УДК 004.41

АДАПТИВНАЯ СИСТЕМА КОНТРОЛЯ И КЛАСТЕРИЗАЦИИ ВИЗУАЛЬНЫХ ДАННЫХ

М.Г. Казаков, Е.Н. Крючкова

В статье рассматривается адаптивная система классификации сложных изображений, основанная на использовании семантических связей между классами. Используется семантический граф понятий, имеющих достаточно постоянное визуальное представление. Для автоматического получения выборки обучающих изображений при настройке системы используются общедоступные поисковые системы, что позволяет автоматическим образом проводить поиск соответствующих изображений.

Ключевые слова: компьютерное зрение, классификация визуальных объектов, обучающая выборка, семантический граф, SIFT-дескрипторы, визуальные особенности.

Актуальность

Задача классификации изображений является одной из наиболее сложных задач компьютерного зрения [1]. Особую сложность задача кластеризации приобретает в условиях, когда система должна адаптироваться к новым категориям объектов, учитывать визуальную изменчивость распознаваемых объектов и семантическую связь между ними. Основу существующих на сегодняшний день методов классификации составляет использование визуальных особенности (visual features) изображений. К таким достаточно универсальным методам можно отнести метод Bag-of-Words (BoW) [2] и методы, основанные на использовании наивного байесовского классификатора [3, 4]. Однако в реальных задачах часто требуется не просто отнести входное изображение к определенному классу, но и локализовать экземпляр этого класса на изображении. Такая задача получила название категоризации через локализацию (categorization by localization) [5].

Алгоритмы решения задач категоризации, как правило, основаны на использовании специальных библиотек изображений. Главным недостатком таких методов является то, что подобные библиотеки построены вручную, что существенно усложняет их расширение. Кроме того, при обработке сложных изображений, на которых представлены объекты разных классов, каждый отдельный объект приходится обрабатывать независимо от других, поскольку библиотечные классы существуют независимо друг от друга. В данной работе предлагается пересмотреть постановку задачи и взять за основу не построенные вручную библиотеки классов изображений, а знания человека о реальном мире, представленные в виде семантического графа. Кроме того, предлагаемая схема основана на использовании общедоступных поисковых систем, что исключает применение ручного труда при построении обучающей выборки. Таким образом, достигается возможность гибкой подстройки на использующиеся типы изображений, решается проблема связности изображений в классах, основанная на представлении человека о мире. Данный подход позволяет строить обучающие выборки, в которых отсутствует информация о расположениях объектов на изображениях, следовательно, ее можно использовать для поиска типов представленных на изображении классов, что по сравнению с традиционным методом решения задачи категоризации более соответствует схеме анализа изображения человеком. Более того, использование связей между классами, имеющимися в обучающей выборке, позволяет анализировать сложные изображения, в которых могут быть представлены экземпляры разных семантически связанных классов. Структура предлагаемой системы приведена на рисунке 1.

Вычисление дескрипторов локальных особенностей изображений

Визуальные особенности изображений в нашей работе представлены SIFT-дескрипторами [6], вычисление которых начинается с представления изображения в виде гауссовой пирамиды. Внутри каждой октавы находятся экстремумы разницы гауссиан $D(x,y,\sigma)$, являющиеся потенциальными центрами локальных особенностей. Эти позиции уточняются, используя разложение в ряд Тейлора разницы гауссиан $D(x,y,\sigma)$ совмещая начало координат с найденной точкой:

$$D(X) = D + \frac{\partial D^{\mathrm{T}}}{\partial X} X + \frac{1}{2} X^{\mathrm{T}} \frac{\partial^2 D}{\partial X^2} X \quad (1)$$

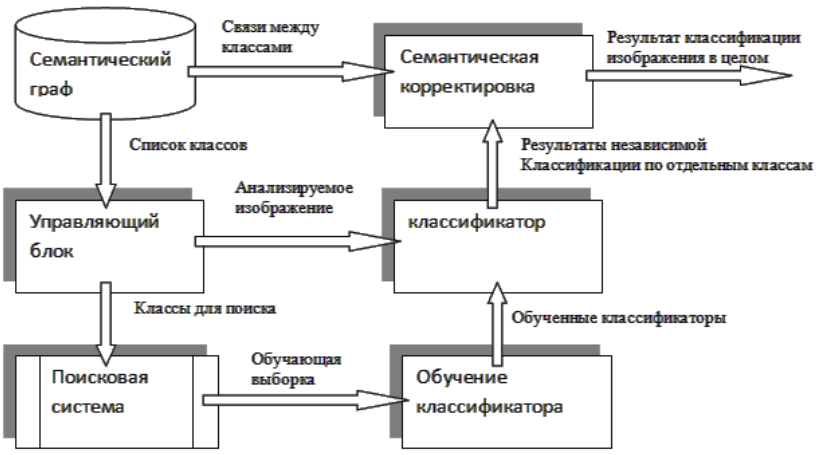

Рисунок 1 – Структура адаптивной системы

Экстремум *X'* вычисляется, приравнивая нулю производную функции *D(X)*:

$$X' = -\frac{\partial^2 D}{\partial X^2}^{-1} \frac{\partial D}{\partial X} \quad (2)$$

Для фильтрации неконтрастных точек *X'*, все точки в которых $|D(X')| < 0.03$ отбрасываются. Помимо малоконтрастных точек отбрасываются малоустойчивые точки, располагающиеся вдоль границ, для чего вычисляется гессиан функции $D(x,y,\sigma)$ на полученном ранее уровне σ и вводится ограничение на него:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \quad \frac{Tr(H)}{Det(H)} < 12 \quad (3)$$

На следующем этапе вычисляется ориентация локальной особенности, что позволяет добиться инвариантности от поворотов изображения. В окрестности *X'* на текущем уровне σ гауссовой пирамиды для каждой точки $L(x,y)$ считаются магнитуды $m(x,y)$ и ориентации $\Phi(x,y)$ градиентов. После составления гистограммы ориентаций градиентов, применяя магнитуду как весовой коэффициент, пиковые значения принимаются за ориентации региона.

При вычислении дескриптора региона, он разбивается на сетку 4x4, учитывая его размер и ориентацию. Для каждой ячейки сетки составляется гистограмма ориентаций градиентов точек из 8 диапазонов таким образом, чтобы каждая точка распределялась по четырем смежным ячейкам исходя из ее расположения. В результате каждая локальная визуальная особенность представляется 128-мерным вектором, а изображение в целом – множеством таких векторов.

Семантический граф

Вершины *V* семантического графа *G = (V, U)* представляют собой понятия, имеющие достаточно постоянное визуальное представление, а взвешенные дуги *U* определяют семантические связи между словами, причем вес дуги должен определять степень семантической близости между словами. Контекстные связи между объектами можно использовать в качестве отдельных этапов при структурном анализе сложного визуального объекта. При таком подходе любое простое или сложное отношение рассматривается как реализация обобщенного отношения, заданного на упорядоченном множестве понятий. Очевидно, что тем дальше друг от друга в графе *G* находятся слова, тем менее они связаны, а, следовательно, и менее связаны между собой соответствующие классифицируемые фрагменты изображения.

Эксперименты показали, что большое количество лингвистической информации затрудняет использование, в частности, попытка использовать граф WordNet [7] привела к построению чрезмерно широкой семантической окрестности анализируемых понятий, причем даже использование большого коэффициента затухания при обходе графа не позволяло выделить сильно связанную семантическую окрестность. В данной работе используется граф, автоматически построенный на основе толкового словаря и словаря синонимов [8], содержащий связи определения, ассоциации и синонимии. Отношение определения связывает слово с обобщающим его понятием. Ассоциативные связи в графе *G* являются признаком использования соответствующих понятий в одной словарной

статье толкового словаря. Связи синонимии являются точным отображением определений из словаря синонимов, однако в нашей задаче они рассматриваются как контекстные ассоциативные связи. Именно контекстные ассоциации играют определяющую роль в распознавании классов сложных изображений, когда два объекта встречаются вместе в одном контексте.

Для каждой вершины $x \in V$ графа G можно ввести в рассмотрение окрестность $\delta(x)$, в которой расстояние от x до любой вершины множества $\delta(x)$ не превышает заданной величины. На выбор соотношения между весами связей определения и ассоциативных связей повлиял тот факт, если больший вес имеют ассоциативные связи, то в $\delta(x)$ оказываются контекстные ассоциации, что в полной мере соответствует задаче выделения семантически связанных объектов на изображении.

Автоматическое формирование обучающей выборки в адаптивной системе

Формирование обучающей выборки нами построено на основе использования Google Custom Search. В результате исключается ручной труд по ее формированию, а набор необходимых классов может быть гибко подстроен под каждую конкретную задачу. Следует отметить, что при таком создании обучающей выборки у системы нет информации о расположении объектов внутри изображений, поэтому выборка не может быть использована для решения задач категоризации через локализацию, однако она вполне удовлетворяет цели семантической кластеризации визуальных данных.

Классификатор и обработка обучающей выборки

Автоматически полученная обучающая выборка предъявляет высокие требования к устойчивости результата работы классификатора при незначительных изменениях обучающей выборки – при появлении в выборке «плохих» экземпляров общий результат работы не должен кардинально ухудшаться. Был проведен ряд экспериментов по применению Bag-of-Words и наивного байесовского классификатора. Наивный байесовский классификатор показал способность работать с объектами, представленными частично. Однако, он неустойчив при изменениях обучающей выборки, и предъявляет повышенные требования к однотипности обучающей выборки для классов с разным количеством извлекаемых визуальных особенностей, при незначительном изменении выборки для одного класса с «плохими» данными способен существенно ухудшить общую работу системы. Эксперименты показали, что этот классификатор является непригодным для работы с автоматически полученной обучающей выборкой. Метод Bag-of-Words показал устойчивость результатов при изменениях обучающей выборки, хотя при обработке изображений с несколькими экземплярами объекта качество классификатора существенно падает.

Каждый класс представляется набором гистограмм, каждая из которых описывает изображения класса в терминах распределения локальных особенностей по визуальным словам. Данные, представленные в таблице 1, демонстрируют хорошее качество автоматически построенной обучающей выборки, что позволяет использовать данный метод для классификации.

Таблица 1 –Пример классификации на основе автоматически построенной обучающей выборки

Класс	Автомобиль	Самолет	Колесо	Мост	Окно	Дом
Автомобиль	100%	20%	33%	24%	25%	13%
Самолет	45%	100%	20%	0%	6%	0%
Колесо	36%	20%	100%	20%	12%	4%
Мост	9%	26%	33%	100%	12%	13%
Окно	0%	6%	13%	8%	100%	30%
Дом	36%	0%	20%	12%	50%	100%

Алгоритм классификации на основе семантического графа

Коэффициент близости найденных классификатором объектов *a* и *b* определяет их визуальное соответствие и вычисляется в процессе обхода семантического графа *G*:

$$S_b = S_a \times E(a,b) \times D, \quad S_a \geq T, \qquad (4)$$

где *E(a,b)* – функция сходства понятий *a* и *b* в семантическом графе, *D* – коэффициент демпфирования при каждом очередном удалении от изначального понятия, *T* – порог, превышение которого означает полное несоответствие анализируемых объектов.

Введем функцию W(a,x) ширины класса, которая равна сумме коэффициентов сходства всех понятий, достижимых из понятия *a* при заданном начальном коэффициенте близости S_a=*x*:

$$W(a,x) = \sum_{i=1}^{N} S_i^a \qquad (5)$$

Здесь S_i^a вычисляется для всех понятий в *G*, достижимых из *a*. В семантическом графе

понятия располагаются с неравномерной плотностью, поэтому значения близости нормируются значением *W(C, 1)*. Для получения сравнимых величин в терминах понятий с учетом ширины классов используется следующее преобразование:

$$Q_C = \frac{W\left(C, 1 - \frac{r_C - r_{\min}}{r_{\max} - r_{\min}} \times 0.5\right)}{W(C,1)}, \qquad (6)$$

где r_C – результат работы классификатора класса *C*, r_{min} – минимальный результат работы всех классификаторов, r_{max} – максимальный результат работы всех классификаторов.

В соответствии с преобразованием (6) более близкий в соответствии с результатом работы классификатора класс получит изначальный коэффициент семантической близости 1.0, а наименее близкий - значение 0.5. Вычисление коэффициентов семантической близости является основой для объединения понятий в кластеры при условии превышения некоторого порогового значения величины пересечения понятий, имеющих ненулевое значение коэффициента близости.

При объединении значениями близости понятий S_i^C в новом кластере $C(C_1,C_2)$ принимаются набольшие значения из оригинальных значений близости:

$$S_i^C = \max(S_i^{C_1}, S_i^{C_2})$$

Значение нормированной ширины Q_C объединенного кластера *C*, полученного из C_1 и C_2, вычисляется по формуле:

$$Q_C = M \times \frac{1 + M \times m^2}{M^3 + m^3}, \qquad (7)$$

где $M = \max(Q_{C_1}, Q_{C_2})$ и $m = \min(Q_{C_1}, Q_{C_2})$.

В качестве результата классификации выбирается кластер с максимальным значением нормированной ширины.

Выводы

На рисунке 2 представлен пример сложного изображения, анализ которого был выполнен по предлагаемой методике. Результаты работы после семантической корректировки представлены в таблице 2, в которой верхняя строка – названия классов, вторая - значения коэффициента близости, вычисленные от значений на начальном интервале [0.50; 1.00], нижняя – несмещенные значения коэффициента. Победитель – кластер «Автомобиль - Колесо», на втором месте кластер «Дверь - Окно», хотя без семантической корректировки с преимуществом 0.05 лидировал класс «Мост».

Таким образом, предложенная методика обработки сложных изображений позволяет обрабатывать сложные изображения с высокой степенью релевантности, исключить ручной труд при построении обучающих выборок.

Рисунок 2 – Пример анализируемого изображения

Таблица 2 –Пример классификации изображения

Стол	Автомобиль, Колесо	Самолет	Цветок	Мост	Небо	Кошка	Облако	Дверь, Окно	Дом
0.50	1.47	0.86	0.78	1.00	0.75	0.50	0.66	1.44	0.91
0.00	0.97	0.36	0.28	0.50	0.25	0.00	0.16	0.94	0.41

СПИСОК ЛИТЕРАТУРЫ

1. Szeliski, R. Computer Vision: Algorithms and Applications. / R. Szeliski. - Springer-Verlang New York, 2010.
2. Csurka, G. Visual categorization with bags of keypoints / G. Csurka [and others] // Workshop on statistical learning in computer vision, 2004.
3. Behmo, R. Towards optimal naive Bayes nearest neighbor / R. Behmo [and others] // European Conference on Computer Vision, 2010.
4. Lowe, D.G. Local naïve Bayes nearest neighbor for image classification / D.G. Lowe // Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012.
5. Harzallah, H. Combining efficient object localization and image classification / H. Harzallah, F. Jurie, C. Schmid // International Conference on Computer Vision, 2009.
6. Lowe, D.G. Distinctive image features from scale-invariant keypoints / D.G. Lowe// International Journal of Computer Vision, 2004. 60(2).
7. Fellbaum, C. WordNet: An Electronic Lexical Database. / C. Fellbaum - Cambridge, MA: MIT Press, 1998.
8. Крайванова, В.А. Построение взвешенного лексикона на основе лингвистических словарей / В.А. Крайванова, А.О. Кротова, Е.Н. Крючкова // Материалы Всероссийской конференции «Знания – Онтологии - Теории», 2011.

*Аспирант **Казаков М.Г.**, - mike.kazakov @gmail.com; к.ф.-м.н., проф. **Крючкова Е.Н.** – kruchkova_elena@mail.ru, тел. (3852) 290-868; - каф. прикладной математики Алтайского государственного технического университета им. И.И. Ползунова.*