

# ИКОНИКА – НАУКА ОБ ИЗОБРАЖЕНИИ

УДК 004.93'1:004.932.2

## ОБЪЕКТНО-НЕЗАВИСИМЫЙ СТРУКТУРНЫЙ АНАЛИЗ ИЗОБРАЖЕНИЙ: ИСТОРИЯ И СОВРЕМЕННЫЕ ПОДХОДЫ

© 2014 г. В. Р. Луцив<sup>\*,\*\*</sup>, доктор техн. наук; Р. О. Малашин<sup>\*,\*\*</sup>, аспирант

<sup>\*</sup>Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург;

<sup>\*\*</sup>Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург

E-mail: vluciv@mail.ru

Рассмотрена история возникновения и развития структурного анализа изображений. Для каждого из исторических этапов проанализированы характерные недостатки и преимущества применяемых структурных методов и предпосылки для их совершенствования. Материал завершается описанием особенностей организации и возможностей наиболее совершенных современных алгоритмов структурного анализа и рассмотрением направлений их дальнейшего улучшения.

**Ключевые слова:** сопоставление изображений, классификация изображений, структурное описание, структурный элемент, структурный анализ.

Коды OCIS: 100.2960, 100.3005, 100.3008, 100.5010, 110.2960, 150.1135.

Поступила в редакцию 23.06.2014.

### Введение

Автоматический анализ изображений, ведущий свою историю от первых моделей перцептронов и компьютерных экспериментов по автоматическому распознаванию зрительных образов, выполненных в середине двадцатого века, стал в настоящее время самостоятельным научным направлением, имеющим важные практические приложения в автоматизации производства и контроля качества, неразрушающей диагностике, медицине, криминалистике, системах безопасности, дистанционном зондировании Земли, системах наведения, робототехнике и многих других областях. Начало интенсивных исследований в области автоматического анализа изображений относится к середине прошлого века, когда уровень развития кибернетики и вычислительной техники позволил применить к изображениям хорошо уже исследованные в статистической теории связи методы узнавания сигналов. Такие методы, основанные на согласованной фильтрации и корреляции, были объектно-независимы (т.е. применимы для поиска произвольного сигнала

по соответствующему ему по форме эталону). Однако при переходе от сигналов – функций одной переменной (времени), к двумерным и трехмерным пространствам реальных сцен, методы эти оказались пространственно не инвариантными: они плохо справлялись с геометрическими трансформациями изображений, характерными для естественных и искусственных зрительных систем.

Патентная литература второй половины двадцатого века предлагает множество устройств и методов, реализующих такой подход (см., например, [1, 2], ссылки на более обширные обзоры таких методов и устройств можно найти в [3]). Корреляционный отклик, вычисленный в декартовой системе координат, достаточно быстро деградирует при взаимном вращении или масштабировании сопоставляемых изображений [4]. В связи с этим были предложены технические решения (см., например, классические работы [5–7]), делающие корреляторы инвариантными к преобразованиям группы подобия. Такие решения основывались на вычислении функции корреляции в полярно-логарифмической системе координат и инвари-

антности амплитуды гармоник спектра Фурье к сдвигу изображения вдоль декартовых координатных осей. При этом преобразование Фурье единообразно применялось ко всей площади изображения, в связи с чем в дальнейшем в настоящей работе вычисленные таким образом инвариантные характеристики упоминаются как интегральные.

В то же время начали разрабатываться и некорреляционные методы распознавания. Они также были заимствованы из статистической теории связи, но основывались на построении дискриминационных поверхностей в многомерных пространствах признаков. В ставшей классикой работе [8] приводится фундаментальное описание этой методики, предлагающей альтернативу ограниченным возможностям корреляционного подхода. Большинство предлагаемых на ранних этапах систем признаков разрабатывались для решения частных задач распознавания и использовали такие характеристики, как длины характерных сечений изображений объектов, точки совпадения контуров объекта и эталона, максимальный размер объекта и другие объектно-специфические параметры. В литературе второй половины двадцатого века можно найти множество таких предложений (см., например, [9–13], ссылки на более обширные обзоры таких источников приводятся в [3]). Некоторые из таких решений (см., например, [14, 15]) объектно-независимы и даже инвариантны к преобразованиям группы подобия. Инвариантные признаки такого типа, как предложены в [14, 15], вычисляются по всей площади изображения, поэтому они также будут упоминаться как интегральные.

В выше перечисленных методах основное внимание уделялось трудностям распознавания изображения, обусловленным его геометрическими трансформациями в применяемых оптических системах. Однако наблюдаемый мир также сам по себе изменяется сложным образом независимо от конструкции отображающей его оптической системы. Уже на первых шагах развития компьютерного зрения стало очевидно, что рассмотренные выше объектно-независимые инвариантные к пространственным преобразованиям интегральные методы распознавания, исходно разработанные для обнаружения сигнала стабильной формы, не достаточно эффективны при анализе сложным образом изменяющихся двумерных и трехмер-

ных образов, соответствующих изображениям реальных сцен. Эти методы имеют общий недостаток: они единообразно извлекают информацию из всей площади изображения, рассматривают изображение как единое целое и, как правило, применяют ко всему изображению одну и ту же модель геометрического преобразования системы пространственных координат. Если сопоставляемые изображения не включают достаточно большого числа общих элементов сцены (имеют недостаточную “площадь перекрытия”), если взаимное положение достаточно большого количества отдельных элементов сцены существенно меняется от изображения к изображению, или знаки взаимных контрастов изображений поверхностей достаточно большого числа объектов меняются на противоположные, интегральные методы сопоставления теряют работоспособность. В частности, корреляционный отклик деградирует под действием преобразований яркости пикселей, связанных с естественными изменениями природных ландшафтов и меняющимися условиями наблюдения [16]. В связи с этим возникла необходимость перехода к структурному сопоставлению, позволяющему анализировать изображения локальных элементов сцены по отдельности и сопоставлять затем с заданной точностью их взаимное положение.

### **Становление структурного подхода к анализу изображений**

Методы пространственной декомпозиции и структурного сопоставления начали применяться уже на ранних стадиях становления науки об автоматическом анализе изображений. В частности, “наложение масок” эталонов выполнялось для распознавания не всего изображения, а отдельных его характерных фрагментов (см., например, [17, 18]), которые могут быть более устойчивы к искажениям. Следующим шагом стал учет координат распознанных локальных фрагментов изображений для окончательного распознавания изображений в целом (см., например, [19, 20]). В патентной литературе второй половины двадцатого века можно найти большое количество упоминаний методов и устройств, основанных на распознавании отдельных фрагментов изображений и анализе их положения; ссылки на более обширные обзоры такой литературы также можно найти в [3]. Кстати, метод, основан-

ный на распознавании характерных локальных фрагментов, аналогичный предложенному в [19], и по сей день является ведущим подходом при построении систем автоматического распознавания отпечатков пальцев. Схожий подход положен в основу распознавания изображений знаменитой нейронной сетью NEOCOGNITRON [21].

Аналогичная тенденция может быть отмечена и в дискриминантных методах распознавания, выполняемого в пространствах признаков. Продемонстрируем это на примерах построения искусственных нейронных сетей, начиная с перцептрона Розенблатта [22], который, как показано в [8], осуществляет построение дискриминационных гиперповерхностей в пространстве признаков, причем в качестве признаков, фактически, выступает информация, поступающая от совокупности всех пикселей изображения, т.е. используется интегральный метод распознавания. На смену перцептрону Розенблатта пришли многослойные перцептроны, обучаемые методом обратного распространения ошибки [23, 24]. В силу недостаточной инвариантности дискриминантных методов к естественным изменениям изображений при непосредственном использовании яркости пикселей в качестве признаков в указанных нейронных сетях стали применять системы инвариантных признаков, основанных на инвариантных моментах, Фурье инвариантах, клино-кольцевых детекторах (см., например, [25–27]). Однако указанные инвариантные признаки также были интегральными, т.к. вычислялись по всей площади распознаваемого изображения, что ограничивало эффективность применяемых нейронных классификаторов. Уже в [26] делаются попытки преодолеть некоторые недостатки интегрального подхода: зрительное поле, по которому вычисляются инвариантные признаки, делится на подполя, чтобы в процессе распознавания присутствующие в большом поле зрения посторонние объекты не вносили “шумовую составляющую” в вычисленные интегральные признаки целевого объекта. В [28, 29] были сделаны попытки устранить недостатки интегрального подхода другим путем: в применяемой искусственной нейронной сети глобальная система связей между последовательными слоями нейронов была заменена совокупностью локально связанных подсистем, позволявших независимо анализировать локальные фрагменты образов.

Авторы [28, 29] отмечают, что такое усовершенствование оказалось плодотворным.

Принцип дополнения глобального анализа локальным достиг апофеоза в нейронных сетях, сконструированных Кунихико Фукусимой [30]. Эти нейронные сети продемонстрировали впечатляющие способности к распознаванию контурных изображений, подвергнутых сложным локальным трансформациям. Однако этот успех был достигнут слишком дорогой ценой: для каждого варианта локального смещения каждого распознаваемого фрагмента изображения был сконструирован отдельный нейронный детектор. В силу этого нейронная сеть достигла внушительных размеров, и дальнейшее ее усовершенствование, например, введение возможностей анализа изображений с глобальными геометрическими трансформациями, такими как аффинные или проективные преобразования было неприемлемо.

Приведенные выше материалы демонстрируют закономерное эволюционирование корреляционных и признаковых методов распознавания изображений в сторону применения структурных подходов. Однако параллельно происходила разработка систем компьютерного зрения, изначально основанных на структурных принципах анализа изображений, в связи с чем целесообразно отметить ряд успешных исследовательских проектов, описание которых было сведено воедино в двухтомнике, посвященном созданию реально функционирующих автономных роботов [32, 33]. Построенные в рамках этих проектов системы компьютерного зрения использовали структурную декомпозицию изображений, основанную не на строгой теории, а скорее на наборах эвристических решений, что тем не менее позволяло созданным роботам достаточно успешно ориентироваться в окружающей обстановке. Несколькими годами позже в печати появляется ряд фундаментальных работ, посвященных теоретическим аспектам построения структурных описаний и выполнения структурного распознавания (см., например, [31, 8, 34, 35]). В таких работах были представлены математические, формально-логические, лингвистические основы для выполнения структурного анализа и синтеза, хотя редко приводились сколько-нибудь ценные результаты практического применения разработанного теоретического аппарата. Можно сказать, что представленный в них формальный аппарат структурного анализа представ-

для более теоретическую, чем практическую ценность, не будучи в состоянии описать все многообразие свойств реальной действительности, хотя, несомненно, некоторые разработанные теоретические конструкции стали основой механизмов обработки знаний в появившихся вскоре экспертных системах (ЭС). Рассмотрим вкратце историю их появления.

### **Применение и развитие методов структурного анализа изображений в экспертных системах**

Не погрешив сильно против истины, можно утверждать, что появление ЭС во многом обязано созданию формально-логических и лингвистических основ представления и обработки знаний, упомянутых в предыдущем разделе, поэтому ЭС согласно общепринятому толкованию этого понятия содержат в качестве подсистемы механизм (engine) формально-логической обработки знаний, а также базу знаний, интерфейс пользователя и интерфейс разработчика, выполняющего наполнение базы знаний и отладку алгоритмов их обработки. Однако в реально созданных системах не всегда удается выделить все эти элементы в явном виде. Будучи, как правило, основаны на формально-логической и символической обработке структурированной информации (а именно такая обработка упоминалась в последней четверти двадцатого века как методы искусственного интеллекта), ЭС в течение нескольких десятилетий упоминались в научных публикациях под рубрикой “искусственный интеллект”. Другое их часто встречавшееся название – “системы, основанные на знаниях” – обязано названию одного из их основных компонентов – базы знаний, которая могла хранить не только информацию о моделях объектов и их взаимосвязях (в частности, структурных взаимосвязях между элементами изображений), но и алгоритмы манипулирования такими данными.

Совокупность корректно взаимодействующих базы знаний, механизма их формально-логической обработки и интерфейсов разработчика и пользователя представляли удобный инструмент теоретического исследования способов описания и распознавания элементов окружающей действительности (синтез и распознавание речи, описание и распознавание сцен, доказательство теорем, исследование игровых стратегий и стратегий поведения и т.п.). Не удивительно, что

вслед за успешным применением таких систем “искусственного интеллекта” для решения достаточно абстрактных формально логических задач было предпринято множество попыток их использования при решении важных практических проблем. Однако бесконечное разнообразие свойств реального мира (в отличие от ранее решавшихся идеализированных абстрактных задач “искусственного интеллекта”) закономерно повлекло за собой расширение количества и объема моделей объектов и алгоритмов их обработки, хранящихся в базе знаний, и расширение и усложнение используемых механизмов формально-логического вывода. Не удивительно поэтому, что попытки практического использования расширенных таким образом “систем основанных на знаниях” обычно упоминаются в контексте выполнения больших исследовательских или военных проектов (см., например, [36–38]), масштабы которых позволяли включать в схему автоматического анализа изображений все новейшие (на тот момент) инструменты семантического анализа (такие, например, как аппарат фреймов), оптимизации решений и параллельных вычислений (в частности, искусственные нейронные сети).

Лавинообразное усложнение решаемых реальных задач закономерно привело к специализации применяемых “систем искусственного интеллекта”. С одной стороны, в рамках одного проекта стали применять несколько самостоятельных специализированных ЭС для решения различных напрямую не связанных задач (см., например, [36, 39, 40]), в частности, – для диалога с человеком, для интерпретации изображений целей и т.п. Иными словами, стало нецелесообразным дальнейшее применение сложного универсального механизма для решения более скромной частной проблемы. С другой стороны, стало очевидным, что возможности механизма абстрактно-логического вывода недостаточны для описания и анализа многообразия свойств реальной действительности (как в частном случае недостаточны возможности механизмов формальной лингвистики для понимания живого языка). В связи с этим, механизмы формально-логического вывода стали дополняться многочисленными эвристиками, специфичными для отдельных практических задач [3]. Вот отдельные такие эвристики: “поиск мобильных наземных целей надо начинать на дорогах”, или “из-за сухих климатических условий возможно образование пылевого шлейфа



маскирующего колеса при движении транспортного средства”.

Такую специфическую практически важную эвристическую информацию предоставляли эксперты – специалисты в определенной области знания, в связи с чем “системы искусственного интеллекта” стало более корректно упоминать как ЭС. С этого момента разработчиков систем компьютерного зрения все меньше занимают способы применения универсальных механизмов формально-логического вывода, эйфория по поводу их возможностей прошла. Реальные усилия переносятся с разработки громоздких исследовательских комплексов, на конструирование компактных практически действующих систем, например, бортовых ЭС. В научно-практических публикациях уделяется внимание как раз особенностям разработанных алгоритмов обработки и анализа изображений – особенностям применяемых эвристик (см., например, [40–48]).

В описаниях современных систем автоматической интерпретации аэрокосмических снимков (см., например, [49, 50]) ЭС или “искусственный интеллект” могут вообще не упоминаться, хотя они решают такие же по масштабности задачи, как описанные двадцать лет назад громоздкие “системы основанные на знаниях”. В то же время такие современные системы все еще часто включают в себя в качестве неотъемлемой части базы данных, хранящие знания (описания) об объектах исследования и автоматического анализа. Исследователей и разработчиков интересуют конкретные применяемые “знания” и основанные на них частные эвристические решения. Это, в частности, относится к такой практически важной проблеме, связанной с анализом реальных плохо предсказуемых и формализуемых ситуаций, как конструирование зрительных систем автономных мобильных роботов (см, например, [51, 52]).

### **Кризис структурных методов, основанных на частных эвристиках**

Ахиллесовой пятой большинства ЭС оказалась чрезвычайно узкая сфера эффективного применения каждой из них. Это закономерно следовало из свойств применяемых эвристических решений, отдельно разрабатываемых для распознавания каждого нового конкретного типа объектов. Кризис эвристического

подхода стал наиболее очевиден при решении задач сопоставления и распознавания изображений, характеризующихся большой степенью неопределенности подлежащего анализу сценария и существенными возможными изменениями облика представленных в них объектов, например, при автоматической обработке аэрокосмических снимков Земли или при построении мобильных автономных роботов, предназначенных для применения на открытой местности. Трудности решения отдельных аспектов такого рода сложных проблем достаточно успешно преодолевались путем применения экспертных знаний и основанных на них эвристических решений. Однако такие успехи в основном были обусловлены частным характером решаемых задач. Nevatia, например, при обнаружении объектов искусственной инфраструктуры (см., например, [53, 54]) искал в аэрофотоснимках прямолинейные фрагменты контуров, складывал из них прямоугольные элементы, на основании которых строил потом взлетно-посадочные полосы, улицы и т.д. Военные ЭС, ориентированные на обнаружение танков, искали в изображении дугообразные контуры, после чего пытались “сложить” из них колеса, учитывая при этом, как упоминалось выше, что пылевые шлейфы могут их частично экранировать. При этом, эвристики, ориентированные на эффективный поиск наземных транспортных средств были непригодны для распознавания самолетов (для решения этой задачи приходилось разрабатывать отдельную ЭС), а алгоритмы, применявшиеся Nevatia, теряли эффективность при анализе природных ландшафтов, не складывающихся так легко, как искусственные, из прямоугольных элементов.

Резюмируя этот этап развития техники компьютерного зрения, можно утверждать, что структурные методы, основанные на эвристиках, успешно применяемые при решении частных задач, не достаточно эффективны при решении более общих проблем. Теоретически, можно было бы строить очень громоздкие ЭС, включающие большие наборы разнообразных эвристик, но для решения указанных максимально сложных задач (особенно для задачи сопоставления аэрокосмических снимков) выстраивание общего алгоритма как набора частных эвристик становится бессмысленным в силу бесконечного разнообразия наблюдаемых природных ландшафтов. Иными словами, для этого класса задач методы, основанные на наборах

частных эвристик, исчерпали свои возможности. Назрела необходимость разработки более общего подхода к структурному анализу. В отличие от ранее применявшихся подходов, и будучи ориентирован на автоматический анализ широкого класса изображений, новый подход не мог основываться на частных эвристиках, следующих из особенностей конкретного класса анализируемых объектов. В основу методов анализа изображений, применяемых в рамках такого подхода, должны были быть положены только наиболее общие ограничения, вытекающие из самых общих законов организации наблюдаемого мира и свойств применяемых систем формирования изображений. И такие эффективные объектно-независимые методы действительно были разработаны. Некоторые из таких методов [55–63], подробно рассмотренные в следующем разделе, разрабатывались при непосредственном участии одного из авторов настоящего обзора главным образом для решения задач автоматического анализа и сопоставления аэрокосмических снимков. Последующие разделы посвящены другим современным методам объектно-независимого структурного анализа.

### **Объектно-независимый структурный анализ с использованием алфавита обобщенных структурных элементов**

В течение последних двух десятилетий коллективом российских ученых при участии одного из авторов настоящей статьи был разработан объектно-независимый подход к структурному анализу изображений [55–65]. Целью разработки было преодоление трудностей анализа изображений, характерных для объектно-специфических алгоритмов, применяемых в ЭС. В основу разработанных методов объектно-независимого структурного анализа были положены только наиболее общие ограничения, следующие из свойств применяемых систем формирования изображения и законов организации наблюдаемого мира. Был синтезирован универсальный алфавит очень простых структурных элементов, имеющих предельно узкую полосу частот пространственного спектра, расположенную вокруг нулевой гармоники. В силу ограниченности спектральной полосы использованные структурные элементы оказались инвариантны к аффинным геометрическим преобразованиям и устойчивы к естественным изменениям

наблюдаемых сцен. При выполнении структурного сопоставления изображений была использована аффинная модель их взаимного геометрического преобразования, применяемая при съемке с достаточно больших расстояний через традиционные зеркальные и линзовые объективы. В рамках реализации предложенного подхода была достигнута возможность сопоставления изображений естественных ландшафтов, сформированных в разные сезоны датчиками различающихся типов (оптических видимого диапазона, инфракрасного диапазона, радиолокационных, дальнометрических). Появилась возможность сопоставления изображения с картой местности или даже с контурным наброском, грубо нарисованным от руки. Проективные преобразования с умеренной нелинейной составляющей также не являются существенным препятствием при использовании этого подхода в задаче сопоставления изображений.

Предельно простая форма структурных элементов позволяет сделать структурное сопоставление инвариантным к выше описанным искажающим факторам и геометрическим трансформациям, но, чтобы сделать сопоставление множества простых (похожих друг на друга) структурных элементов достаточно однозначным, в процессе структурного анализа максимально учитывается их взаимное пространственное положение, а его изменение ограничивается допустимой (аффинной) моделью геометрического преобразования, примененной ко всей площади анализируемого изображения. Это допущение делается некорректным при анализе трехмерных сцен, в которых изображение каждой поверхности претерпевает преобразование, описываемое индивидуальной геометрической моделью [59, 60]. Эта трудность преодолевается другими методами структурного анализа, описанными в следующих разделах.

### **Автоматическое сопоставление изображений по локальным информативным точкам и регионам**

Почти одновременно с исследованиями, описанными в предыдущем разделе, и независимо от них разрабатывались методы, основанные на поиске и анализе локальных информативных точек и локальных информативных регионов изображения. Наибольшую известность среди

них получили методы структурного сопоставления, использующие так называемые признаки SIFT (scale invariant feature transform) или SURF (speeded-up robust features) (см. основополагающие работы [66, 67]). Многочисленные примеры более поздних модификаций и аппаратной реализации SURF- и SIFT-алгоритмов другими авторами (см., например, в [68, 69]) свидетельствуют о возрастающей популярности этих методов. В этом случае также выполняется объектно-независимый структурный анализ изображений, но структурные элементы имеют более сложную форму по сравнению с элементами, рассмотренными в предыдущем разделе. Это позволяет сделать структурное сопоставление достаточно однозначным без привлечения дополнительной информации о взаимном пространственном положении структурных элементов. В общем случае модель взаимного геометрического преобразования сопоставляемых изображений может вообще не использоваться в процессе сопоставления, что дает возможность оперировать с изображениями трехмерных сцен, в частности, для автоматической навигации автономных наземных мобильных роботов (см., например, [59, 70, 71]). Однако в отличие от метода обобщенных структурных элементов, описанного в предыдущем разделе и основывающемся на сопоставлении главным образом на анализе взаимного положения структурных элементов, такие алгоритмы, как SIFT и SURF, описывают структурные элементы совокупностью “микроскопических” особенностей фрагмента изображения, окружающего ключевую точку. Сопоставление таких описаний надежно при анализе изображений гладких поверхностей искусственных объектов, например, при осуществлении навигации внутри помещений. В то же время, поверхности естественных объектов, например, крон деревьев, имеют ноздреватую структуру, изображение которой радикально меняется на “микро-уровне” при изменении ракурса наблюдения или направления освещения. Описание ключевых точек при этом недопустимо искажается. К аналогичным эффектам приводят сезонно-суточные изменения естественных ландшафтов. В этих условиях методы, основанные на анализе ключевых точек, в их классическом варианте становятся неприменимыми для автоматического управления беспилотными аппаратами вне помещений. Справедливости ради необходимо отметить, что предложены модификации алгорит-

мов сопоставления ключевых точек, во многом преодолевающие выше описанные “слабости” алгоритма SIFT. Дополнение описаний, построенных в рамках этого метода, структурной информацией о взаимном пространственном положении структурных элементов внутри локальных регионов группирования сделало результаты сопоставления изображений более робастными [72] и даже позволило применять эти методы для анализа аэрокосмических снимков [73]. Однако возможности методов, основанных на сопоставлении ключевых точек, все еще ниже, чем у методов, описанных в предыдущем разделе, если речь идет об изображениях, имеющих очень сильные локальные отличия, например, при сопоставлении реального снимка с чертежом или картой. Сопоставление ключевых точек также становится ненадежным при сильном изменении ракурса наблюдения объектов трехмерных сцен.

Характерной особенностью алгоритмов SIFT и SURF является применяемый в них метод определения позиций структурных элементов: ключевые точки (центры локальных регионов, описываемых SIFT-дескрипторами) выбираются там, где результаты фильтрации изображения с использованием детекторов Харриса или Гессе [74, 75] имеют экстремумы. Анализ изображения производится последовательно для разных коэффициентов его масштабирования, а содержание локального региона, описываемого SIFT-дескриптором, нормализуется относительно вращения. Такой способ описания структурных элементов делает их инвариантными к смещениям, масштабированию и вращению в плоскости изображения. Альтернативный метод расположения SIFT-дескрипторов применяется алгоритмом DenseSIFT [76]. Анализ в общем случае производится без изменения масштаба изображения, ключевые точки равномерно распределены по всей его площади и располагаются в узлах сетки ортогональных линий, следующих с одинаковым шагом. Такой формат использования SIFT-дескриптора больше соответствует описанию не формы объектов, а текстур их поверхностей, особенно, если учесть, что SIFT-дескрипторы обычно рассчитываются на пространственном разрешении, близком к предельному.

Необходимо отметить, что SIFT и SURF [66, 67] не единственные в своем роде, но являются наиболее известными в ряду аналогичных методов [77–80], которые можно в целом охарак-

теризовать как методы, основанные на поиске и анализе локальных информативных точек и регионов изображения, например, “максимально стабильных экстремальных регионов” [80]. Некоторые авторы [81] предлагают методы автоматического обучения системы признаков для описания информативных точек. Однако своей известностью SIFT- и SURF-алгоритмы обязаны своей оптимальности с точки зрения эффективности практического применения. Она достигается на основе разумного компромисса между вычислительной эффективностью этих методов и получаемой с их помощью надежностью структурного сопоставления изображений.

Нельзя не упомянуть метод ASIFT [82, 83], согласно которому классический алгоритм SIFT дополнен поиском наилучшего решения среди всех возможных вариантов анизотропного по направлению масштабирования изображения, соответствующих всевозможным пространственным положениям видеокамеры относительно анализируемой трехмерной сцены. За счет этого достигается инвариантность метода SIFT к группе аффинных преобразований, не включающей зеркальные отражения, в связи с чем новый метод и получил название “affine SIFT” или ASIFT. Предложен также метод [84] сопоставления ключевых точек в динамически меняющихся трехмерных сценах.

Классификация изображений с использованием алгоритма Вайолы-Джонса

Начиная с этого раздела, мы переходим от рассмотрения алгоритмов сопоставления описаний двух конкретных изображений к алгоритмам классификации, выполняющим сопоставление анализируемого изображения с усредненным абстрактным описанием эталонного изображения какого-либо “известного” алгоритму-классификатору класса объектов. Рассматривая объектно-независимые по принципу функционирования структурные классификаторы, нельзя не упомянуть адаптивно обучаемые классификаторы, предложенные Вайолой и Джонсом [85]. Конструкции в виде нескольких прилегающих друг к другу однородных и различающихся по яркости прямоугольных полей, используемые этими авторами в качестве структурных элементов и упоминаемые как вейвлеты Хаара, имеют достаточно простую форму, что не способствует достижению высокой однозначности выполняемого на их основе структурного распознавания. В связи с этим

есть все основания полагать, что основой успеха, достигнутого этими авторами при решении задачи обнаружения лиц, является используемый ими алгоритм AdaBoost (сокращение от Adaptive Boosting) – алгоритм усиления классификаторов, путем объединения их в комитеты. Основой высокой производительности алгоритма Вайолы-Джонса является организация процесса классификации в виде каскада последовательно применяемых решающих правил, позволяющая отбрасывать большинство неверных решений на самых ранних стадиях структурного анализа. Эффективность обучения классификатора во многом обязана применяемой стратегии повторного предъявления для обучения наиболее сложных для распознавания примеров. Неоспоримым достоинством алгоритма Вайолы-Джонса является его способность самостоятельно (в отличие от некоторых других рассмотренных ниже классификаторов) определять количество, тип, масштаб и положение структурных элементов в соответствии с параметрами и положением характерных особенностей анализируемого изображения. Как будет показано ниже, именно “каскад Вайолы-Джонса”, а не их структурные элементы, эффективно заимствуется другими авторами при построении более эффективных решений.

Согласно алгоритму Вайолы-Джонса структурный анализ и классификация производится внутри окна, сканирующего всю площадь изображения при использовании широкого диапазона коэффициентов масштабирования, чем достигается робастность результатов классификации объектов по отношению к их положению и размерам. Однако результат классификации остается устойчивым к вращению объекта в плоскости изображения только в диапазоне  $\pm 22^\circ$  относительно эталонного положения. Такой способ анализа изображений (в скользящем окне с переменным масштабом) и плохая устойчивость к вращению характерны и для ряда рассмотренных ниже методов, описывающих изображения гистограммами направления градиентов.

### **Классификация изображений по гистограммам направления градиентов**

Методика применения дескрипторов DenseSIFT была предложена и развита в работах, описывающих изображения гистограммами направления (ориентации) градиентов (Histograms of Oriented Gradients) – HOG-дескрипторами.



В данном случае анализ изображения с использованием пирамиды масштабов также ведется в сканирующем окне, сеткой ортогональных линий окно поделено на ячейки, из которых составлены частично налагающиеся блоки, в которых строятся гистограммы направления градиентов яркости. В отличие от SIFT-дескриптора в данном случае содержимое блоков не нормализуется относительно вращения, что делает использование HOG-дескрипторов ограниченно устойчивым к поворотам объектов в плоскости изображения. Однако сохраненная информация об абсолютном направлении градиентов увеличивает возможности дескриптора при классификации изображений, подвергающихся сложным изменениям. Конкатенацией гистограмм направлений градиентов, вычисленных внутри составляющих окно блоков, формируется многокомпонентный вектор признаков, описывающий содержание сканирующего изображение окна. Классификация текущего содержимого окна выполняется машиной опорных векторов, обучаемой на большом количестве “положительных” и “отрицательных” примеров изображений объектов заданных классов, что, как демонстрируется в базовой работе Дэйлала и Триггса [86], обеспечивает впечатляющие результаты применения HOG-дескрипторов для обнаружения пешеходов в разнообразных ракурсах и позах. Однако классификацию изображений, применяемую Дэйлалом и Триггсом, можно лишь условно назвать структурной (изображение внутри окна неизменным путем делится на отдельно описываемые блоки). Семантически значимые или хотя бы чем-либо отличающиеся друг от друга части объектов не рассматриваются по отдельности. Существенное взаимное перемещение частей объекта приведет к невозможности его правильной классификации. Эта проблема решается алгоритмами, рассмотренными в следующих разделах.

### **Классификация изображений на основе каскадов гистограмм направления градиентов**

В [87] многообещающие возможности HOG-дескрипторов, наконец, дополняются реальным разделением образа, анализируемого в сканирующем окне, на структурные элементы, соответствующие реально существующим локальным особенностям изображения. Это делается

путем выделения семантически значимых подокон внутри главного окна неизменного размера, сканирующего изображения на разных уровнях пирамиды разрешений. Положения и размеры подокон адаптивно определяются на этапе обучения классификатора аналогично тому, как это делалось в каскаде Вайолы-Джонса, однако, в отличие от простых прямоугольных структурных элементов, состоящих из однородных по яркости полей, теперь содержимое каждого подокна описывается отдельным HOG-дескриптором. Авторы HOG-каскада показывают, что за счет его введения достигается заметное улучшение качества классификации изображений как по сравнению с каскадом Вайолы-Джонса, так и по сравнению с простым применением HOG-дескрипторов Дэйлалом и Триггсом. Кроме того, использование каскада классификаторов способствует ускорению принятия решения на этапе классификации.

### **Классификация изображений на основе моделей с отдельно настраиваемыми частями**

Вслед за алгоритмом Вайолы-Джонса и HOG-дескрипторами Дэйлала и Триггса HOG-каскад, рассмотренный в предыдущем разделе, сохраняет ограниченную инвариантность процесса классификации к вращению в плоскости изображения и изменению ракурса наблюдения. Кроме того, используемые в HOG-каскаде подокна, хотя и соответствуют по своему положению и размерам реально существующим особенностям изображения объекта, в общем случае никак не связаны с его семантически значимыми частями. Поэтому при формировании таких подокон (и соответствующих им структурных элементов) не учитываются особенности характерного для конкретного объекта перемещения его частей. Выше перечисленные трудности в значительной мере разрешаются алгоритмом, использующим модели с отдельно настраиваемыми частями [88]. В соответствии с терминологией статьи-первоисточника назовем его здесь алгоритмом DTPBM (Discriminatively Trained Part Based Models).

Этот алгоритм вслед за алгоритмом Дэйлала и Триггса описывает изображение HOG-дескрипторами в сканирующем окне, используя пирамиду масштабных уровней (пространственных разрешений), а классифицируются полученные дескрипторы машиной опорных

векторов. Однако наряду с дескриптором всего окна (всего содержащегося в окне объекта) формируются также дескрипторы нескольких его подокон, реально соответствующих семантически значимым частям объекта. Наиболее характерное (среднее) положение таких частей определяется для объектов конкретного класса при обучении машины опорных векторов, а отклонение частей от таких характерных положений “штрафуется” при попытке отнесения к этому классу неизвестного классифицируемого изображения.

При обучении классификатора распознаванию объектов конкретного класса изображения этих объектов предъявляются ему в нескольких ракурсах. При этом, несколько субклассификаторов автоматически отделяют “свой” ракурс от остальных и обучаются распознавать только его. Таким образом частично решается как проблема устойчивости классификации по отношению к ракурсу наблюдения объекта, так и проблема устойчивости к вращению в плоскости изображения, поскольку объект, повернутый на определенный угол, можно интерпретировать как наблюдаемый с другого ракурса.

Здесь необходимо отметить, что в этой секции мы перешли к рассмотрению самых совершенных современных алгоритмов структурной классификации. В частности, алгоритм DTPBM – один из победителей соревнования программ-классификаторов 2011 года проводившегося в системе ImageNet [89]. При классификации изображения в один из 1000 известных ей классов программа совершала только около 15% ошибок. При этом объекты классификации предъявлялись в самых разнообразных ракурсах и модификациях, в разнообразном контекстном окружении, частично загороженными. Это колоссальный успех! Нужно отметить, что при классификации учитывались дополнительные контекстные ограничения (например, характерное положение объекта в кадре) и применялась специальная стратегия размещения границ сканирующего окна относительно границ объекта.

### **Классификация изображений искусственными нейронными сетями “с глубоким обучением”**

Перейдем теперь к алгоритму – главному победителю [90]. Он продолжает лидировать в соревнованиях системы ImageNet и в 2012, и в

2013 году. Это – так называемая искусственная нейронная сеть “с глубоким обучением”. В этой многослойной сети подобно тому, как это четверть века назад было сделано Ямадой [28, 29] и Фукусимой [30], глобальные системы связей из слоя в слой заменены локальными, что позволяет более эффективно справляться с изменением взаимного пространственного положения частей объектов и локальными вариациями их формы. Обучение сети с использованием большой репрезентативной выборки изображений объекта распознавания (десятки тысяч примеров на один класс объектов) позволяет ей затем с минимальным количеством ошибок распознавать “известный ей объект” в самых разнообразных ракурсах и модификациях и в разном контекстном окружении. В отличие от изоциренных алгоритмов классификации, рассмотренных в предыдущих разделах, в процессе обучения сеть сама определяет количество, характерные формы и позиции частей объекта. Авторы сети называют ее “сверточной” (convolutional), поскольку детали объекта в результате обучения запоминаются в виде усредненных локальных фильтров, не содержащих высоких пространственных гармоник (подобно описанным в одном из предыдущих разделов обобщенным объектно-независимым структурным элементам). Кроме того, при распознавании помимо локальных особенностей формы объекта сеть принимает во внимание и локальные цветовые параметры. Правда, обучение такой сети занимает неделю на компьютере с двумя графическими ускорителями.

### **Заключение**

Мы рассмотрели одни из самых лучших современных алгоритмов структурного сопоставления и структурной классификации изображений. Уже в наши дни ими достигнуты впечатляющие результаты, а их возможности в области распознавания изображений приблизились к человеческим, а в ряде случаев и превзошли их. Необходимо, однако, констатировать, что эти лучшие алгоритмы при анализе изображений реальных сцен до сих пор почти не используют информацию о третьей пространственной координате. Фактически, анализируются только плоские проекции объектов. В то же время, уже в базовой статье по алгоритму SIFT [66] продемонстрировано, как использовать информацию о наклонах поверх-

ностей, а это – уже шаг к трехмерному анализу. Использование информации о третьей пространственной координате позволило бы разрешить трудности применения описанного выше алфавита обобщенных структурных элементов при объектно-независимом структурном анализе трехмерных сцен. Изопсанный алгоритм классификации изображений на основе моделей с раздельно настраиваемыми частями вместо искусственно заданных ракурсов наблюдения, возможно, смог бы более объективно анализировать изображения объектов, сфото-

графированных с произвольного направления. Современные средства формирования видеоданных позволяют эффективно измерять третью координату, есть все основания полагать, что это позволит сделать очередной важный шаг в структурном анализе изображений.

Исследование выполнялось при частичной финансовой поддержке Совета по грантам Президента Российской Федерации (грант МД-1072.2013.9) и при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

\* \* \* \* \*

## ЛИТЕРАТУРА

1. Тайдзо И., Фудзицу к.к. Устройство для распознавания изображений // Патент Японии № 56-33747. 1973.
2. Hajime industries, Ltd. Pattern discrimination method // Заявка Великобритании № 2035645. Abstracts. 1980. № 4766.
3. Луцив В. Автоматический анализ изображений. Объектно-независимый структурный подход. Саарбрюкен, Германия: Ламберт Академик Паблшинг, 2011. 300 с.
4. Casasent D., Furman A. Sources of correlation degradation // Applied Optics. 1977. V. 16. № 6. P. 1652–1661.
5. Кейсесент Д. Оптическое когерентное распознавание образов // ТИИЭР. 1979. Т. 67. № 5. С. 131–148.
6. Casasent D., Psaltis D. Deformation-invariant, space-variant optical pattern recognition // Progress in Optics. 1978. V. XVI. P. 291–356.
7. Chen Q., Defrise M., Deconinck F. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1994. V. 16. № 12. P. 1156–1167.
8. Ту Дж., Гонсалес Р. Принципы распознавания образов. Пер. с англ. / Под ред. Журавлева Ю.И. М.: Мир, 1978. 411 с.
9. Чавчанидзе В.В. Способ классификации объектов по N признакам // Патент СССР № 596979. 1978.
10. Агаджаян Ж.М. Устройство для считывания и распознавания изображений микрообъектов // Патент СССР № 898466. 1982.
11. Коге гидзюцу инте. Схема идентификации контурной информации // Заявка Японии № 56-27913. Токке Кохо. 1981. № 6-698.
12. Ниппон дэнки к.к. Устройство для распознавания изображений // Заявка Японии № 57-21750. Токке Кохо. 1982. № 6-544.
13. Коге гидзюцу инте. Способ выделения характерных признаков линий изображения // Заявка Японии № 56-15544. Токке Кохо. 1981. № 6-389.
14. Maitra S. Moment invariants // Proc. IEEE. 1979. V. 67. № 4. P. 697–699.
15. Hu M.-K. Visual pattern recognition by moment invariants // IRE Trans. on Information Theory. 1962. V. IT-8. P. 179–187.
16. Жеребко А.К., Луцив В.Р. Согласованная фильтрация в естественных и искусственных нейронных сетях // Оптический журнал. 1999. Т. 66. № 9. С. 69–72.
17. Синдзи Я., Хитати сейсакусе к.к. Устройство распознавания изображений посредством их пофрагментного сравнения с эталонами // Заявка Японии № 53-17021. Токке Кохо. 1978. № 47-47204.
18. Kashioka S., Eijri M., Mese M., Miyatake T., Hitachi Ltd. Pattern position detecting system // Патент США № 4091394. 1978.
19. Calspan corporation. Fingerprint identification method and apparatus // Заявка Великобритании № 1577797. Abridgments. 1980. № 4779.

20. *Schroeder E.N., Lexington K.* Adaptive alignment for pattern recognition system // Патент США № 4204193. 1980.
21. *Fukushima K.* Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position // *Biological Cybernetics*. 1980. V. 36. № 4. P. 93–202.
22. *Розенблатт Ф.* Принципы нейродинамики (перцептрон и теория механизмов мозга). М.: Мир, 1965. 480 с.
23. *Rumelhart D.E., Hinton G.E., Williams R.J.* Learning internal representations by error propagation // *Parallel Distributed Processing*. 1986. V. 1. P. 318–362.
24. *Rumelhart D.E., Hinton G.E., Williams R.J.* Learning representations by back-propagating errors // *Nature*. 1986. V. 323. P. 533.
25. *Casasent D.P., Barnard E.* Adaptive clustering optical neural net // *Applied Optics*. 1990. V. 29. № 17. P. 2603–2615.
26. *Farr K.B., Hartman R.L.* Optical-digital neural network system for aided target recognition // *Proc. SPIE*. 1995. V. 2485. P. 141–154.
27. *McAulay A., Kadar I.* Image compression and automatic target recognition // *Proc. SPIE*. 1989. V. 1099. P. 74–82.
28. *Yamada K., Kami H., Tsukumo J., Temma T.* Handwritten numeral recognition by multilayered neural network with improved learning algorithm // *IJCNN Intern. Joint Conf. on Neural Networks*. USA. D.C., 1989. P. II-259–II-266.
29. *Yamada K.* Improved learning algorithm for multilayer neural networks and handwritten numeral recognition // *NEC Research and Development*. 1990. July. № 98. P. 81–88.
30. *Fukushima K.* Neural network model for selective attention in visual pattern recognition and associative recall // *Applied optics*. 1987. V. 26. № 23. P. 4985–4992.
31. *Фу К.* Структурные методы распознавания образов. Пер. с англ. / Под ред. Айзермана М.А.. М.: Мир, 1977. 319 с.
32. Интегральные роботы. Сборник статей. Вып. 1. Пер. с англ. / Под ред. Поздняка. Г.Е.. М.: Мир, 1973. 421 с.
33. Интегральные роботы. Сборник статей. Вып. 2. Пер. с англ. / Под ред. Поздняка. Г.Е.. М.: Мир, 1975. 449 с.
34. *Гренандер У.* Синтез образов. // Лекции по теории образов. Т.1. Пер. с англ. / Под ред. Журавлева Ю.И.. М.: Мир, 1979. 384 с.
35. *Гренандер У.* Анализ образов. // Лекции по теории образов. Т.2. Пер. с англ. / Под ред. Журавлева Ю.И.. М.: Мир, 1981. 448 с.
36. *Corby N.R., Mundy J.L., Vrobel P.A.* An environment for intelligence analysis // *Proceedings: 18<sup>th</sup> Image Understanding Workshop*. 1988. 6–8 April. Cambridge, Massachusetts, USA. P. 342–350.
37. *Fisher M.A., Bolles R.C.* Image understanding research at SRI international // *Proceedings: 18<sup>th</sup> Image Understanding Workshop*. 1988. 6–8 April. Cambridge, Massachusetts, USA. P. 53–61.
38. *Lawton D.T., Levitt T.S., Gelband P.* Knowledge based vision for terrestrial robots // *Proceedings: 18<sup>th</sup> Image Understanding Workshop*. 1988. 6–8 April. Cambridge, Massachusetts, USA. P. 103–110.
39. *Степанов А.А.* Развитие искусственного интеллекта в Западной Европе и США // *Новости зарубежной науки и техники. Системы авиационного вооружения*. 1990. № 15. С. 26–30.
40. *Matsuyama T.* Expert systems for image processing: Knowledge based composition of image analysis processes // *Computer vision, graphics and image processing*. 1989. V. 48. Is. 1. P. 22–49.
41. *Dijk J., van Eekeren A.W.M., Rojas O.R., Burghouts G.J., Shutte K.* Image processing in aerial surveillance and reconnaissance: from pixels to understanding // *Proc. SPIE*. 2013. V. 8897. P. 88970A-1–88970A-17.
42. *Bourma H., Dekker R.J., Shoemaker R.M., Mohamoud A.A.* Segmentation and wake removal of seafaring vessels in optical satellite images // *Proc. SPIE*. 2013. V. 8897. P. 88970B-1–88970B-11.
43. *Beech E.* Sight Unseen // *Flight International*. 2 February. 1989. V. 135. № 4151. P. 34–35.
44. *Бочкарев А.М., Почуев С.И.* Экспертные системы – электронные консультанты летчика (современное состояние проблемы, перспективы) // *Зарубежная радиоэлектроника*. 1989. № 10. С. 42–51.
45. *Ergul M., Alatan A.A.* An automatic geo-spatial object recognition algorithm for high resolution satellite images // *Proc. SPIE*. 2013. V. 8897. P. 88970J-1–88970J-9.



46. McKeown D.M., Wilson Jr., Harvey A., Ewixson L. Automating knowledge acquisition for aerial image interpretation // Computer vision, graphics and image processing. 1989. V. 46. № 1. P. 37–81.
47. Fisher M.A., Strax T.M. Recognizing objects in a natural environment: a contextual vision system (CVS) // DARPA image understanding workshop. May 23–26. 1989. Palo Alto, California, USA. P. 774.
48. De Catrel J.M., Surdu J.R. Practical recognition of armored vehicles in FLIR // Proc. SPIE. 1995. V. 2485. P. 200–208.
49. Ponteau B. Vector-to-raster change detection for property assessment // Earth observation magazine. 2005. V. 14. Is. 1. P. 10–13.
50. Shackelford A.K., Davis C.H. Automated processing of high resolution satellite imagery for feature extraction and mapping of urban areas // Earth observation magazine. 2005. V. 14. Is. 1. P. 17–19.
51. Murillo A.S., Guerrero J.J., Sagiés C. SURF features for efficient robot localization with omnidirectional images // IEEE international conference on robotics and automation. 2007. April. Rome, Italy. P. 3901–3907.
52. Jeong W.Y., Lee K.M. Visual SLAM with line and corner features // Intelligent Robots and Systems. IEEE/RSJ International Conference. 9–15 October. 2006. Beijing, China. P. 2570–2575.
53. Huertas A., Cole W. and Nevatia R. Detecting runways in aerial images // Proceedings of Sixth International Conference on Artificial Intelligence AAAI-87. 1987. July 13–17. Seattle, Washington, USA. P. 712–717.
54. Nevatia R., Lin C., Huertas A. A system for building detection from aerial images // Automatic extraction of man-made objects from aerial and space images. Basel: Birkhauser Verlag, 1997. P. 77–86.
55. Lutsiv V.R., Dolinov D.S., Zhrebko A.K., Novikova T.A. Using artificial neural networks in image processing problems // J. optical technology. 1997. V. 64. № 2. P. 112–118.
56. Lutsiv V.R., Malyshev I.A., Pepelka V.A. Automatic fusion of the multiple sensor and multiple season images // Proc. SPIE. 2001. V. 4380. P. 174–183.
57. Lutsiv V.R., Malyshev I.A., Pepelka V.A., Potapov A.S. The target independent algorithms for description and structural matching of aerospace photographs // Proc. SPIE. 2002. V. 4741. P. 351–362.
58. Lutsiv V.R., Malyshev I.A., Potapov A.S. Hierarchical structural matching algorithms for registration of aerospace images // Proc. SPIE. 2003. V. 5238. P. 164–175.
59. Луцев В.Р., Лапина Н.Н., Малышев И.А., Потапов А.С. Особенности сопоставления изображений в задачах определения местоположения мобильного робота // Оптический журнал. 2010. Т. 77. № 11. С. 25–34.
60. Lutsiv V., Potapov A., Novikova T., Lapina N. Hierarchical 3D structural matching in the aerospace photographs and indoor scenes // Proc. SPIE. 2005. V. 5807. P. 455–466.
61. Lutsiv V.R., Andreev V.S., Gubkin A.F., Iljashenko A.S., Kadykov A.B., Lapina N.N., Malyshev I.A., Novikova T.A., Potapov A.S. Algorithms for automatically processing and analyzing aerospace pictures // J. optical technology. 2007. V. 74. № 5. P. 307–322.
62. Луцев В.Р. Объектно-независимый подход к структурному анализу изображений // Оптический журнал. 2008. Т. 75. № 11. С. 26–34.
63. Lutsiv V., Malyshev I. Image structural analysis in the tasks of automatic navigation of unmanned vehicles and inspection of Earth surface // Proc. SPIE. 2013. V. 8897. P. 88970F-1–88970F-15.
64. Potapov A.S., Gamayunova O.S. Information criterion for constructing the hierarchical structural representations of images // Proc. SPIE. 2005. V. 5807. P. 443–454.
65. Potapov A.S. Comparative analysis of structural representations of images based on the principle of representational minimum description length // J. Opt. Techn. 2008. V. 75. № 11. P. 715–720.
66. Lowe D.G. Distinctive image features from scale-invariant keypoints // Int. J. Computer Vision. 2004. V. 60. № 2. P. 91–110.
67. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded Up Robust Features // Proc. 9<sup>th</sup> European Conf. on Computer Vision. 2006. May 7–13. Graz, Austria. P. 404–417.
68. Rudinac M., Lenseigne B., Jonker P. Keypoint extraction and selection for object recognition // Proceedings of MVA2009IAPR Conference on Machine Vision Applications. 2009. May 20–22. Yokohama, Japan. P. 191–194.
69. Qui J., Huang T., Huang Y., Ikenaga T. A hardware accelerator with variable pixel representation & skip mode prediction for feature point detection. Part of SIFT algorithm // Proceedings of MVA2009IAPR Conference on Machine Vision Applications. 2009. May 20–22. Yokohama, Japan. P. 239–242.
70. Choi S.W., Yoon J.S., Redkov V., Son H.S., Na S.I., Kim Y.G., Lee S.S., Baek S.M., Potapov A., Choi Y.J., Yi D.H., Lee J.H., Kim Y.B., Lutsiv V. Apparatus and method for detecting rotation-invariant features for localization of a mobile robot // Patent of Korea №KR20110009547. 2011.

71. Луцив В.Р., Малышев И.А. Автоматический анализ и сопоставление изображений в задачах навигации беспилотных аппаратов и мониторинга земной поверхности // Труды Юбилейной Всероссийской Научно-Технической Конференции "Моделирование Авиационных Систем". 2011. 12–14 апреля. Москва. С. 197–206.
72. Ascani A., Frontoni E., Mancini A., Zingaretti P. Feature group matching for appearance-based localization // IEEE/RSJ International Conference on Intelligent Robots and Systems. 2008. Sept. 22–26. Nice, France. P. 3933–3938.
73. Casetti A., Frontoni E., Mancini A., Ascani A., Zingaretti P., Longhi S. A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications // J. Intelligent Robot Systems. 2011. V. 61. P. 157–168.
74. Harris C., Stephens M. A combined corner and edge detector // Proceedings of the 4-th Alvey Vision Conference. 1988. Manchester, UK. P. 147–151.
75. Mikolajczyk K., Schmid C. Scale & affine invariant interest point detectors // Intern. J. Computer Vision. 2004. V. 60. Is. 1. P. 63–86.
76. Van de S. Koen E.A., Gevers T., Snoek C.G.M. Evaluating color descriptors for object and scene recognition // IEEE transactions on pattern analysis and machine intelligence. 2010. V. 32. № 9. P. 1582–1596.
77. Lindeberg T., Garding J. Shape-adapted smoothing in estimation of 3D shape cues from affine deformations of local 2D brightness structure // Image and Vision Computing. 1997. V. 15. № 6. P. 415–434.
78. Kadir T., Zisserman A., Brady M. An affine invariant salient region detector // Proceedings of the 8-th European Conference on Computer Vision. 2004. Prague, Czech Republic. P. 404–416.
79. Tuytelaars T., Van Gool L. Matching widely separated views based on affine invariant regions // Intern. J. Computer Vision. 2004. V. 59. Is. 1. P. 61–85.
80. Matas J., Chum O., Urban M., Pajdla T. Robust wide baseline stereo from maximally stable extremal regions // BMVC. 2002. P. 384–393.
81. Potapov A.S., Malyshev I.A., Puysha A.E., Averkin A.N. New paradigm of learnable computer vision algorithms based on representational MDL principle // Proc. SPIE. 2010. V. 7696. P. 769606-1–769606-11.
82. Morel J.M., Yu G. ASIFT: A new framework for fully affine invariant image comparison // SIAM J. Imaging Sciences. 2009. V. 2. Is. 2. P. 438–469.
83. Yu G., Morel J.M. A fully affine invariant image comparison method // Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 2009. Taipei, Taiwan. P. 1597–1600.
84. Peterson M.V. Clustering of a set of identified points on images of dynamic scenes, based on the principle of minimum description length // J. optical technology. 2010. V. 77. № 11. P. 701–706.
85. Viola P., Jones M.J. Rapid object detection using a boosted cascade of simple features // Proceedings of IEEE Computer Vision and Pattern Recognition Conference. 2001. 8–14 December. Kauai, HI, USA. P. I-501–I-518.
86. Dalal N., Triggs B. Histograms of oriented gradients for human detection // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005. San Diego, CA, USA. P. 886–893.
87. Zhu Q., Avidan S., Yeh M., Cheng K. Fast human detection using a cascade of histograms of oriented gradients // Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006. N.Y., USA. V. 2. P. 1491–1498.
88. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part based models // IEEE Trans. Pattern Analysis and Machine Intelligence. 2010. V. 32. № 9. P. 1627–1645.
89. <http://www.image-net.org> (5 мая 2014 г.).
90. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // NIPS. 2012. P. 1106–1114.