

УДК 004.896:621.865

ГРАФЫ КАК ИНСТРУМЕНТ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ СЦЕН

А.Д. Московский (*moscowskyad@gmail.com*)
НИЦ «Курчатовский институт»

Аннотация. В работе рассматривается метод распознавания сцен как часть общей системы технического зрения. В качестве инструмента для решения данной задачи предлагается использовать графовый механизм, т.к. данное описание естественно для представления сцен. Также в работе применен метод недоопределенных моделей для разрешения ситуаций с неполными входными данными.¹

Ключевые слова: распознавание, техническое зрение, реконструкция сцен, недоопределенные модели, доопределение атрибутов.

Введение

В работе рассматривается вопрос распознавания и анализа сцен. Распознавание сцен является классической задачей искусственного интеллекта, и является крайне актуальной для робототехники, прежде всего - мобильной. Под анализом трехмерной сцены (по соответствующему ей двухмерному изображению) изначально понималась задача выделения на изображении областей, принадлежащим разным объектам и фону, а также разбиение сложных объектов на составные части. Одними из первых работ на эту тему были [Roberts, 1965], [Guzman, 1967] и [Huffman, 1971]. В основном это были первые шаги по объединению простых объектов в более сложные структуры. Важным этапом стала работа [Marvin Minsky, 1975], где помимо самих объектов рассматривались отношения между ними. Такое представление было введено для решения задачи распознавания сцен субъектом, находящимся в движении. С тех пор в литературе обычно под понятием сцена подразумевается набор объектов и набор отношений между объектами [Карпов, 2011]. Несмотря на развитие методов получения трехмерных данных по видеокамере с использованием стереозрения [Scharstein, 2001] или structure-from-motion [Jebara, 1999], а так же появления таких датчиков как Kinect, распознавание сцен является по-

¹ Работа выполнена при частичной поддержке гранта РФФИ 15-07-07483

прежнему актуальной задачей. Из интересных работ можно привести метод [Brown, 2011], использующий алгоритм SIFT для задачи классификации сцен. Также распознавание типов сцен по характерным признакам рассматривается в работе [Parizi, 2012], а алгоритм, позволяющий работать с большими объемами распознаваемых паттернов сцен представлен в [Zhou, 2014]. Одной из проблем распознавания сцен, особенно методами получения трехмерных данных, является слабая возможность работы с динамическими сценами. Существует несколько исследований, описывающих работу со сценами, изменяющимися во времени, см. например [Yu, 2011]. Однако в большинстве работ этот вопрос остается не рассмотренным, в то время как данный критерий является крайне важным в задачах мобильной робототехники.

В настоящей работе предлагается метод распознавания сцен, с применением элементов теории графов и механизма недоопределённых моделей. При этом будет предполагаться, что задача поиска отдельных объектов, из которых формируются сцены, является решенной, так, как это описано в [Московский, 2016].

1. Постановка задачи

Пусть имеются данные от системы распознавания, представляющие собой информацию о найденных объектах. Также есть начальные условия, сообщающие системе описания всех сцен, которые требуется распознать. Как говорилось выше, описание сцены представляет собой набор объектов и набор отношений между ними. Отношения подразумеваются в первую очередь пространственные («быть над», «быть под», «быть на заданном расстоянии») и сравнительные («быть больше», «быть меньше»). Задать сцену можно посредством матрицы, где номера строк и столбцов соответствуют объектам в базе, а элементы матрицы соответствуют отношениям (при условии, что объекты идентифицированы и упорядочены).

| | O ₁ | O ₂ | O ₃ | O ₄ |
|----------------|----------------|----------------|----------------|----------------|
| O ₁ | - | R ₂ | R ₆ | R ₁ |
| O ₂ | R ₂ | - | R ₃ | R ₅ |
| O ₃ | R ₆ | R ₃ | - | R ₄ |
| O ₄ | R ₁ | R ₅ | R ₄ | - |

Рис. 1. Матричное описание сцены

Задачей является сопоставить текущий наблюдаемый набор с одной из заранее заданных сцен. С одной стороны существует ряд способов решения

данной задачи, с другой с учетом особенностей мобильной робототехники, задача приобретает ряд сложностей, связанных с проблемами визуальных сенсоров и помех создаваемыми окружающей средой. К этим помехам можно отнести заслонение объектов, неполную освещённость, неудачный ракурс и различные шумы. В связи с этим информация об объектах собирается частичная, не полная, но разрешаться эта проблема может различными путями, например, с помощью недоопределённых моделей, введённых Нариньяни [A.S.Narin'yanı, 1997]. Предложенный в [Московский, 2016] метод позволяет строить предположения о нахождении объектов на изображении, когда входная информация неполная или искаженная. Объекты, поступающие на вход алгоритма, помимо своих естественных характеристик (положение на изображении, тип объекта) содержат еще и коэффициент уверенности, сообщающий насколько точно объект был обнаружен в терминах системы. Поскольку информация об объектах может подаваться также в неполном виде, то метод распознавания сцен требуется также снабдить механизмом недоопределенных моделей, что бы повысить его эффективность.

2. Описание подхода

2.1 Описание сцены

Формализуем описание сцены. Сцена задается набором объектов

$$O(t) = \{o_1, o_2, \dots\} \quad (1)$$

и набором отношений

$$R(t) = \{r_1^2, r_i^j, \dots\} \quad (2)$$

Сцены, которые представлены в базе, и те которые требуется распознать, будем называть эталонными сценами. Помимо матричной формы задания сцен, удобно представлять сцену в виде графа, где вершинами являются объекты, а ребрами – отношения.

$$G = (O, R) \quad (3)$$

По каждой эталонной сцене строится свой граф G_{ref} . Этот граф является неориентированным и обязан быть связным.

2.2 Распознавание сцен

На каждом шаге алгоритма на вход поступает ряд распознанных объектов, а также отношений, регистрируемых между ними. Назовем это текущей сценой. Текущая наблюдаемая сцена представляется также в виде графа G_{cur} путем добавления всех объектов и отношений на него. Таким образом, мы получаем граф, содержащий полную информацию о

наблюдаемой сцене. Далее требуется распознать, присутствуют ли среди данного графа какие-либо подграфы, соответствующие эталонным сценам. Это все сводится к естественной задаче поиска подграфа. Однако в силу определенной выше специфики задачи, информация может быть неполная, т.е. могут быть неправильно определены или вообще отсутствовать некоторые объекты и отношения на данной сцене. Также из-за того, что объекты на изображении естественным образом разрознены, то граф текущей сцены не обязан быть связным. Поэтому прежде чем решать задачу поиска подграфа, требуется произвести предобработку графа.

Первый шаг предварительной обработки графа заключается в разбиении графа на связные подграфы.

$$G_{ref} \rightarrow \{G_1, G_2, G_3, \dots, G_k\} \quad (4)$$

Данная процедура нужна, чтобы определить базовые структуры сцены, от которых можно оттолкнуться при дальнейшем анализе. Далее каждый из связных подграфов требуется сравнить с графами эталонных сцен. Связный подграф может иметь совершенно произвольное количество элементов, зависящее от условий наблюдения и наблюдаемой сцены, в то время как граф эталонной сцены всегда фиксирован. Перед тем как определять похожесть, надо учесть все возможные варианты, которые может представлять связный подграф текущей сцены. Предполагается, что на текущей сцене обычно находится больше объектов одного типа, чем должно быть, поэтому будем искать эталонный граф как подграф текущего. Однако на текущем графе могут и отсутствовать часть объектов, составляющих эталонную сцену. В связи с этим создается временный граф (для каждого связного подграфа текущей сцены), который получается путем удаления с эталонного графа вершин объектов, не представленных на текущем связном подграфе. Соответственно удаляются и все связные с этими вершинами ребра.

$$G_{temp} = G_i \setminus \{o_p \dots o_k\} \quad (5)$$

Эта операция гарантирует, что текущий граф будет содержать меньшее или равное связному подграфу текущей сцены количество вершин. После того как временный граф подготовлен, можно выполнить его поиск на соответствующем ему связном подграфе. Это задача поиска изоморфизма подграфов. Среди многих решений данной задачи, был выбран алгоритм VF2 и его реализация на языке C++, представленная в библиотеке `igraph` [Документация `igraph`, 2017]. Алгоритм VF2 возвращает карту соответствий вершин временного графа и связного подграфа текущей сцены. По этой карте можно отобразить объекты, формирующие эталонную сцену на текущей сцене.

$$VF2(G_i, G_{temp}) = G_{wanted} \quad (6)$$

Описанный процесс позволяет подготовить подграф для задачи поиска изоморфизма подграфов в случае потерянных объектов. Если отношения потеряны, то данная схема не даст положительных результатов. Чтобы решить эту проблему, предлагается преобразовать эталонный граф и граф текущей сцены в их реберное представление. Это представление меняет ребра и грани местами. Имея такое представление можно применить описанную выше процедуру.

Если изоморфный подграф был найден, и на временном графе не пришлось удалять объекты, то задача распознавания сцены была бы решена. Однако в случае с отсутствием некоторых объектов требуется проводить дальнейшую обработку результатов алгоритма поиска изоморфных подграфов.

2.3 Доопределение элементов сцены

Наиболее интересен случай, когда не было распознано несколько отношений, но сами объекты присутствуют на текущей сцене (предположим, из общей картины выпало столько отношений, что один объект не попал на связный подграф). В таком случае механизм, описанный выше, возвращает те объекты на изображении, которые соответствуют эталонной сцене, но без потерянного объекта. Далее выполняется поиск данного объекта, и тут возникает та самая неопределенность. Изолированных от сцены объектов может быть несколько, и каждый потенциально может являться частью сцены, но поскольку отношения для него распознано не было, то нельзя определить какой именно (или же ни одного из представленных). Для статических задач распознавания на отдельном изображении эта задача не очень простая и требует дополнительной входной информации, например, контекста. Однако в данной работе предполагается использование метода в робототехнических системах, что позволяет говорить о распознавании на последовательности изображений. Также как в случае распознавания объектов, предлагается прибегнуть к механизму недоопределённых моделей (n-моделей) [Нариньяни, 1998]. Напомним, что в случае с объектами, система отслеживает их перемещение от кадра к кадру и, таким образом, позволяет определить соответствия между объектами на разных стадиях наблюдения.

Обобщенная вычислительная модель данной методологии задается четырьмя множествами:

$$M=(V,W,C,R) \quad (7)$$

где V – множество всех объектов, W - множество функций присваивания, C - множество функций проверки корректности, R - множество ограничений. Объектами множества V в рассматриваемом случае являются объекты,

получаемые от системы распознавания. Функциями присваивания W выступает в данном случае механизм, относящий объекты к эталонным сценам. Функции S выполняют проверку корректности выбранных изолированных объектов в соответствии с ограничениями R . Смысл представленной вычислительной модели - в итерационном применении данных функций к объектам. Функции и ограничения должны задаваться таким образом, что бы итерационное применение как минимум не расширяло область неопределенности, однако сходимость метода не гарантируется.

Ограничения заданы таким образом, что в сцену попадает только тот изолированный объект, который уже был к ней причастен на предыдущих шагах распознавания. Этот подход позволяет поддерживать представление о сцене в случае ухудшения условий наблюдения. Этот же подход работает аналогичным образом, если изолированных объектов несколько в сцене.

Также интересен случай, когда нераспознанные отношения разбивают сцену на две части, и каждая часть при этом выступает как распознанная часть эталонной сцены. Данный случай разрешается также как с изолированными объектами, только добавляется проверка на не повторение одних и тех же объектов в общей сцене.

После того, как все распознающие процедуры были произведены, для каждой сцены требуется рассчитать коэффициент доверия, подобно тому, как это делается для объектов.

$$CC_S = \frac{1}{N_{ratio}} \left(\sum_{ratio} \frac{1}{2} (CC_{Oi} + CC_{Oj}) + \sum_{isolated} \frac{CC_{Ok}}{N_{con.ratio} + 1} \right) \quad (8)$$

Первое слагаемое коэффициента уверенности представляет собой сумму по всем распознанным отношениям полусумм коэффициентов уверенности связанных отношением объектов к общему числу отношений. Второе слагаемое - это сумма по всем изолированным объектам их коэффициентов уверенности к количеству исходящих отношений из данного объекта плюс один. Вся сумма делится на общее число отношений в сцене. Если все отношения и объекты были распознаны, то коэффициент уверенности становится равным единице.

Общая схема, описанного в работе метода представлена на рисунке 2.



Рис. 2. Общая схема предложенного метода распознавания сцен

3. Эксперименты

Исследование работоспособности данного метода производилось следующим образом. Генерировался граф с заданным количеством вершин и случайными связями, это была эталонная сцена. Далее этот граф подвергался различному искажению: случайно удалялись или добавлялись вершины и грани. Полученный граф играл роль наблюдаемой сцены. Далее для полученных графов применялся описанный в работе метод. На рис. 3 представлена зависимость коэффициента доверия (8) от искажений, внесенных в граф.

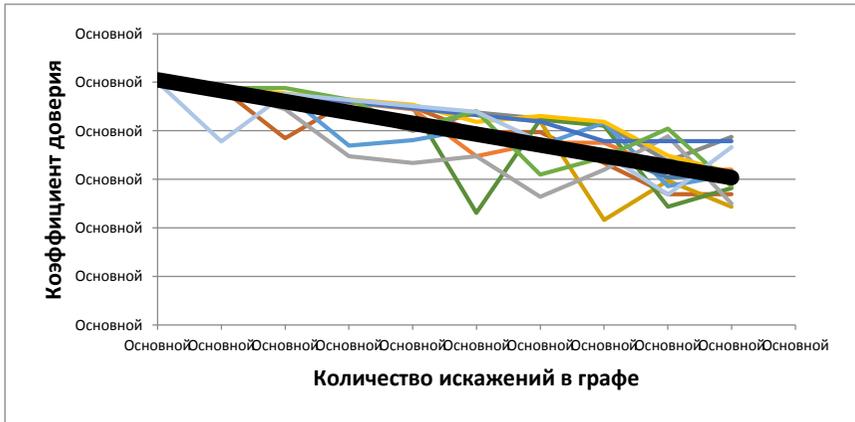


Рис. 3. Зависимость коэффициента уверенности от количества искажений для набора графов с 20-25 вершинами, черной жирной линией указана линейная аппроксимация

Из графика видно, что метод достаточно стабильно справляется с распознаванием, даже когда граф отличается почти на половину элементов (коэффициент доверия в районе 0.5).

Заключение

Был разработан метод распознавания сцен для мобильных робототехнических систем, использующий графовый механизм и H-модели. Проведенные эксперименты показали не только работоспособность подхода, но также достаточную скорость работы алгоритма для использования в бортовых системах мобильных роботов. При этом приемлемая скорость обработки (2-5 Гц) достигается на выборках, содержащих не более ста объектов на текущей сцене.

Список литературы

- [Карпов, 2011] Карпов В.Э. Об одном механизме реконструкции схем // VI-ая Международная научно-практическая конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте», Сб. науч. трудов. В 2-т., Т.1, М.: Физматлит. , 2011.
- [Московский, 2016] Московский А.Д. Об одном методе распознавания объектов с не полностью определенными признаками // Всероссийский научно-практический семинар «Беспилотные транспортные средства с элементами искусственного интеллекта» (БТС-ИИ-2016), Казань. 2016. С. 137–146.
- [Нариньяни, 1998] Нариньяни, А.С., Телерман, В.В., Ушаков, Д.М., Швецов, И.Е. Программирование в ограничениях и недоопределенные модели // Информационные технологии. 1998. Т. 7.
- [A.S.Narin'yani, 1997] A.S.Narin'yani, S.B.Borde, D.A.Ivanov. Sub-Definite Mathematics and Novel Scheduling Technology Programs // Artif. Intell. Eng. 1997. Т. 11.
- [Brown, 2011] Brown M., Susstrunk S. Multi-spectral SIFT for scene category recognition // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. , 2011. С. 177–184.
- [Guzman, 1967] Guzman A. Some Aspects of Pattern Recognition by Computer // MAC-TR-37. 1967.
- [Huffman, 1971] Huffman D. Impossible Objects as Nonsense Sentences // Mach. Intell. 6. 1971.
- [Jebara, 1999] Jebara T., Azarbayejani A., Pentland A. 3D structure from 2D motion // IEEE Signal Process. Mag. 1999. Т. 16. № 3. С. 66–84.
- [Marvin Minsky, 1975] Marvin Minsky. A Framework for Representing Knowledge // Psychol. Comput. Vis. 1975.
- [Parizi, 2012] Parizi S.N., Oberlin J.G., Felzenszwalb P.F. Reconfigurable models for scene recognition // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. , 2012. С. 2775–2782.

- [**Roberts, 1965**] Roberts L. Machine Perception of Three Dimensional Solids / под ред. Т. J. Cambridge: , 1965. Вып. The MIT Pr.
- [**Scharstein, 2001**] Scharstein D., Szeliski R., Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // Proceedings - IEEE Workshop on Stereo and Multi-Baseline Vision, SMBV 2001. , 2001. С. 131–140.
- [**Yu, 2011**] Yu, X., Fermuller, C., Teo, C. L., Yang, Y., & Aloimonos, Y. (2011). Active scene recognition with vision and language. In 2011 International Conference on Computer Vision, ICCV 2011. Pp. 810-817.
- [**Zhou, 2014**] Zhou B., Khosla A., Lapedriza A. Learning Deep Features for Scene Recognition using Places Database // Adv. Neural Inf. Process. Syst. 27. 2014. С. 487–495.
- [**Документация igraph, 2017**] Документация igraph [Электронный ресурс]. URL: <http://igraph.org/c/#docs> (дата обращения: 14.05.2017).