

САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математическое обеспечение и администрирование
информационных систем

Системное программирование

Назаренко Владимир Владимирович

**Выделение объектов на
видеопоследовательности**

Выпускная квалификационная работа

Научный руководитель:
ст. преп. СМЕРНОВ М. Н.

Рецензент:
ПЕНКРАТ Н. А.
менеджер проектов, ООО “Ланит-Терком”

Санкт-Петербург
2018 г.

SAINT PETERSBURG STATE UNIVERSITY

Software and Administration of Information Systems
Software Engineering

Vladimir Nazarenko

Object detection in a video sequence

Master thesis

Scientific advisor:

sr. Lecturer MIKHAIL SMIRNOV

Reviewer:

NICKOLAY PENKRAT

Project Manager, Lanit-Tercom LLC

St Petersburg
2018

Содержание

1	Постановка задачи	6
2	Обзор	7
2.1	Основные определения	7
2.2	Карта глубины	7
2.2.1	Semi-global Matching	8
2.2.2	AD-Census	9
2.2.3	CVS	9
2.2.4	Сравнение алгоритмов построения карты глубины	10
2.3	Существующие подходы к детектированию препятствий с помощью стереокамеры	10
2.4	Подход Stixel World	11
2.5	Подходы к детектированию маркеров дорожной разметки	13
2.6	Подходы к валидации алгоритмов систем помощи водителю	13
3	Алгоритм поиска препятствий движению автомобиля	13
4	Алгоритм детектирования маркеров дорожной разметки	14
5	Апробация	14
5.1	Данные	14
5.2	Оценка качества работы алгоритмов	14
5.2.1	Методика оценки алгоритма детектирования препятствий	14
5.2.2	Результаты оценки алгоритма детектирования препятствий	15
5.2.3	Сравнение результатов с реализацией модели Stixel World без использования карт диспаратета	15
6	Заключение	15

Введение

На дорогах общего пользования происходит большое количество дорожно-транспортных происшествий (ДТП). Согласно исследованию [13] причиной большого количества ДТП является водитель, внимательность которого ослаблена в следствие усталости, приёма препаратов, выполнения задач, не связанных с управлением транспортным средством и других схожих факторов.

В настоящее время широкое распространение получили системы помощи водителю (ADAS). Такие системы, например, предупреждают водителя об ограничении скорости на участке дороге (с помощью распознавания соответствующих дорожных знаков), о пересечении маркеров дорожной разметки, об опасности столкновения с различными объектами. Существуют исследования, экспериментально доказывающие практическую полезность таких систем [14].

Важным элементом таких систем являются различные сенсоры. От типа сенсора в том числе зависит спектр алгоритмов, применимых для решения задач системы помощи водителю. Наиболее распространёнными сенсорами для решения перечисленных выше задач являются лидары, радары, ультразвуковые датчики и оптические системы видимого спектра: стерео- и монокамеры. Область применения каждого из типов сенсоров ограничена. Так, радары обладают низкой точностью определения формы и расстояния до объекта. Лидары обладают низкой точностью в плохих погодных условиях. Кроме того, высокая стоимость лидаров делает решения на их основе недоступными для массового сегмента автомобильной промышленности. Ультразвуковые датчики способны обнаруживать препятствия только на небольших расстояниях. Использование оптических систем требует использования сложных для определения расстояния до объектов.

Стереокамера сочетает в себе невысокую стоимость, возможность с высокой точностью определять геометрию и расстояние до объектов, простоту монтажа.

На кафедре Системного Программирования Санкт-Петербургского Государственного Университета в совместных исследовательских проектах с компанией Prosense, Южная Корея разрабатываются алгоритмы для системы помощи водителю. Одним из требований к этой системе является низкая стоимость и возможность установки системы без существенных модификаций конструкции автомобиля. Одной из частей этой системы является так называемый ***сенсор безопасного движения*** – подсистема, предупреждающая водителя о потенциальных столкновениях и пересечении маркеров дорожной разметки.

В данной работе мы сфокусировались на разработке и апробации алгоритмов для сенсора безопасного движения. В связи с требованиями к разраба-

тываемой системе, а именно ограничением на стоимость и простоту монтажа, в качестве сенсора мы выбрали стереокамеру, состоящую из двух откалиброванных камер видимого диапазона.

Существует, как минимум, два класса алгоритмов для решения задач помощи водителю, использующих оптические сенсоры: нейросетевые алгоритмы и алгоритмы на основе методов классического компьютерного зрения. В данной работе решено было использовать алгоритмы на основе классического компьютерного зрения. Связано это со следующими проблемами нейросетевых алгоритмов.

- Сложность модификации нейросетевых алгоритмов.
- Неуниверсальность нейросетевых алгоритмов.
- Сложность интерпретации решений нейросетевых алгоритмов.

Строго говоря, под предупреждением водителя о потенциальном столкновении мы понимаем детектирование на изображении **препятствий** – любых объектов, которые делают невозможным или опасным проезд через занимаемую ими область пространства. Типовыми примерами препятствий являются люди, автомобили, столбы, здания. Также мы считаем препятствиями особенности рельефа (холмы) и различные мелкие объекты, такие как бордюры. Слова “объект” и “препятствие” для нас являются синонимами. Под **детектированием препятствий** мы понимаем выделение препятствий на изображении в том или ином виде. Например, в виде описывающего прямоугольника или в виде области на изображении, движение в которой безопасно – **безопасной области**.

Также отметим, что термины “выделение объектов”, “детектирование объектов” и “сегментация изображения на объекты” мы считаем эквивалентными.

Под **детектированием маркеров дорожной разметки** мы понимаем задачу выделения на изображении таких маркеров, как одиночная сплошная линия, двойная сплошная линия, прерывистая линия, бордюры.

1 Постановка задачи

Целью данной работы является разработка и реализация, на основе подходов классического стерео-зрения и классической обработки изображений, алгоритмов для “сенсора безопасного движения”. Для достижения этой цели в рамках работы были сформулированы следующие задачи.

- Разработать и реализовать алгоритм поиска препятствий движению автомобиля на изображении, полученном с помощью стереокамеры.
- Разработать и реализовать алгоритм поиска маркеров дорожной разметки на изображении, полученном с помощью стереокамеры.
- Провести апробацию разработанных алгоритмов.

2 Обзор

2.1 Основные определения

Единственным сенсором, который мы используем для решения поставленных задач является стереокамера.

Стереокамера это система из двух камер, расположенных на небольшом расстоянии друг от друга, взаимное расположение и калибровка которых известны в любой момент времени. Ключевым свойством такой системы является возможность оценки расстояния до объектов на изображении за счёт параллакса.

Оптический поток – это отображение пикселей изображения в пару целочисленных значений, соответствующих движению объекта реального мира, спроецированного в пиксель, в плоскости камеры.

С использованием стереокамеры становится возможным вычислить **карту глубины** – отображение, сопоставляющее пикселю исходного изображения расстояние от оптического центра камеры до точки в пространстве, которая была спроецирована в данный пиксель. Карта глубины может быть плотной, в таком случае подразумевается, что большинству пикселей сопоставлено значение глубины, либо неплотной – значение глубины сопоставлено лишь некоторым пикселям. В случае неплотной карты глубины, как правило, значения глубины сопоставлены так называемым ключевым точкам – точкам, в которых на изображении имеется выраженный перепад яркостей.

В настоящее время алгоритмы построения карты глубины используют **ректифицированные изображения** – это пара изображений, к которым применено перспективное преобразование, модифицирующее исходную пару изображений, симулируя расположение плоскостей обеих камер в одной и той же плоскости.

Основным шагом вычисления карты глубины является вычисление диспаритета для каждого пикселя. **Диспаритет** это смещение между координатами пикселя, соответствующего одной и той же точке пространства на первом и втором сенсорах стереокамеры. Как правило, диспаритет вычисляется для пары ректифицированных изображений. В таком случае диспаритет – это одно число, смещение по горизонтальной оси.

2.2 Карта глубины

Построение карты глубины является популярным методом предобработки пары изображений, полученных с помощью стереокамеры, и используется

во многих работах, авторы которых решают задачу детектирования препятствий.

Существуют альтернативные методы предобработки для вычисления геометрии и расстояния до объектов на изображении [4][8], основанные на использовании одной камеры, однако они проигрывают построению карты глубины по снимку, полученному с помощью стереокамеры либо в точности, либо в качестве результатов. Поэтому в данной работе мы придерживаемся подхода на основе построения карты глубины.

Способ построения карты глубины критически важен для алгоритма детектирования препятствий, разработанного в рамках данной работы. Различные алгоритмы отличаются друг от друга как качеством, так и скоростью работы. Разработка собственного алгоритма построения карты глубины находится за пределами данной работы, однако нами были рассмотрены следующие реализации алгоритмов.

- Реализация алгоритма Semi-global Matching из открытой библиотеки компьютерного зрения OpenCV[7].
- Закрытая реализация алгоритма AD-Census.
- Закрытая реализация алгоритма вычисления разреженной карты глубины CVS.

2.2.1 Semi-global Matching

Алгоритм Semi-global Matching [6] (SGM) широко применяется для вычисления карты глубины.

Данный алгоритм минимизирует целевую функцию, заданную в выражении 1 на паре ректифицированных изображений, полученных с помощью стереокамеры.

$$E(D) = \sum_p [C(p, D_p) + \sum_{q \in N_p \wedge |D_p - D_q| = 1} P_1 + \sum_{q \in N_p \wedge |D_p - D_q| > 1} P_2] \quad (1)$$

Здесь p это координаты пикселя на левом изображении, $C(p, D_p)$ – значение функции схожести левого и правого изображения стереокамеры, P_1 и P_2 – константные штрафы, N_p – окрестность точки p на левом изображении, D_p и D_q – значения диспаратетов для пикселей p и q .

Особенностью алгоритма SGM является то, что минимизация целевой функции производится сразу по нескольким направлениям, как это изображено на рисунке 1.

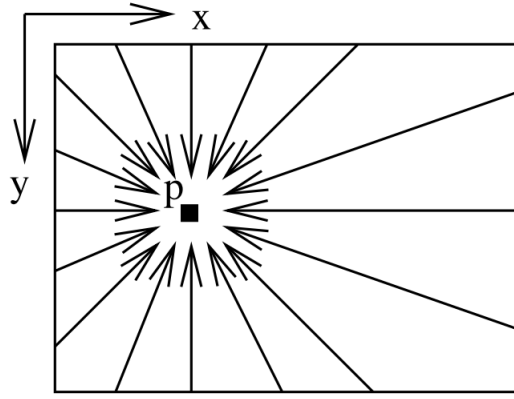


Рис. 1: Схема направлений оптимизации в алгоритме SGM [6]

В целях предобработки пары изображений, полученных с помощью стереокамеры, нами была использована общедоступная реализация алгоритма SGM из открытой библиотеки OpenCV. В данной реализации в качестве функции схожести $C(p, D_p)$ используется сумма абсолютных разниц (SAD) в окрестности пикселя.

2.2.2 AD-Census

Алгоритм AD-Census [11] является модификацией алгоритма SGM для систем массового параллелизма, таких как графические ускорители.

В отличие от алгоритма SGM, в алгоритме AD-Census используется функция схожести, основанная на взвешенном среднем двух компонент: абсолютной разности интенсивностей (компонента AD) и расстоянии Хэмминга (компонента Census). Кроме того, авторы предлагают ряд модификаций оригинального алгоритма SGM для практического уменьшения количества требуемых вычислений.

В целях предобработки пары изображений, полученных с помощью стереокамеры, нами была использована закрытая реализация алгоритма AD-Census, выполненная инженерами компании Ланит-Терком.

2.2.3 CVS

Алгоритм CVS является запатентованной разработкой компании Ланит-Терком. Данный алгоритм позволяет строить разреженную карту глубины, где значения глубины сопоставляются ключевым точкам – пикселям на изображении, соответствующим выраженным перепадам яркости на изображении.

2.2.4 Сравнение алгоритмов построения карты глубины

TODO: Изображение – качественное сравнение алгоритмов расчёта карты глубины

2.3 Существующие подходы к детектированию препятствий с помощью стереокамеры

Существует большое количество исследований в области детектирования неклассифицированных объектов с помощью стереокамеры. В рамках работы было проведено изучение соответствующих работ. Перечислим далее работы, наиболее релевантные нашей цели.

Авторы [5] предложили использовать геометрическую зависимость между оптическим потоком и расстоянием до объекта, вычисленным с помощью алгоритмов построения карты глубины. Как отмечают авторы, данный подход неустойчив к движению камер в плоскости кадра и к неточностям алгоритмов расчёта оптического потока и карты глубины. К плюсам предложенного решения можно отнести высокую скорость работы, без учёта вычисления оптического потока и расчёта карты глубины.

В работе [9] авторы представили способ определения рельефа дорожной поверхности без использования карты глубины. Также авторы показали принципиальную возможность выделять отдельные препятствия с помощью представленного ими алгоритма. Однако точность алгоритма невысока, особенно в случае наличия большого количества препятствий на изображении. Тем не менее, способ динамического определения положения дорожного полотна, предложенный в этой работе, широко используется. В том числе, он был использован в нашем исследовании.

Развивая подход Labayrade [9], авторы [15] предлагают улучшения для алгоритма, позволяющие применять алгоритм в ситуациях бездорожья. Тем не менее авторы приводят только качественную оценку своего алгоритма, из которой следует вывод, что алгоритм применим только в ситуациях с небольшим количеством препятствий на изображении, как и подход Labayrade [9].

В работе [1] предложен подход выделения препятствий на изображении, основанный на расчёте неплотного оптического потока с помощью KLT-трекера и фильтра Калмана и расчёте неплотной карты глубины. Авторы данной работы не приводят способа выделения отдельных объектов на изображении, также в данной работе не рассмотрено выделение препятствий, не имеющих собственного движения.

В статье [12] авторы предлагают вместо вычисления карты глубины сегментировать изображение на вертикальные полосы, для каждой из которых

вычисляются две горизонтальных границы, соответствующие основанию ближайшего препятствия и его высоте. Затем, используя эту сегментацию, можно выделить объекты не прибегая к сложным вычислениям. Данный подход показался нам перспективным и многие идеи нашего исследования были заимствованы из данной работы, поэтому опишем этот алгоритм подробно.

2.4 Подход Stixel World

Опишем подход, предложенный в статье [12], подробнее, так как наша работа основана на данном подходе.

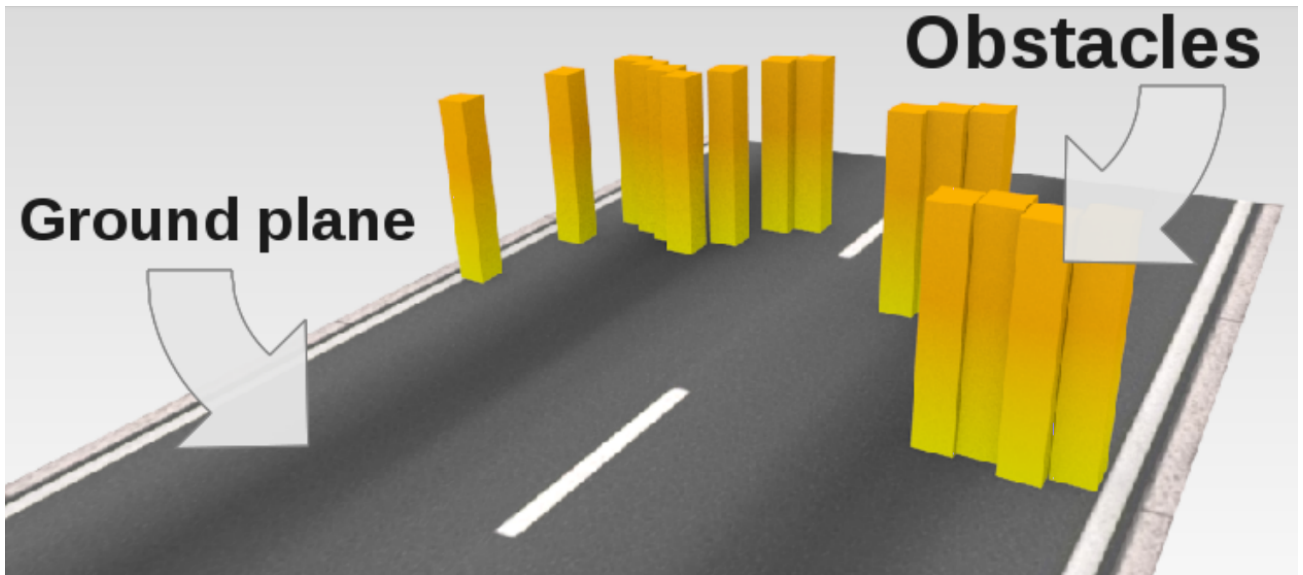


Рис. 2: Упрощённое представление мира в модели Stixel World [2]

В модели Stixel World препятствия описываются упрощённой моделью: предполагается, что любое препятствие может быть хорошо приближено параллелепипедом. В таком случае на изображении препятствия могут быть хорошо приближены набором вертикальных прямоугольников с сопоставленными им значениями глубины. На рис. 2 изображена схема упрощённой модели препятствий, принятой в модели Stixel World.

Таким образом, модель Stixel World предполагает сегментацию изображения на стиксели – прямоугольники фиксированной ширины. Каждому стикселю соответствует 4 следующих значения.

1. Индекс стикселя.
2. Координата строки изображения – нижняя граница стикселя.
3. Координата строки изображения – верхняя граница стикселя.

4. Значение расстояния. При этом предполагается, что расстояние одинаково до всех пикселей, принадлежащих стикселю.

Пример сегментации изображения, получаемый с использованием модели Stixel World, представлен на рис. 3.

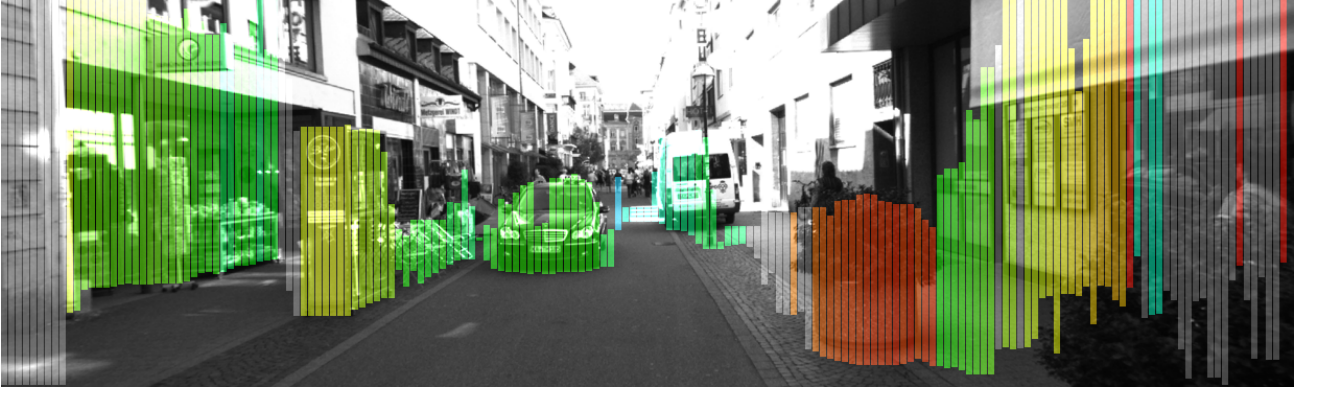


Рис. 3: Пример сегментации препятствий с помощью стикселей. Цветом закодировано расстояние.

Для построения модели Stixel World по двум изображениям, полученным с помощью стереокамеры, авторы предлагают минимизировать сумму двух целевых функций.

$$d_s^*(u) = \operatorname{argmin}_{d(u)} \sum_u c_s(u, d(u)) + \sum_{u_a, u_b} s_s(d(u_a), d(u_b)) \quad (2)$$

$$c_s(u, d) = c_o(u, d) + c_g(u, d) \quad (3)$$

$$c_o(u, d) = \sum_{v=v(h_o, d)}^{v(d)} c_m(u, v, d) \quad (4)$$

$$c_g(u, d) = \sum_{v=v(d)}^{|V|} c_m(u, v, f_{ground}(v)) \quad (5)$$

$$s_s(d_a, d_b) = \begin{cases} \infty, & \text{если } d_a < d_b - 1 \\ c_o(u_a, d_a) & \text{если } d_a = d_b - 1 \\ 0, & \text{если } d_a > d_b - 1 \end{cases} \quad (6)$$

В формулах 2 – 6 приняты следующие обозначения.

u – порядковый номер стикселя.

v – индекс строки изображения.

d_s^* – вычисленное значение диспаратета для стикселя под номером u .

h_0 – минимальная высота объекта на сцене, в метрах.

$c_m(u, v, d)$ – значение функции схожести в точке (u, v) для диспаратета d .

$f_{ground}(v)$ – диспаратет плоскости дороги для строки v изображения.

u_a, u_b – индексы соседних стикселей.

Также авторы предлагают способ оценки высоты объектов с помощью оценки высоты стикселей. Для этого вводится схожая целевая функция, которую мы опустим для краткости.

Такая целевая функция допускает минимизацию на ректифицированной паре изображений с помощью техник динамического программирования, что и было предложено авторами модели Stixel World.

Данный подход требует значительно больше вычислительных ресурсов, в отличие от предыдущих, однако авторы [2] сообщают, что алгоритм может быть оптимизирован для работы в реальном времени.

2.5 Подходы к детектированию маркеров дорожной разметки

TODO

2.6 Подходы к валидации алгоритмов систем помощи водителю

TODO

3 Алгоритм поиска препятствий движению автомобиля

В данном разделе описан предложенный и реализованный в виде прототипа алгоритм поиска препятствий с помощью стереокамеры. Основные идеи данного подхода взяты из модели Stixel World [12].

В предлагаемом нами алгоритме мы, вслед за авторами работы [12], представляем сцену, видимую на пара изображений, полученных с помощью стереокамеры, в виде модели Stixel World. Однако, в отличие от [12], вычисление параметров модели происходит не на паре изображений, а на карте диспаратетов.

Такая модификация ведёт к следующей модификации целевых функций.

TODO

4 Алгоритм детектирования маркеров дорожной разметки

TODO

5 Апробация

5.1 Данные

Для оценки качества работы алгоритма детектирования препятствий, были выбраны следующие наборы данных.

- Тестовый набор для алгоритмов детектирования объектов KITTI [3].
- Тестовый набор для алгоритмов детектирования объектов Lost and Found [10].

Набор данных KITTI состоит из приблизительно 8000 изображений, полученных с помощью стереокамеры, установленной на крыше автомобиля. Изображения были получены в рамках движения автомобиля по дорогам общего пользования Германии. Каждое изображение проаннотировано прямоугольниками, описанными вокруг объектов из следующих классов: велосипедист, пешеход и автомобиль.

TODO

5.2 Оценка качества работы алгоритмов

5.2.1 Методика оценки алгоритма детектирования препятствий

Так как выбранные наборы данных содержат аннотацию в виде описывающих прямоугольников, было решено применить следующий подход оценки.

1. На изображении оцениваются параметры модели Stixel World.
2. Для каждого описывающего прямоугольника из аннотации из всех стикселей выбираются те, горизонтальная координата которых заключена между левой и правой границей прямоугольника.
3. Для выбранных стикселей вычисляется медиана нижней границы.
4. Отличие вычисленной медианы от нижней границы описывающего прямоугольника из аннотации принимается метрикой качества для данного изображения.

5. Метрики качества агрегируются для всей коллекции.

TODO

5.2.2 Результаты оценки алгоритма детектирования препятствий

TODO

5.2.3 Сравнение результатов с реализацией модели Stixel World без использования карт диспаратета

TODO

6 Заключение

В рамках данной работы были достигнуты следующие результаты.

- На основе подхода Stixel World и классического стереозрения разработан и релизован на языке C++ прототип алгоритма решения задачи поиска препятствий движению автомобиля на видеопоследовательности, полученной со стереокамеры, закреплённой на лобовом стекле автомобиля.
- На основе методов классической обработки изображений разработан и реализован на языке Python прототип алгоритма поиска маркеров дорожной разметки на изображении, полученном с камеры, закреплённой на лобовом стекле автомобиля.
- Выполнена апробация разработанных алгоритмов на наборах данных KITTI и tuSimple, а также на собственных данных.

Список литературы

- [1] 6d-vision: Fusion of stereo and motion for robust environment perception / Uwe Franke, Clemens Rabe, Hernán Badino, Stefan Gehrig // Joint Pattern Recognition Symposium / Springer. — 2005. — P. 216–223.
- [2] Benenson Rodrigo, Timofte Radu, Van Gool Luc. Stixels estimation without depth map computation // Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on / IEEE. — 2011. — P. 2010–2017.
- [3] Geiger Andreas, Lenz Philip, Urtasun Raquel. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite // Conference on Computer Vision and Pattern Recognition (CVPR). — 2012.
- [4] Godard Clément, Mac Aodha Oisin, Brostow Gabriel J. Unsupervised Monocular Depth Estimation with Left-Right Consistency // CVPR. — 2017.
- [5] Heinrich Stefan. Fast obstacle detection using flow/depth constraint // Intelligent Vehicle Symposium, 2002. IEEE / IEEE. — Vol. 2. — 2002. — P. 658–665.
- [6] Hirschmuller Heiko. Accurate and efficient stereo processing by semi-global matching and mutual information // Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on / IEEE. — Vol. 2. — 2005. — P. 807–814.
- [7] Itseez. Open Source Computer Vision Library. — <https://github.com/itseez/opencv>. — accessed in July, 2017.
- [8] Koenderink Jan J, Van Doorn Andrea J. Affine structure from motion // JOSA A. — 1991. — Vol. 8, no. 2. — P. 377–385.
- [9] Labayrade Raphael, Aubert Didier, Tarel J-P. Real time obstacle detection in stereovision on non flat road geometry through” v-disparity” representation // Intelligent Vehicle Symposium, 2002. IEEE / IEEE. — Vol. 2. — 2002. — P. 646–651.
- [10] Lost and Found: detecting small road hazards for self-driving vehicles / Peter Pinggera, Sebastian Ramos, Stefan Gehrig et al. // Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on / IEEE. — 2016. — P. 1099–1106.

- [11] On building an accurate stereo matching system on graphics hardware / Xing Mei, Xun Sun, Mingcai Zhou et al. // Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on / IEEE. — 2011. — P. 467–474.
- [12] Pfeiffer David, Franke Uwe. Efficient representation of traffic scenes by means of dynamic stixels // Intelligent Vehicles Symposium (IV), 2010 IEEE / IEEE. — 2010. — P. 217–224.
- [13] Staubach Maria. Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems // Accident Analysis & Prevention. — 2009. — Vol. 41, no. 5. — P. 1025–1033.
- [14] Studying effects of advanced driver assistance systems (ADAS) on individual and group level using multi-driver simulation / Christian Maag, Dominik Muhlbacher, Christian Mark, Hans-Peter Kruger // IEEE Intelligent Transportation Systems Magazine. — 2012. — Vol. 4, no. 3. — P. 45–54.
- [15] The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA grand challenge on TerraMax / Alberto Broggi, Claudio Caraffi, Pier Paolo Porta, Paolo Zani // Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE / IEEE. — 2006. — P. 745–752.