

УДК 004.421

## **Методология определения формы объектов по данным видеоряда камеры**

**Ефимов А.И.\*, Ильин В.Н.\*\***

*Московский авиационный институт (национальный исследовательский университет), МАИ, Волоколамское шоссе, 4,  
Москва, А-80, ГСП-3, 125993, Россия*

*\*e-mail: [alhimic23@yandex.ru](mailto:alhimic23@yandex.ru)*

*\*\*e-mail: [vnil2005@yandex.ru](mailto:vnil2005@yandex.ru)*

### **Аннотация**

В данной статье описывается методология определения формы объемных объектов по данным, полученным из видеоряда камеры и области её применения. Также затронуты методы предварительной обработки изображения, особенности поиска ключевых особенностей и их описания по сравнению с поиском на плоских объектах. Отдельно рассмотрено применение полученных данных для расширения сферы использования дополненной реальности и увеличения реалистичности конечного изображения. Кроме того описаны открывающиеся возможности стереометрии по одному видеопотоку и отмечены дальнейшие шаги по улучшению рассмотренных методик.

**Ключевые слова:** компьютерное зрение, анализ формы объекта, дополненная реальность, визуальное распознавание объектов.

В настоящее время в связи с быстрым развитием робототехники, 3D технологий и технологий дополненной реальности становится весьма актуальной задача расширения возможностей компьютерного зрения. Обычно эта задача решается путём анализа информации, поступающей из видеокамеры в форме видеоряда, состоящего из последовательности кадров. В данной статье описываются подходы к анализу формы объектов по данным, полученным из видеоряда камеры и возможные области применения результатов этого анализа.

Анализ формы объекта видеоряда серьезно расширяет возможности компьютерного зрения и имеет ряд технических применений. Во-первых, он позволяет строить трехмерную модель детали без дорогостоящих устройств, которую например можно в дальнейшем использовать для воспроизведения объекта на 3D принтере. Во-вторых, анализ расширяет возможности технологии дополненной реальности за счет более полной получаемой информации. На данный момент хорошо изучены алгоритмы анализа плоских поверхностей[1][2][7], но алгоритмы анализа объемных объектов более сложной формы не изучены в достаточной степени. И алгоритмы компьютерного зрения в этой области по надежности особенно сильно уступают человеческому зрению.

Это связано с тем что данный анализ сталкивается с большим количеством сложностей по сравнению с распознаванием плоских объектов. Действительно, как видно из рис.1, при перемещении и вращении плоского объекта поведение точек, если пренебрегать перспективой, поддается законам аффинного преобразования, т.е. гораздо легче проводить параллели между 2 наборами точек через их взаимное расположение, кроме того алгоритмы поиска точек в данном случае работают

намного устойчивее. При вращении объекта сложной формы наблюдается сильная вариация набора ключевых особенностей и их взаимного расположения в зависимости от ракурса камеры, кроме того элементы объекта могут перекрывать друг друга. Помимо этого серьезное влияние оказывает вариация тонов в зависимости от угла падения света и особенностей освещения, влияния шума и бликов на полированных поверхностях. Также большое влияние оказывает присутствие в сцене динамических объектов. Данные особенности влияют как на выборку детекторов, так и на содержание дескрипторов.

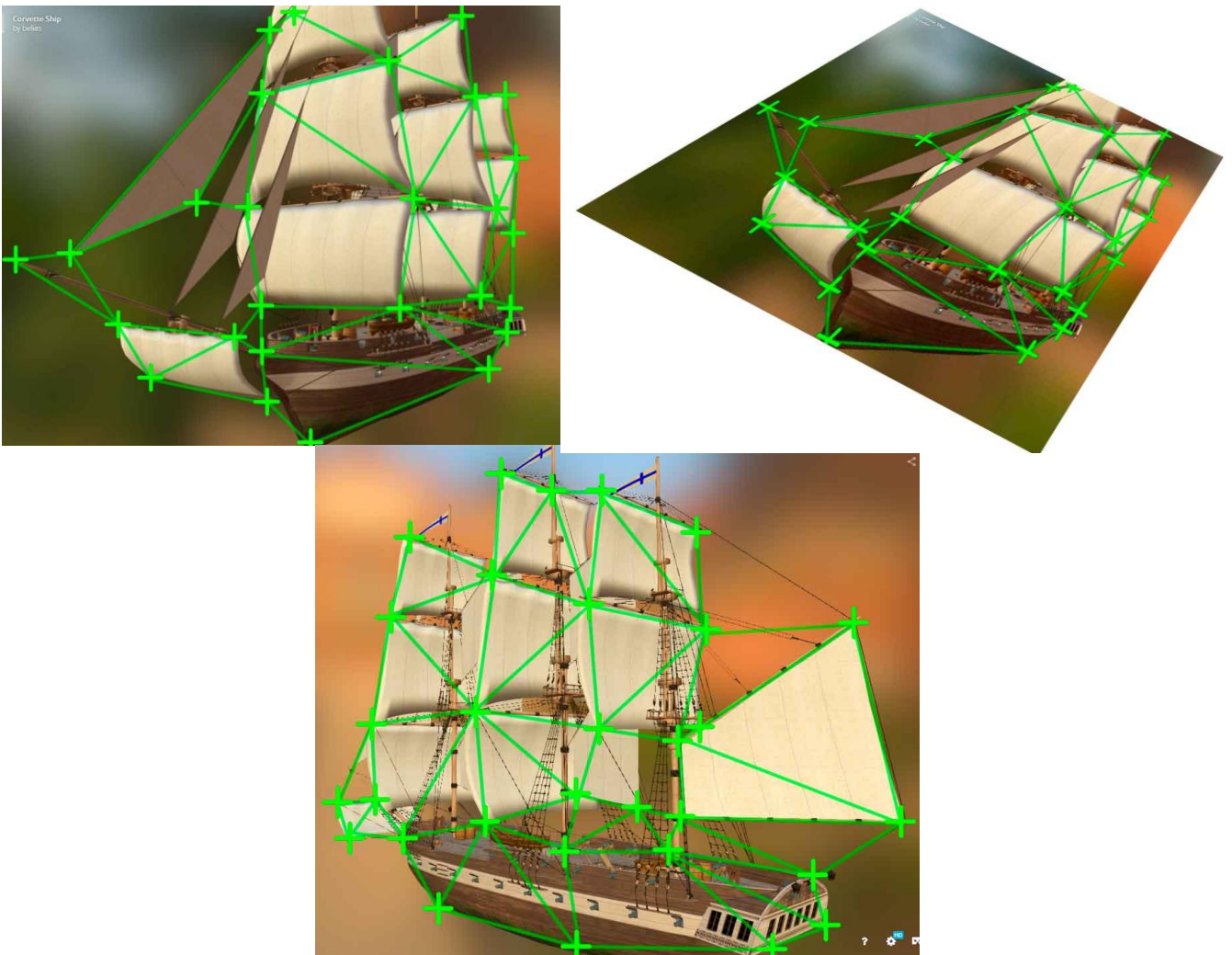


Рис. 1 Поведение ключевых точек

На данный момент на рынке присутствует большое число продуктов, определяющих форму объекта, однако сферы их применения весьма ограничены. Самый точный из предлагаемых подходов требует наличия дорогостоящего 3D сканера, однако этот вариант не всегда приемлем. Другие варианты с использованием дальномеров в связке с камерами, например Kinect, не являются столь же легко применимыми как использование только камеры. Продукты же выпущенные на данный момент, основанные на анализе набора фотографий или видеоряда(такие как: 3DSOM, 3D Photo Builder Professional, PhotoModeler Scanner) не являются достаточно точными, несмотря на широту представленных подходов и многочисленность самих продуктов.

Перед изложением методологии анализа объемных объектов сложной формы дадим некоторые определения. Видеоряд - последовательность изображений (кадров) на видеозаписи. В данном случае это видеозапись одной сцены с разных ракурсов, либо же достаточно большой набор фотографий одной и той же сцены. Кадр - (здесь)фрагмент видеоряда являющийся отдельным изображением, либо одна фотография. Силуэт объекта - граница объекта с текущего ракурса съемки. Ключевая особенность - это некоторый участок картинки, который является отличительным для заданного изображения, это может быть точка, блоб (точка с окрестностью), граница, область. Детектор - алгоритм, осуществляющий поиск ключевых особенностей на изображении. Deskриптор - алгоритм, описывающий ключевые особенности, причем одна и та же ключевая особенность в идеале на разных кадрах должна описываться одинаково, а разные никогда не иметь одинакового описания.

Предлагаемая в данной работе методика анализа видеоряда состоит из следующих последовательно выполняемых пяти шагов:

1. Предобработка изображения. Она уменьшает влияние шума и особенностей освещения на результаты распознавания.
2. Привязка к окружающему пространству. На изображении осуществляется поиск плоского неподвижного маркера. Относительно полученных данных определяется позиция камеры в данном кадре.
3. Поиск стабильных ключевых точек. Происходит поиск ключевых точек методом FAST-ER. Проанализировав набор кадров, осуществляется поиск точек со стабильными мировыми координатами.
4. Построение силуэта объекта. Опираясь на стабильные точки, строится примерная карта глубины изображения и ищутся ключевые точки, расположенные на её перепадах. Проанализировав градиенты на отрезках между ними строится приближенный силуэт объекта на конкретном кадре.
5. Систематизация данных для получения и уточнения полной геометрии объекта.

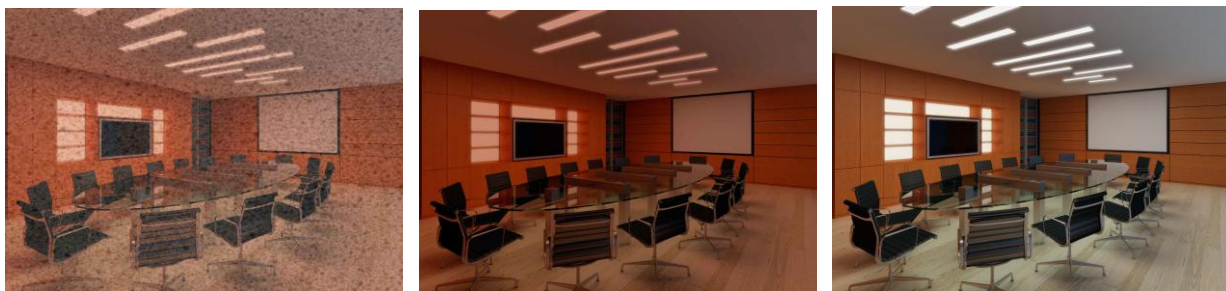


Рис. 2 Предварительная коррекция видеоряда

Шаг первый, предобработка изображения. Перед запуском алгоритмов распознавания для их большей устойчивости важно максимально подавить шум и компенсировать разность освещения, выровнять изображение по яркости, а также произвести цветокоррекцию (если при распознавании учитывается цвет). Для подавления шума логично использовать комбинацию гауссова и медианного фильтров. Медианный фильтр компенсирует шум типа “соль перец”, однако ведет к пропаданию тонких линий, гауссов фильтр позволяет компенсировать гауссов шум, но при этом размывает изображение. Компенсация освещения происходит по принципу вычитания из кадра карты освещенности, полученной применением фильтра гаусса большого разрешения для этого же кадра. Для лучшего эффекта может использоваться многомасштабный вариант данного алгоритма при котором берется взвешенная сумма карт освещенности (обычно 3-х) с фильтрами разного разрешения. Важно, что при данном подходе лучше сохраняется исходный цвет. На рисунке 2(б) демонстрируется результат подавления шума данными фильтрами на исходном изображении 2(а). В случаях, когда оттенок цвета влияет на алгоритм распознавания важно также выполнить цветокоррекцию, для чего используется модель серого мира, либо, если в сцене используется маркер или любой другой объект с точно известным цветом, происходит ориентирование на разницу цвета маркера при белом освещении и при освещении сцены видеоряда. На рисунке 2(в) выполнена цветокоррекция методом серого мира.

Шаг второй, привязка к окружающему пространству. Для реализации процесса распознавания формы в первую очередь требуется разработать алгоритм определения позиции камеры в каждый момент времени. Данная задача достаточно

сложна без информации об окружающем пространстве поэтому, для упрощения задачи сначала используется распознавание плоского маркера. Это проще ибо поведение ключевых точек на плоскости близко к законам аффинного преобразования. Данный подход более точен, т.к. заранее известны свойства объекта к которому осуществляется привязка, в отличие от анализа сразу всех ключевых точек (блобов, областей) видеоряда. Кроме того обычно точно известны размеры маркера и следовательно возможно оценить размеры и дальность других объектов, а также выполнить калибровку камеры. Также сложно подобрать достаточно корректный дескриптор при сложной форме объекта, однако при наличии строгой привязки, в качестве дескриптора используется смещение относительно маркера.

Шаг третий, поиск стабильных ключевых точек. Для поиска ключевых точек в первую очередь нужно определить наиболее подходящий детектор, для данной цели лучше всего подходят детекторы углов, либо комбинированные детекторы углов и блобов. Детекторы областей подходят для плоских простых изображений с четкими границами, либо объектам, близким к плоским. Детекторы блобов при оценочном тестировании показали меньшее качество повторяемости по сравнению с детекторами углов.

После расчета траектории движения камеры, производится построение ключевых точек, для каждого кадра и анализ их смещения на временном промежутке. Для этой цели идеально подходит метод FAST-ER, так как он ориентирован на поиск угловых точек, следовательно точки будут находиться как на границах объектов так и на крупных особенностях объекта. Помимо этого, алгоритм

устойчив к повороту, сдвигу, масштабированию и обладает хорошим свойством повторяемости. Кроме того, полученные ключевые точки обладают классификацией “силы” что важно при наличии многих шагов в процессе анализа. Также важным достоинством является установка 2-х порогов, при превышении верхнего точка начинает распознаваться и продолжает распознаваться, пока её вес не опустится ниже нижнего порога, что позволяет избежать частого исчезания ключевых точек.

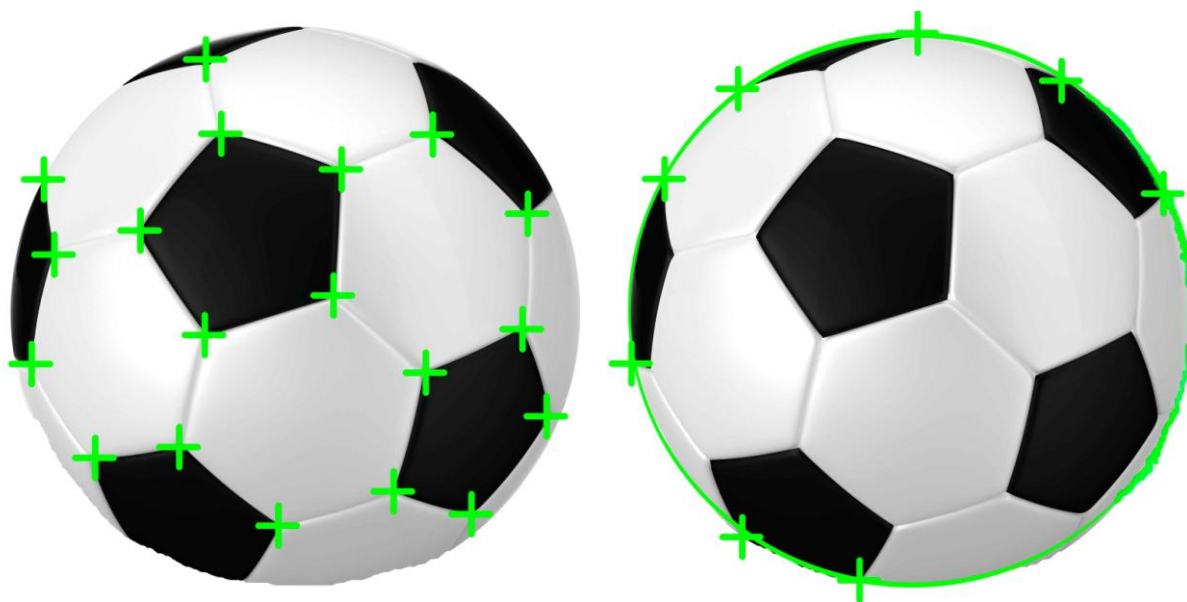


Рис. 3 А и В множества ключевых точек

После этого необходимо выделить из полученного набора точек те, которые не меняют своего положения в 3-х мерном пространстве при изменении ракурса камеры, данные точки выделены на рисунке 3(а). Обозначим это множество как А. К ним относятся локальные особенности самого объекта, это могут быть сильные особенности текстуры, достаточно острые углы, а также не закруглённые края



объекта. Если точка стабильна то по смещению её позиции на 2-х кадрах, зная смещение камеры между ними, определяется приближенное расстояние до этой точки, а используя угол обзора камеры определяется приближенная позиция точки. Для уточнения позиции берется статистика результатов для разных пар кадров. Если результаты соотносятся с нормальным распределением, то точка является стабильной и используется медианная позиция из набора для дальнейших расчетов. По полученным позициям точки группируются, как правило близко расположенные точки относятся к одному объекту и служат хорошим ориентиром для определения его границ на кадре. В качестве дескриптора данных точек используется их глобальная позиция в декартовых координатах, в качестве начала координат берется маркер, оси  $X$ ,  $Y$  берутся по его сторонам а ось  $Z$  перпендикулярно его плоскости.

Шаг четвертый, построение силуэта объекта. В начале данного шага строится приближенная карта глубины и выделяются точки границ объектов, для чего выделяются ключевые особенности, расположенные на сильных перепадах карты глубины, с учётом результатов предыдущих кадров. Эти точки выделяются из множества, полученного детектором FAST-ER (при этом в этот набор могут войти некоторые точки из множества  $A$ ) либо использовать в качестве детектора комбинацию метода Хариса и его модификацию, ориентированную на поиск границ, а не углов. Полученное множество обозначим  $B$ . Данное множество можно наблюдать на рисунке 3(б). По полученным точкам строится ломаная линия, корректность которой определяется сравнением карты глубины с обеих сторон линии. При стабильном значении с одной стороны линии, и сильном перепаде между значениями с разных сторон, линия считается корректной. При отсутствии

данных признаков берется статистика этой линии на других кадрах и в случае её корректности при более ранних наблюдениях, считаем ее корректной, в противном случае игнорируем. Далее уточняется форма объекта, с помощью оценки градиентов изображения перпендикулярных отрезкам этой ломанной. Проходя по каждому отрезку, оценивается изменение цвета и определяется смещение границы объекта от заданного отрезка. После этого получаем примерный силуэт объекта.

Шаг пятый, систематизация. Множество силуэтов, полученных на предыдущем шаге наиболее точно оценивает форму объекта. По мере наполнения этого множества данные систематизируются и форма объекта уточняется.

Зная примерную форму объекта, становится возможным сравнить его с набором эталонов и получить информацию по нему. Так-же по данной оценке определяется его положение в пространстве, что дает возможность осуществить привязку графики при использовании технологии дополненной реальности. Также данный подход возможен для составления самой таблицы эталонов, для чего нужно оценивать только один объект, чаще всего находящийся в центральной части видеоряда с высоким качеством распознавания с условиями близкими к идеальным. Максимальная точность построения эталона важна, т.к. в противном случае его погрешность будет усугублять погрешность при последующем поиске на видеоряде.

При высоконагруженном распознавании целесообразно производить анализ по уже отснятым видео из-за слишком тяжелых вычислений для расчета в реальном времени. Например, это имеет смысл для вышеописанного построения эталонов. Однако чаще приходится иметь дело с распознаванием объектов в реальном времени и выводу актуальной информации на данный момент. При этом крайне

важно оптимизировать данный процесс. Для чего используются полноценные части выбранного алгоритма только на отдельных опорных кадрах. На промежуточных используется тот факт, что изображение резко не меняется и следовательно не нужно производить полный перерасчет параметров, а ключевые точки не будут перемещаться далеко от своих первоначальных позиций. В качестве опорных кадров можно выбирать либо по порядку, пропуская определенное количество, однако лучше выбирать наиболее удачные: с меньшим влиянием шума, лучшим освещением, меньшей смазанностью, оптимальным набором ключевых точек, для чего используется быстрый вариант оценки качества. Так же в процессе наблюдения многие вещи уже можно прогнозировать, что также уменьшает процессорное время на анализ. Также группировка основной части ключевых точек в определенных областях, сильно повышает производительность и конечную точность оценки.



Рис. 4 Использование карты глубины в дополненной реальности

Одной из основных областей применения алгоритмов на основе вышеописанных методик является технология дополненной реальности. Получаемая

информация является весьма полезной для неё сразу с нескольких позиций. Во-первых, распознав объект и определив его позицию имеется возможность вывести как текстовую, так и графическую информацию с привязкой к нему и для этого не потребуется обклеивать его набором маркеров. Во-вторых, повышается реализм виртуальной сцены за счет анализа карты глубины. При обычных условиях виртуальная сцена просто выводится поверх видеоряда и может выглядеть частью реального мира только до тех пор, пока отрисованный контент не перекрывает элементы, находящиеся ближе имитируемой позиции объекта. Данный эффект виден, если взглянуть на рисунок 4(а), где автомобиль прорисовывается поверх столба, расположенного ближе, и не вписывается в окружающий мир. Относительно точная карта глубины позволит избежать данного эффекта, это видно на рисунке 4(б). В третьих, используя карту глубины можно относительно корректно реализовать стереометрию видеоряда на очках виртуальной реальности(данный аспект более подробно описан ниже).

Относительно дальнейших перспектив по развитию данного подхода можно выделить несколько аспектов. Например, уточнение карты глубины, учитывая изменение яркости освещения на разных частях объекта, однако данный подход осложняется падающими тенями и неоднородной текстурой объекта. Также важным дальнейшим шагом является определение движения камеры при пропадании маркера из обзора камеры. При этом алгоритм опирается на данные об окружающих объектах, которые успел получить. Самыми полезными данными в этом случае будут стабильные в трехмерном пространстве точки, корректность которых уже удалось проверить. Теоретически при отсутствии погрешностей и движущихся

объектов, достаточно 3 подобных точек в каждый момент времени. Однако для достижения оптимальной точности потребуется более мощное множество. Еще более сложным шагом является полный отказ от маркеров, для чего потребуются ещё более сложная методика, так как отсутствует опорная точка в окружающем пространстве. В этом случае движение камеры будет определяться анализом перемещения наиболее характерных среди стабильных ключевых точек и для анализа формы будет необходимо не попарное сравнение кадров как описано выше, а сравнение 3-х и более кадров. Дополнением к этому является серьезное усложнение дескрипторов, а также сложность классификации ключевых точек, выделения в них множеств  $A$  и  $B$ . Правда, при наличии дополнительных устройств, например акселерометра, гироскопа и компаса данная задача может быть упрощена, а погрешность результатов заметно уменьшена, однако даже с их использованием задача остается сложной. Подход с отказом от маркера расширит применимость алгоритмов на основе этой методики, однако неизбежно уменьшит его точность и повысит требования к ресурсам.

Так же в качестве оценки перспектив следует учесть еще один аспект применения. При использовании технологии дополненной реальности важной стороной вопроса является удобство вывода информации для человека. В первую очередь это упирается в выбор типа устройства. Чаще всего выбор падет на смартфон или планшетный компьютер. Однако, использование очков виртуальной реальности во многих случаях может быть более актуально. Наиболее актуальный вариант заключается в использовании бюджетных моделей очков с вставкой смартфона с креплениями, не закрывающими камеру. Видеоряд с камеры и

наложенная на него дополнительная информация выводятся на экран. Минусами данного подхода являются неудобства самой маски очков и полное отсутствие обзора при зависании системы.

Следует так же учесть, что на вход приложения приходит только один видеопоток, а изображения на экране генерируются для каждого глаза отдельно и они должны как можно ближе совпадать с тем, что пользователь может увидеть каждым глазом в отдельности. Поэтому как отдельную проблему можно выделить реализацию стереоскопического зрения с использованием только 1 камеры. Одним из вариантов является использование карты глубины, полученной выше, смещение пикселей в зависимости от дальности и закраска “слепых” зон градиентами, сливающимися с остальным изображением. Более сложный подход заключается в закрашивании “слепых” зон изображениями с предыдущих кадров. Стереометрия сцены дополненной реальности рассчитывается тривиальными формулами, её реалистичное отображение зависит лишь от калибровки. Таким образом, появляется еще одно применение алгоритмов, основанных на данной методике, - реализация более реалистичной дополненной реальности на устройствах виртуальной реальности, оснащенных только одной камерой. Данный аспект важен так как не требует создания специальных очков с 2 камерами, которые практически не представлены на широком рынке. Конструкции с одной камерой наоборот широко представлены и обладают широким спектром цены/качества что позволяет использовать как экономичные варианты, так и варианты с лучшей производительностью, эргономичностью, надежностью в зависимости от поставленных целей и предоставленных ресурсов.

Описанные выше подходы открывают широкие возможности для практического применения и дальнейшего развития систем компьютерного зрения. Распознавание формы способно приблизить принцип анализа объектов производимый компьютером, к принципу анализа, происходящего в мозге человека, и приблизить возможности компьютерного распознавания к возможностям человеческого мозга.

### **Библиографический список**

1. Andrew Blake, Michael Isard. Active Contours, Luxembourg, Springer, 2000, 306 p.
2. Prince S.J.D. Computer Vision: Models, Learning, and Inference, Cambridge, Cambridge University Press, 2012, 600 p.
3. Scott Krig. Computer Vision Metrics: Survey, Taxonomy, and Analysis, Springer, Luxembourg, 2014, 418 p.
4. Asim Bhatti. Current Advancements in Stereo Vision, InTech, Rijeka, Croatia, 2012, 236 p.
5. Aapo Hyvarinen, Jarmo Hurri, Patrik O. Hoyer. Natural Image Statistics, Luxembourg, Springer, 2009, 448 p.
6. Xiong Zhihui. Computer Vision, InTech, Rijeka, Croatia, 2008, 548 p.
7. Robert Laganière. OpenCV 2 Computer Vision Application Programming Cookbook, Birmingham, Packt Publishing Ltd, 2011, 290 p.
8. Лунев Е.М., Неретин Е.С., Дяченко С.А., Дуброво А.И. Разработка программно-алгоритмического обеспечения прототипа системы синтетического видения для

перспективных объектов авиационной техники // Труды МАИ. 2016. № 86. URL:

<http://trudymai.ru/published.php?ID=66366>