

РАЗРАБОТКА ЭЛЕМЕНТОВ ПРЕДСТАВЛЕНИЯ ПЕРЦЕПТИВНЫХ ЗНАНИЙ В СИСТЕМАХ КОМПЬЮТЕРНОГО ЗРЕНИЯ

© 2010 г. А. С. Рожков

Национальный исследовательский университет информационных технологий,
механики и оптики, Санкт-Петербург

E-mail: alexisrozhkov@gmail.com

Проведена разработка основы для создания расширяемой системы взаимодействия базы знаний с априорно заданными процедурами обработки изображений. Экспериментальные исследования предложенной модели на изображениях из тестовой выборки свидетельствуют о возможности использования данной системы для создания элементов представления перцептивных знаний для систем интерпретации изображений.

Ключевые слова: обработка изображений, интерпретация сцен, представление знаний.

Коды OCIS: 100.2960

Поступила в редакцию 08.06.2010

Введение

За последнее время возрос интерес к когнитивным роботам, одной из областей применения которых может быть создание “роботов-помощников”, которые бы выполняли опасные или трудоемкие задания. Возможности когнитивных роботов включают в себя обработку данных, полученных в результате восприятия окружающей среды, принятие решений и планирование своих дальнейших действий. Одним из ключевых аспектов функционирования данных роботов является получение полных и достоверных сведений об окружающей среде, большая часть которых будет получена за счет зрительного восприятия. Системы компьютерного зрения, удовлетворяющие этим критериям, активно разрабатываются [1], однако результаты получены только в тех областях, в которых свойства рассматриваемых сцен или изображений лежат в довольно узких пределах. Основной причиной этого является отсутствие общего подхода к проблемам компьютерного зрения.

В системах компьютерного зрения сначала всегда проводятся обработка и анализ изображений, которые в зависимости от приложения могут быть разных типов: статические и динамические, двух- и трехмерные, с одной или нескольких камер и др. В результате анализа

составляется некое символическое описание изображения, которое затем можно использовать, начиная от определения взаимного положения и ориентации предметов сцены и кончая принятием решений и навигацией.

Принято разделять типы описаний изображений [2] на низкоуровневые, описывающие изображения, состоящие из численных массивов, созданные соответствующей аппаратурой или программным обеспечением. Это и есть изображения в их обычном виде, хотя это могут быть и карты высот, томограммы, электрокардиограммы, радиосигналы и др. К этому типу представления изображений принято относить и результаты преобразований изображений, например, преобразованием Фурье. Следующий тип – это среднеуровневые описания, более известные как “промежуточные символические представления” (ISR), использующие символические обозначения для областей, линий и поверхностей, выделенных из низкоуровневого изображения с помощью алгоритмов сегментации и группировки. И, наконец, высокоуровневые описания, где изображения представляются такими семантическими категориями, как “дом”, “дорога”, “люди” и проч. Помимо назывной компоненты в этом уровне может присутствовать и процедурная, характеризующая связь объекта с другими объектами и его поведение.

Использование высокоуровневого описания вместо исходного низкоуровневого предпочтительнее, потому что, во-первых, оно будет занимать меньший размер (хотя он будет зависеть от “разрешения” описания, т. е. количества атрибутов, которое будет иметь каждый объект) и, во-вторых, поиск и анализ проще и быстрее проводить в структурированной базе данных, которой и является высокоуровневое описание.

Как правило, задачи, которые ставятся перед системами компьютерного зрения, формулируются в высокоуровневых терминах, характеризующих объекты и ситуации [3]. В традиционных методах анализа изображений, не использующих в явном виде представления знаний, эти знания тоже используются, но неявно, в процедурной форме [4]. Это усугубляет проблему так называемой “семантической пропасти” между исходными данными и конечной целью. К примеру, для получающих все большее распространение роботов-спасателей цель может формулировать в виде: “Пройти по коридору, зайти во вторую слева дверь и установить наличие людей в данной комнате”. При процедурном подходе анализ подобных целей может стать чрезмерно сложной задачей, требующей изначально заложенных знаний о геометрии помещения, в то время как при декларативном подходе достаточно определить основные предикаты, такие как определение положения и выполнение действий, после чего они могут быть применены по отношению к объектам или областям, обнаруженным в результате сегментирования изображения со зрительных сенсоров.

Явное представление знаний позволяет создавать легко расширяемые системы, хотя и в этих случаях знания обычно закладываются априорно, причем процедуры обработки и анализа изображений оказываются “внешними” по отношению к системам представления знаний, строящимся в стиле традиционных экспертных систем, часто работающих в предположении о “замкнутости мира” [5]. В результате, преодоление “семантической пропасти” не выполняется.

Таким образом, актуальной является задача детального исследования таких представлений знаний в системах компьютерного зрения, которые бы захватывали и начальные уровни представления изображений. В данной работе проводится предварительный анализ способов представления знаний систем компьютерного зрения, основанных на знаниях. Рассматриваются механизмы интерпретации изображений, объединенные

с динамически строящимися базами знаний, основанными на логических представлениях.

Представление знаний в системах компьютерного зрения

Представление знаний – это вопрос, который связывает воедино такие области, как когнитивную психологию и искусственный интеллект. Если ставить целью имитацию зрения живых организмов с помощью компьютера, то представление знаний тоже должно в какой-то мере повторять процессы, происходящие при накоплении и организации знаний животными или человеком.

Проблема представления знаний стала доминирующей в области искусственного интеллекта с середины 1970-х годов, чему способствовала бурно развивающаяся в то время компьютерная лингвистика. Системы, основанные на знаниях, нашли применение как экспертные системы, которые при использовании знаний о предметной области оказались способны строить формальные описания задач, сформулированных на ограниченном естественном языке для одной узкой предметной области.

Существуют несколько подходов к организации и хранению знаний. Процедурный подход характеризуется необходимостью в явном виде иметь описание процедуры получения конкретного результата. Процедурные знания описывают порядок и характер преобразования объектов предметной области и составляют процедурную предметную модель. Как правило, под процедурными представлениями подразумевается неявное использование знаний, включенное в процедуры вывода (“рассуждений”), реализованные в виде компьютерных программ.

В противоположность процедурному подходу при декларативном подходе дается лишь набор сведений, вывод на основе которых осуществляется программным модулем, отделенным от самой базы знаний. Декларативные знания определяют содержательную, или семантическую часть предметных знаний и порождают семантическую предметную модель, которая представляет собой последовательность семантических фактов, т. е. фактов, простейших по составу, имеющих предметный смысл.

У декларативного подхода по сравнению с процедурным есть некоторые преимущества. В нем знания не привязаны к механизмам логического (или иного) вывода, что упрощает их независимую модификацию и позволяет исполь-

зовать единожды разработанную систему для работы со знаниями, описывающими независимые предметные области. При данном подходе язык представления знаний, как правило, имеет более простой синтаксис, что упрощает работу пользователей, не являющихся программистами, а также автоматизацию создания подобных баз. Для систем компьютерного зрения важным является внутренняя интерпретируемость знаний, представленных в декларативной форме, т. е. система оказывается осведомленной об имеющихся у нее знаниях. Декларативные представления обычно описывают знания как совокупность однотипных элементов, например, логических формул, т. е. воплощают модульный принцип.

С другой стороны, при декларативном представлении процесс вывода гораздо труднее контролируется и оптимизируется, поскольку процедурная компонента этого процесса в явном виде не представляется. Для систем компьютерного зрения, имеющих дело с большим объемом информации, этот аспект может оказаться принципиальным и требует отдельного анализа. Кроме того, декларативные представления менее выразительны и не всегда достаточно прозрачны, поскольку их интерпретация может зависеть от конкретной реализации системы манипулирования знаниями.

В силу этого напрашивается вывод о целесообразности использования декларативного метода описания знаний в системах компьютерного зрения, так как несмотря на некоторую сложность его внедрения, он позволяет делать эти системы более гибкими, что и является нашей целью.

На данный момент лучше всего исследованы семантические знания, поскольку они тесно связаны с работами в области традиционных экспертных систем. Однако, если стремиться преодолеть “семантическую пропасть”, имеет смысл реализовать систему перцептивных знаний, которые являются неотъемлемым элементом зрительной системы не только у человека, но и у высокоразвитых животных.

Восприятие (перцепция, от лат. *perceptio*) – познавательный процесс, формирующий субъективную картину мира. В общем случае перцептивные знания – это знания, полученные организмом в течение жизни, или же, в данном случае, функционирования компьютерной программы. У биологического организма это знания, полученные от органов чувств, при компьютерной реализации это данные, полученные от

сенсоров, к которым обращается программа. За счет постоянного накопления перцептивных знаний, происходящего при таком подходе, расширяется база фактов программы, из которой средствами логического вывода можно синтезировать новые знания, не связанные с априорно заданными. Это дает такой программе большой выигрыш по сравнению с классическими подходами к системам компьютерного зрения.

Реализация системы перцептивного знания

В данной работе разработаны базовые компоненты системы представления перцептивных знаний, связываемых с процедурами обработки изображений, формирующих промежуточные символьные описания. Изображения при высокоуровневом описании, вероятно, представляются в терминах областей, их форм, пространственных отношений, цветовых и прочих характеристик.

Таким образом, следует провести сегментацию изображения для выделения на нем объектов, или разбиения его на области. Затем каждому объекту, или области сопоставляется набор его свойств – от простейших, таких как цвет, форма и пространственное положение, до более сложных, таких как состояние или принадлежность к некоторому классу. Затем над накопленной базой фактов могут проводиться операции вывода. В рамках данной работы используется логический вывод в рамках исчисления предикатов первого порядка в силу его наибольшей проработанности и наличия программных реализаций с открытым кодом.

В качестве примера рассмотрим работу алгоритма определения состояния светофора на изображении, оперируя исключительно высокоуровневыми объектами и абстрактными понятиями, которые автоматически задаются программным модулем при сегментации и анализе изображения и формировании его высокоуровневого представления. Скажем, на нашем изображении после сегментирования обнаружен набор объектов, включающий светофор и лампы, принадлежащие ему. Каждый объект имеет параметры цвета и принадлежности к другому объекту, а также параметры взаимного расположения. На языке Пролог данную базу фактов можно представить следующим образом:

```
color(obj1, red). %% красная лампочка светофора горит
```

```

color(obj2, black). %% желтая и зеленая вы-
ключены
color(obj3, black).
color(obj4, black). %% цвет светофора
top(obj1, obj2). %% красная лампочка выше
желтой
top(obj2, obj3). %% а желтая выше зеленой
partOf(obj1, obj4). %% все лампочки принад-
лежат светофору
partOf(obj2, obj4).
partOf(obj3, obj4).
...

```

Здесь obj1, ..., obj4 – это термы в языке Пролог, которые одновременно являются метками областей в модуле обработки изображений и порождаются этим модулем в процессе обработки текущего изображения. Термы red и black в данном примере заданы априорно, как и двуместные предикаты color, top, partOf. Введение в базу правила вида color(obj1, red) означает, что предикат color следует считать истинным, когда его аргументы равны obj1 и red соответственно.

То, что данные накапливаются автоматически – несомненный плюс, однако сторонним эффектом этого является вероятная избыточность части данных в контексте решаемой задачи, как, например, цвет светофора в этом примере. Теперь алгоритм определения состояния светофора может оперировать высокоуровневыми понятиями, такими как объекты и их свойства. Данный пример этого алгоритма на языке Пролог будет выглядеть следующим образом:

```

trafficStop(X) :- partOf(A, X), %% найти три
лампочки светофора
partOf(B, X),
partOf(C, X),
top(A, B), %% при условии, что A – это верх-
няя лампочка
top(A, C),
color(A, red). %% проверить состояние верх-
ней лампочки

```

Здесь истинность предиката trafficStop для некоторой области изображения задается через значения истинности других предикатов. Примененный к объекту светофора данный предикат будет возвращать значение “истина” в случае, если его верхняя лампочка горит красным. На основе этого шаблона несложно написать предикаты для определения остальных двух состояний светофора, да и вообще реализацию проверки различных состояний на дороге, несмотря на то, что изначально накапливаемые знания не имели какой-либо специализированной направленности. Это свидетельствует о

возможности использования общих алгоритмов работы с низкоуровневыми изображениями и преобразования их в высокоуровневые, а затем подключения модуля с описаниями алгоритмов для решения узкого круга задач, что удовлетворяет поставленной цели.

Далее проведем анализ реализации системы перцептивных знаний на примере простой системы интерпретации искусственных изображений. В качестве основных перцептивных “понятий” используем предикаты, отвечающие за цвет области на изображении, ее форму и положение (слева, справа, выше, ниже) относительно других областей.

Эксперимент

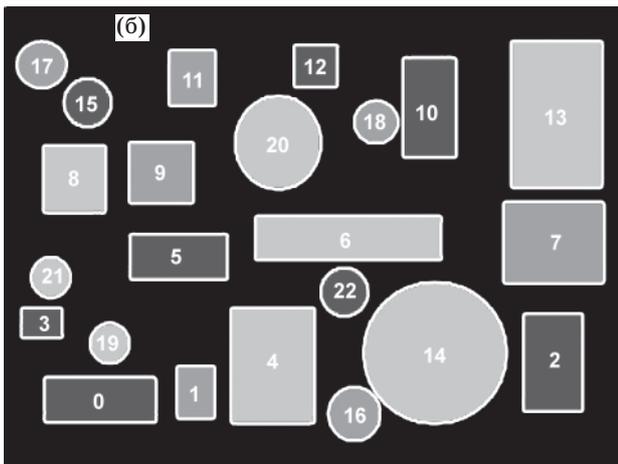
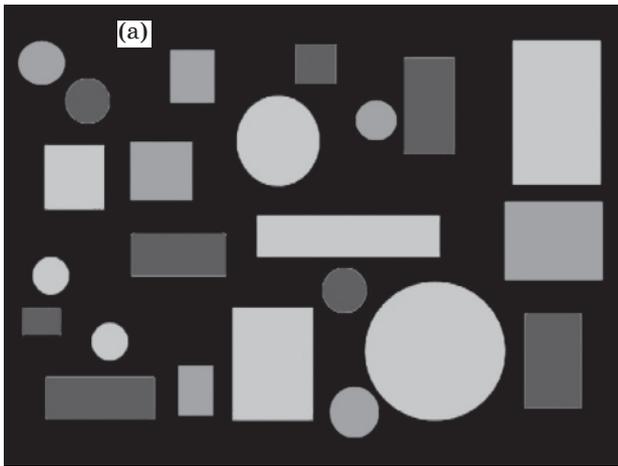
Для исследования описанного подхода был создан программный модуль, на вход которого подается изображение набора кругов и прямоугольников различных цветов на однотонном фоне (рис. 1а).

В первую очередь происходит сегментация, в результате чего формируется массив из объектов, каждый из которых характеризуется такими свойствами, как цвет, форма и набор ближайших фигур (рис. 1б).

Затем формируется база фактов для интерпретатора баз знаний (рис. 1в), после чего пользователю предлагается ввести критерии для выборки объектов на рисунке посредством сложных запросов к интерпретатору языка Пролог. Помимо априорных знаний о ближайших объектах заданы предикаты коммутативности положения (А левее В и В левее С, следовательно, А левее С и т. д.), на основе которых можно определить взаимное положение любых объектов относительно друг друга. Благодаря этому появляется возможность выполнения довольно сложных запросов, например, “выделить все красные прямоугольники выше синего круга”.

Программный модуль был реализован в программной среде С++ с подключением библиотек SWI-Prolog для реализации логического вывода и OpenCV для операций с изображениями.

Объектом интереса в данной работе является организация обратной связи между декларативной и процедурной компонентами модуля. В качестве процедурной компоненты выступает алгоритм обработки изображений, порождающий базу знаний, а затем вносящий в нее изменения, а в качестве декларативной компоненты – механизм логического вывода из существующей



(в) 119 right(obj20,obj4).
 120 left(obj20,obj12).
 121 higher(obj20,obj9).
 122 lower(obj20,obj18).
 123 color(obj21,red).
 124 shape(obj21,circle).
 125 right(obj21,obj17).
 126 left(obj21,obj8).
 127 higher(obj21,obj22).
 128 lower(obj21,obj5).
 129 color(obj22,blue).
 130 shape(obj22,circle).
 131 right(obj22,obj12).
 132 left(obj22,obj8).
 133 higher(obj22,obj3).
 134 lower(obj22,obj21).
 135
 136 left(X, Y) :- left(X, Z), left(Z, Y).
 137 right(X, Y) :- right(X, Z), right(Z, Y).
 138 higher(X, Y) :- higher(X, Z), higher(Z, Y).
 139 lower(X, Y) :- lower(X, Z), lower(Z, Y).

Исходное изображение – а, сегментированное изображение – б, пример базы фактов – в.

базы. Таким образом, данный программный модуль обладает плюсами обоих подходов.

Результаты тестирования

Результаты тестирования модуля показали, что механизм логического вывода, встроенный в интерпретатор SWI-Prolog, оказался не способен

выводить истинность даже несложных предикатов при введении правил коммутативности положения, ссылаясь на малый размер стека. Так предикат вида “test(X):- color(Y, red), shape(Y, rect), left(Y, obj15), right(X,Y).” не давал значения “истина” для областей, удовлетворявших этому условию даже при размере стека в 512 Мб.

К составлению запросов и априорных предикатов, как уже было замечено ранее, следует подходить с особым вниманием и осторожностью, чтобы в создаваемой системе правил учитывались неявные предположения, вносимые интерпретатором, для устранения возможности алгоритмически неразрешимых запросов. Например, такое очевидное правило как “left(X, Y) :- right(Y, X).” может привести к непредсказуемому результату, хотя и не всегда явно ошибочному, что усложняет процесс отладки.

Имеет смысл пересмотреть определение таких свойств предметов, как их взаимное расположение. В данный момент оно определяется по положению их центров, однако это не совсем верно с точки зрения человека, поскольку сдвиг на несколько миллиметров центра одного объекта относительно центра другого не делает его расположенным слева от исходного; это особенно сильно заметно на объектах, существенно различающихся по размеру. Возможно, в данном случае будет оправдано использование нечеткой логики, что потребует пересмотра подхода к организации хранения свойств о взаимном расположении предметов.

Заключение

Подводя итоги, можно сказать, что предложенный в статье подход показал свою состоятельность в ходе экспериментов с тестовым модулем на заданном наборе изображений, однако имеется место для усовершенствования и оптимизации. Задачей первостепенной важности является разработка многофункционального модуля для проведения операций сегментирования и создания высокоуровневого описания изображений, представленных в низкоуровневом виде. Из этого вытекает и необходимость выделения списка необходимых свойств, достаточно полно характеризующих заданный объект и устанавливающих связи между объектами. Определение каждого свойства потребует достаточно сложных алгоритмов, поскольку даже, казалось бы, тривиальная задача определения цвета предмета осложняется тем, что объекты могут иметь разную текстуру или альбедо, в результате чего

на некоторых кадрах бликующий черный объект будет выглядеть светлее, чем белый объект в тени на других кадрах. Кроме того, требуется более мощная и гибкая система логического вывода, которая могла бы позволить динамическую работу с базами фактов, а также вывод из обширных баз с помощью сложных предикатов. Полезным может быть выведение алгоритма, по которому будут синтезироваться правила для часто повторяющихся фактов. На основе этого могут быть выведены факты, например, “если предмет расположен на потолке, то, скорее всего, это люстра”, или некие общие правила, например, коммутативности положения. Еще одной важной задачей можно считать создание модуля, осуществляющего формирование предикатов на основе синтаксических конструкций, что откроет большие перспективы по использованию подобной системы в тех системах, где оператором будет пользователь, далекий от программирования или даже компьютеров вообще.

Построение систем перцептивных знаний в результате обучения является перспективным в области компьютерного зрения, поскольку позволяет существенно повысить гибкость этих систем и сгладить проблему “семантической пропасти”. В настоящее время подобные исследования ведутся [6], однако в них, как правило, не используются системы декларативного представления знаний. Таким образом, актуальной представляется дальнейшая разработка пред-

ставленного в настоящей работе подхода с автоматическим формированием общих правил в процессе обучения.

ЛИТЕРАТУРА

1. Helmer S., Meger D., Viswanathan P., McCann S., Dockrey M., Fazli P., Southey T., Muja M., Joya M., Little J., Lowe D., Mackworth A. Semantic Robot Vision Challenge: Current State and Future Directions // IJCAI-09 Workshop on Competitions in Artificial Intelligence and Robotics. Pasadena, California, USA. July 11–13, 2009. 7 p.
2. Rares A., Reinders M.J.T., Hendriks E.A. Image Interpretation Systems // Technical Report (MCCWS 2.1.1.3.C), MCCWS project, Information and Communication Theory Group. TU Delft. 1999. 32 p.
3. Liedtke C.-E., Grau O., Groue S. Use of explicit knowledge for the reconstruction of 3-D object geometry // Inter. Conf. on Computer Analysis of Images and Patterns. Prague, Czech Republic. September 1995. P. 580–587.
4. Groue S., Tonjes R. A Knowledge Based Approach to Automatic Image Registration // Proc. Inter. Conf. on Image Processing. Washington, DC, USA. October 26–29, 1997. V. 3. P. 228–231.
5. Поманов А.С. Технологии искусственного интеллекта: учеб. пособие / СПб: СПбГУ ИТМО, 2010. 218 с.
6. Deb Roy. Learning Visually-Grounded Words and Syntax for a Scene Description Task // Computer Speech and Language. 2002. V. 16. № 3–4. P. 353–385.