен с многоуровневым движением объектов реализует две задачи: задачу синтаксического анализа последовательности изображений X (с расширенным содержимым основного $V_{O,E,G}$ и вспомогательного $S_{S,LM,GM}$ словарей) в соответствии с уравнением (1) и задачу синтаксического анализа сцены SC . Рассмотрим их более подробно.

Целью синтаксического анализа последовательности изображений X является распознавание динамических объектов, которые целесообразно разделить на две большие группы:

- объекты, состоящие из регионов с постоянными цветовыми, текстурными характеристиками в определенных условиях освещения и обладающие фиксированным набором проекций во фронтальной плоскости, контур которых может изменяться в соответствии с аффинной или проективной группой преобразований (объекты техногенного происхождения);

- объекты, состоящие из регионов с постоянными цветовыми, текстурными характеристиками в определенных условиях освещения и обладающие произвольным набором проекций во фронтальной плоскости с произвольно изменяющимся контуром (объекты антропометрического происхождения). При этом регионы характеризуются постоянными на некотором временном интервале относительными направлениями и значениями скоростей движения.

Эти группы характеризуются несколько отличающимися признаками, причем разброс проекций антропометрических объектов компенсируется локальными признаками движения отдельных статистически однородных регионов. Вопросы распознавания объектов с ограниченным допустимым количеством проекций наиболее хорошо проработаны, и можно предложить следующую формальную схему распознавания таких объектов. Предположим, что каждый образ представлен только одним изображением. Назовем его исходным эталоном образа и обозначим через v^j . Пусть задано также множество допустимых преобразований G_b исходного эталона, параметризованное по мешающему параметру b . Результатом применения преобразования G_b к эталону v^j является преобразованный эталон

$$V(j, b) = G_b v^j.$$

Множество значений, которые принимает эталон $V(j, b)$ при фиксированном значении j и возможных значениях $b \in B$, принято называть областью эталонов j -го образа. Наблюдаемые изображения представляют собой реализации многомерной случайной величины с известным распределением вероятностей $P(X/V(j, b))$, зависящим от преобразованного эталона $V(j, b)$ как от многомерного параметра. Величина $V(j, b)$ является математическим ожиданием или модой этого распределения.

Такая формальная схема позволяет при известном распределении $P(\mathbf{X}/V)$ и известной зависимости преобразованного эталона от параметров j и b решить задачу распознавания. Для решения задачи применим метод максимального правдоподобия. Для принятия решения о значении параметра j необходимо найти максимум функции правдоподобия по параметрам j и b и взять в качестве решения d то значение j , при котором достигается этот максимум:

$$\mathbf{V}^*(\mathbf{X}) = \arg \max_j \max_{b \in B} P(\mathbf{X}/V(j, b)). \quad (2)$$

Решение d не изменится, если заменить функцию правдоподобия любой другой функцией от параметра, значения которой связаны со значениями $P(\mathbf{X}/V(j, b))$ монотонно возрастающей зависимостью, т. е. если $g(\mathbf{X}, V)$ – любая функция от параметров \mathbf{X} и V , удовлетворяющая условию

$$P(\mathbf{X}/V) = f(g(\mathbf{X}, V)), \quad (3)$$

где $f(\cdot)$ – монотонно возрастающая функция, то правило (2) может быть заменено следующим выражением:

$$\mathbf{V}^*(\mathbf{X}) = \arg \max_j \max_{b \in B} g(\mathbf{X}, V(j, b)). \quad (4)$$

Положение принципиально не изменится, если функция $f(\cdot)$ будет монотонно убывающей, но при этом в выражениях (2) и (4) операцию максимизации следует заменить операцией минимизации. Поскольку решающее правило (4) можно понимать как правило нахождения таких значений j и b , при которых сходство наблюдаемого изображения \mathbf{X} с преобразованным эталоном $V(j, b)$ максимально, то значение всякой функции $g(\mathbf{X}, V)$, удовлетворяющей условию (3), можно понимать как меру сходства эталона V с изображением \mathbf{X} . Решением является значение j , найденное в соответствии с выражением (4).

Синтаксический анализ последовательности изображений объектов с неограниченным количеством проекций происходит намного сложнее. Здесь невозможно задать заранее множество допустимых преобразований, требуется рекуррентная процедура слежения за регионами с локальными признаками движения с их последующим группированием в единый видеообъект. В этом случае изменяется структура эталона $\mathbf{V}^*(\mathbf{X})$, состоящего из множества регионов, каждый из которых обладает своим набором локальных признаков движения. Таким образом, рекуррентная процедура нахождения эталонного изображения $\mathbf{V}_i^*(\mathbf{X})$ на i -м шаге, порождаемого грамматикой \mathbf{GR}_D , максимизирует функцию сходства следующего вида:

$$\mathbf{V}_i^*(\mathbf{X}) = \mathbf{V}_{i-1}^*(\mathbf{X}) + \gamma_i f_V(\mathbf{X}_{i-1}, \Delta_i), \quad (5)$$

где $i, i-1$ – шаги аппроксимации; γ_i – некоторая функция, зависящая от шага аппроксимации (например, последовательность положительных чисел); Δ_i – изменяемая в процессе анализа последовательности изображений величина; $\mathbf{V}_i^*(\mathbf{X}) = \arg \max_j P(\mathbf{X}/V(j))$ – функция сходства на i -м шаге.

Функция (5) является разновидностью метода стохастической аппроксимации для решения задачи обучения распознаванию образов. Для организации рекуррентной процедуры (5) важную роль играет выбор функции потерь. Так, например, можно применить следующее пра-

вило: если с помощью некоторой разделяющей функции изображение классифицируется правильно, то штраф равен нулю. Если классификация произведена неверно, то назначается штраф, величина которого пропорциональна расстоянию от вектора, соответствующего распознаваемому изображению, до разделяющей гиперплоскости.

Целью синтаксического анализа сцены **SC** является распознавание событий, производимых одиночными объектами, взаимодействующими объектами, а также определение жанра динамической сцены. Эти вопросы относятся к проблемам понимания изображений и анализу сцен. В случае сложной сцены до распознавания событий следует создать модель многоуровневого движения, т. е. определить количество значимых уровней (в простейшем случае принимают решение о существовании двух уровней – переднего плана и фона) и соотнести каждый распознанный видеообъект с тем или иным уровнем. Такая задача является наиболее востребованной для случаев виртуальной 3D-реконструкции в картографии, системах навигации и т. д., когда видеодатчик установлен на перемещающуюся платформу, и происходит относительное движение всех объектов сцены. При этом возникает впечатление, что объекты, расположенные ближе к камере, «перемещаются» быстрее, чем удаленные объекты. В этом случае модель глобального движения можно назвать моделью многоуровневого движения, определяющей набор различных, но внутренне подобных уровней движения, которые отвечают твердым телам, расположенным на различном расстоянии от движущейся камеры и k сегментам на изображении [3]. Предположим, что уровни движения заданы в параметрической форме, и имеется h уровней движения. Для последовательности изображений требуется определить: а) уровень движения, к которому относится каждый видеообъект; б) значения параметров каждого уровня. По известному уровню движения определяются значения параметров уровня, и наоборот, зная значения параметров, можно установить, к какому уровню движения относится видеообъект.

При синтаксическом анализе сцены временные события, накопленные в виде некоторой предыстории, уже допускают разделение на классы движений и их интерпретацию на понятийном уровне. Движения в видеопоследовательностях с учетом их повторяемости во времени и пространстве можно разделить на три класса: временные текстуры, активные действия и события. Временные текстуры определяются как статистические регулярности в пространстве и времени (морские волны, движение облаков, листва, птиц в стае и т. д.). Активные действия интерпретируются как некие повторяемые во времени (но не в пространстве) структуры (хождение людей, танцы, движения отдельных животных, насекомых и т. д.). События состоят из изолированных простых движений, не повторяющихся во времени и пространстве (выражения лица, приход в помещение, бросание мяча и т. д.). Для динамически изменяемой сцены появляются дополнительные характеристики объектов, такие как предыстория движения объекта и процедура ассоциации, используемой для окончательного формирования понятий и для интерпретации сцены в терминах этих понятий. Предыстория движения объекта как функция координат от вре-

мени может быть подвергнута грубой аппроксимации, поскольку здесь требуется установить характер движения, а не его конкретные характеристики. Затем она интерпретируется как некое событие движения объекта на понятийном уровне, при этом существенную роль играет анализ временных отношений между объектами (рукопожатие, обсуждение, агрессивное действие и т. д.) [4].

Для интерпретации сцены следует воспользоваться процедурой ассоциации, которую характеризуют два показателя [5]:

– величина ассоциации как мера сходства, учитывающая близость векторов признаков объектов сцены, близость относительных преобразований этих объектов и важность объектов;

– след ассоциации как подмножество объектов, охваченных глобальным событием движения.

Для формирования ассоциаций с каждым объектом базы знаний O_j , помимо его непосредственного описания, сопоставлен набор дополнительных характеристик, значения которых вычисляются в соответствии с общей динамикой сцены:

– близость $r_j^t = r(O_j, O^t)$ к объекту интереса, определяющая принадлежность следу;

– близости $q_k^t = q(V_k, V^t)$ отношений с соседними (по связям) объектами O_k ;

– величина ассоциации $a_j^t = \max_k(0, a_j^{t-1} - c, q_k^t \cdot a_k^t + r_j^t)$, ($c \ll 1$), определяющая важность, по максимуму которой на след ассоциации обращается особое внимание.

Формирование понятий основано на пропорциональном величинам ассоциации увеличению весов объектов, принадлежащих следам ассоциаций. Понятия строятся в виде часто встречающихся подструктур, состоящих из достаточно важных объектов.

Таким образом, нами были рассмотрены основы проектирования формальных грамматик в рамках структурного подхода распознавания образов. Предложена структура системы синтаксического распознавания образов, состоящая из модуля предварительной обработки, модуля описания и модуля синтаксического анализа. Подроб-

но рассмотрена двумерная грамматика М. И. Шлезингера для распознавания бинарных графических примитивов в статических сценах. Показано, что для распознавания сложных сцен с многоуровневым движением объектов целесообразно использовать трехуровневую грамматику, включающую основной словарь объектов, временных событий, жанров сцены, вспомогательный словарь структурных элементов, локальных признаков движения и глобальных признаков движения, предикат построения регионов, предикат построения объектов и предикат временных событий. Предложены процедуры распознавания объектов, основанные на допустимых преобразованиях и рекуррентной процедуре стохастической аппроксимации в зависимости от количества допустимых проекций видеообъектов во фронтальной плоскости. Для интерпретации сложной сцены с многоуровневым движением разработана процедура ассоциации, вычисляющую близость векторов признаков объектов сцены.

Библиографические ссылки

1. Фаворская М. Н. К вопросу об использовании формальных грамматик при распознавании объектов в сложных сценах // Решетневские чтения : материалы XIII международного науч. конф. Ч. 2. Красноярск, 2009. С. 540–541.
2. Шлезингер М. И. Синтаксический анализ двумерных зрительных сигналов в условиях помех // Кибернетика. 1976. № 4. С. 76–82.
3. Фаворская М. Н. Вероятностные методы сегментации видеопотока как задача с недостающими данными // Вестник СибГАУ. Вып. 3 (16). Красноярск, 2007. С. 4–8.
4. Video Event Classification Using Bag of Words and String Kernels / L. Ballan, M. Bertini, A. del Bimbo, G. Serra // ICIP'09, 2009. P. 170–178.
5. Фаворская М. Н. Локальные пространственно-временные признаки событий в видеопоследовательностях // Теорет. и прикл. вопросы соврем. информ. технологий : материалы X международного науч.-техн. конф. Ч. 2. Улан-Удэ, 2009. С. 461–466.

M. N. Favorskaya, A. M. Popov

FORMING OF CONTEXT-DEPENDENT GRAMMAR FOR DESCRIBING OF COMPLEX SCENE WITH MULTI-LEVEL MOVING OF OBJECTS

The problems of forming of the context-dependent grammar which describes structural information of both image and interaction of the images in a complex scene are considered. The work suggests the application of three-level grammar based on the task of image sequences syntactic analysis (with extended contents of the main and auxiliary dictionaries), and the task of scene syntactic analysis with multi-level moving of objects.

Keywords: context-dependent grammar, syntactic analysis, multi-level movin.

© Фаворская М. Н., Попов А. М., 2010