

Алгоритмы стереозрения на основе параллакса движения монокулярной камеры бокового обзора¹

Е.И. Ершов, В.Н. Карнаухов, М.Г. Мозеров

Институт проблем передачи информации, Российская академия наук, Москва, 127051, Россия

Поступила в редколлегию 24.11.2015

Аннотация—В данной работе предлагается оценка применимости алгоритмов стереосопоставления для случая параллакса движения монокулярной камеры. Предложены теоретические оценки искажения степени эпиполярных ограничений в зависимости от относительного поворота систем координат снимков. Разработан алгоритмический способ оценки качества вычисляемой карты диспаратности. Описан новый алгоритм стереосопоставления для случая параллакса монокулярной камеры.

КЛЮЧЕВЫЕ СЛОВА: стереозрение, стереосопоставление, латеральное стерео, стереодвижение.

ВВЕДЕНИЕ

Идея использовать камеру бокового обзора робота для определения глубины сцены достаточно привлекательна, так как позволяет улучшить параметры навигации. Кроме того, наличие информации о глубине позволяет в некоторых случаях определить форму объектов сцены, и как следствие помочь распознаванию этих объектов [1, 2]. Идея не нова и была предложена, например, в статьях [3, 4]. Класс подобных задач был назван стереодвижением. Стоит отметить, что, помимо вышесказанного, задачи стереодвижения интересны авторам в контексте ряда задач, подразумевающих использование монокулярной камеры: в задаче группового патрулирования наземных роботов оснащенных боковыми камерами [5], в задаче определения собственного движения БПЛА с заранее известной картой [6], а так же в задаче автоматической классификации транспортных средств [7]. Чтобы понять, в чём новизна нашего настоящего исследования рассмотрим кратко некоторые ключевые моменты известных решений задач этого класса. Основное допущение: если камера движется прямолинейно и её оптическая ось перпендикулярна вектору движения, то в этом случае два положения камеры во времени могут быть рассмотрены для неподвижной сцены, как хорошее приближение к эпиполярному стерео (см. Рис. 1(а)). Напомним, что эпиполярное стерео представляет собой задачу стереосопоставления, исходящую из ключевого предположения, что оптические оси правого и левого объектива стереосистемы перпендикулярны линии соединяющей оптические центры (база оптической системы B), кроме того все три линии (база и оптические оси) лежат в одной плоскости, а вертикальные координатные оси обеих оптических камер перпендикулярны линии базы.

Чем же привлекательно эпиполярное стерео? Все современные алгоритмы стереосопоставления (а это тысячи опубликованных методов) работают в подавляющем числе случаев в рамках эпиполярной модели [8, 9]. Эпиполярная модель позволяет: а) легко вычислять глубину точек трёхмерной сцены используя только значения диспаратностей; б) упрощает задачу двумерного сопоставления к одномерному; в) существуют робастные методы калибровки стерео

¹ Исследование выполнено за счет гранта Российского научного фонда (проект № 14-50-00150)

системы, позволяющие всего один раз настроить блок проективного преобразования входных стереоизображений к эпиполярным. Эпиполярность системы нарушается при движении робота, обусловленном с поворотом камеры вокруг одной из трёх осей системы координат, связанной с камерой (см. Рис.1(б).)

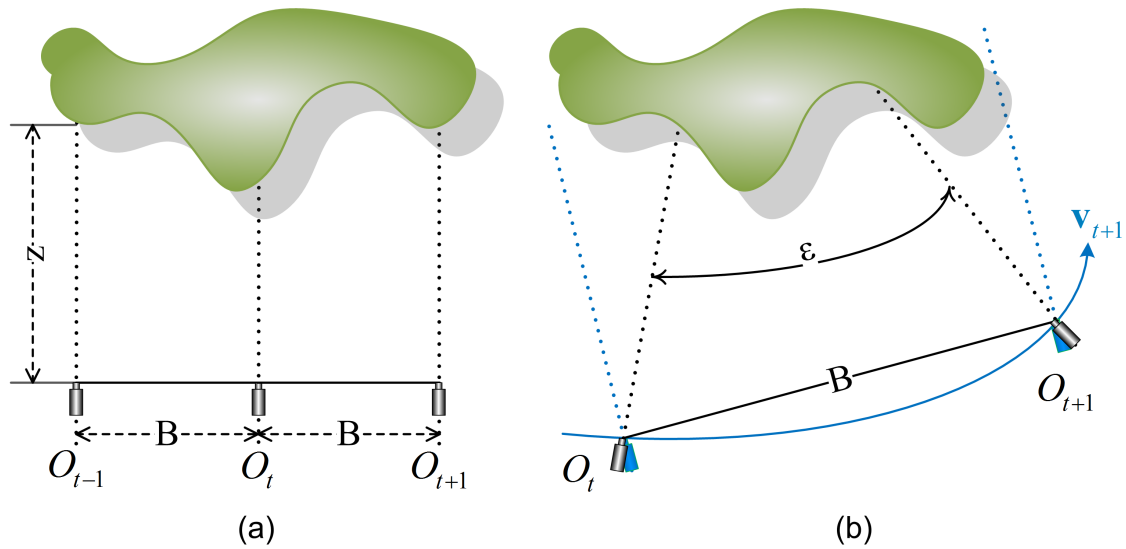


Рис. 1. Схема стереодвижения: (а) эпиполярное приближение; (б) поворот оптической оси камеры.

Отметим, что у стереодвижения есть свои определённые преимущества по сравнению с классическим стерео: если частота кадров большая, можно варьировать как длину базы, так и число кадров, используемых в стереосопоставлении. Такой подход даёт возможность преодолевать проблему окклюзий, характерную для классического подхода. Однако не всегда стереодвижения являются хорошим приближением к эпиполярной модели. Поэтому в этой работе мы поставили три следующие цели:

1. Оценить в каких случаях допустимо использовать эпиполярное приближение для решения задач этого класса.
2. Предложить и оценить возможность простой и быстрой ректификации стереоизображений с помощью матрицы поворота вместо фундаментальной матрицы. Предложенная техника является новой, что обусловлено спецификой задачи. В самом деле, одноразовое вычисление фундаментальной матрицы для оптической стереосистемы обычно не является проблемой. С другой стороны, оценка матрицы поворота часто не требует (в отличие от фундаментальной матрицы) нахождения точек соответствия на самих изображениях, а вычисляется с помощью навигационных гироскопических устройств.
3. Выбрать метод классического стереосопоставления и адаптировать его к стереодвижениям с учётом возможности получения более чем двух изображений.

Статья организована следующим образом: в разделе 1 получены теоретические оценки максимальной точности восстановления карты диспаратности для случая ненулевого относительного угла между системами координат камер и предлагается алгоритмический способ оценки качества карты диспаратности, в разделе 2 описан новый алгоритм стереосопоставления для случая параллакса движения монокулярной камеры, в разделе 3 приведены экспериментальные результаты по исследованию применимости алгоритмов стереосопоставления для случая параллакса движения монокулярной камеры и, наконец, в заключении приведено краткое описание полученных результатов.

1. ОЦЕНКА ЭПИПОЛЯРНОГО ПРИБЛИЖЕНИЯ

Чтобы понять, как поворот камеры при движении может исказить оценку диспаратностей рассмотрим более подробно, что такое эпиллярная модель. Схема эпиллярного стерео иллюстрируется на Рис. 2. Условие совмещения реальной точки трёхмерной сцены для двух

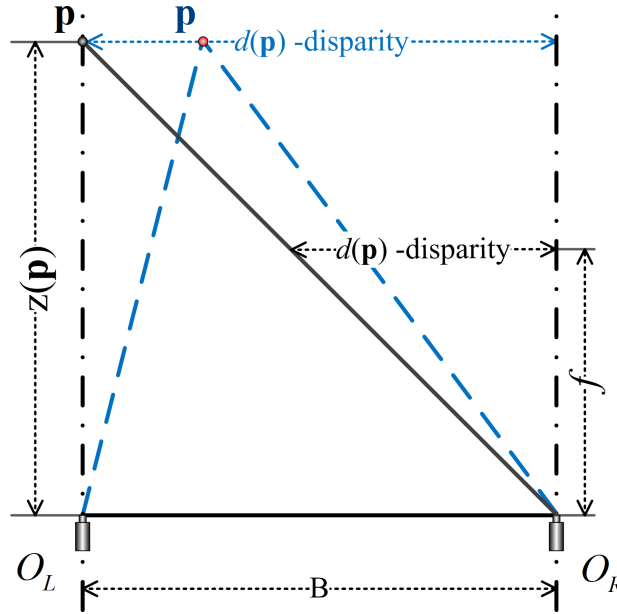


Рис. 2. Схема эпиллярного стерео.

проекций этой точки на левом и правом стереоизображениях означает:

$$x_L(\mathbf{p}) = x_R(\mathbf{p}) - d(\mathbf{p}), \quad (1)$$

или

$$d(\mathbf{p}) = x_R(\mathbf{p}) - x_L(\mathbf{p}), \quad (2)$$

где $x_R(\mathbf{p})$ и $x_L(\mathbf{p})$ – координаты по оси X точки \mathbf{p} на левом и правом стереоизображении соответственно и $d(\mathbf{p})$ – диспаратность. Тогда из Рис. 2 следует, что по оси X соотношение глубины и диспаратности является инвариантным:

$$\frac{z(\mathbf{p})}{B} = \frac{f_{pix}}{d_{pix}(\mathbf{p})}, \quad (3)$$

или

$$z(\mathbf{p}) = \frac{f_{pix}B}{d_{pix}(\mathbf{p})}. \quad (4)$$

где f_{pix} – фокусное расстояние в пикселях и $d_{pix}(\mathbf{p})$ – диспаратность в пикселях. Иными словами, вычисляемая глубина физической точки сцены не зависит от координат проекции этой точки, а только от разности этих координат по оси X .

Что бы понять соотношение измеряемых величин для реальных изображений, допустим для примера, что размер изображения по координате равен 640 пикселям, а угол обзора камеры

составляет 60 градусов. В этом случае фокусное расстояние будет приблизительно равно 555 пикселям. Действительно из Рис. 2 следует что

$$f_{pix} = \frac{X_{pix}}{2} \operatorname{ctg} \frac{\varphi}{2} = \sqrt{3} \times 320 \approx 555.$$

Из Рис. 3 и уравнения (4) видно, что минимальная глубина для стереосистемы при угле обзора в 60 градусов приблизительно равна базе стереосистемы. Однако в этом случае эффективная апертура измерения сокращается до щели шириной в один пиксел, а максимальная диспаратность увеличивается до горизонтального размера изображения. Практически все современные

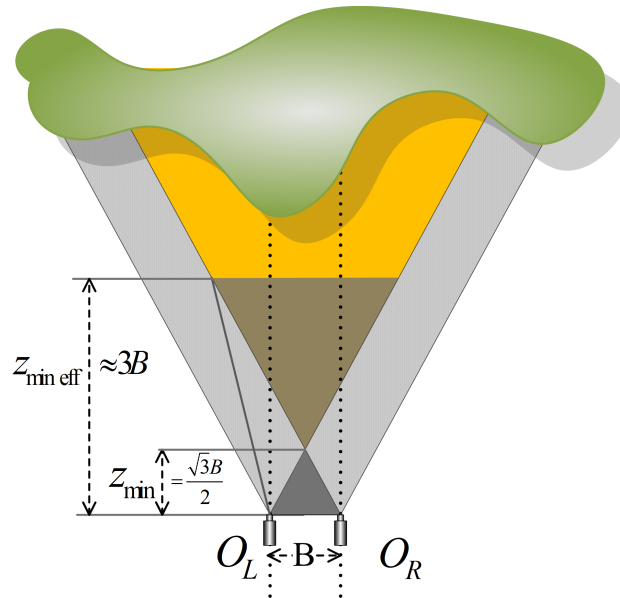


Рис. 3. Рабочий диапазон глубин z для оптической стереосистемы.

алгоритмы стерео ориентируются на базы данных, где эффективная апертура больше или равна 75% горизонтального размера изображения, что соответствует максимальному значению диспаратности в $X/4$. То есть удобно говорить об оценке минимальной эффективной глубины $z_{min\ eff}$, которая обычно превышает расстояние в три базы $z_{min\ eff} > 3B$ (см. Рис. 3). Из выражения (4) также следует, что $z_{max} = f_{pix}B$, то есть максимальная оценка глубины достигает $555B$ для нашего примера, однако ошибка, что следует из выводов ниже, того же порядка. Из выражения (4) следует также, что ошибка измерения глубины обратно пропорциональна квадрату диспаратности:

$$err(z) = dz = \frac{f_{pix}B}{d_{pix}^2}, \quad (5)$$

То есть диапазон ошибок (для нашего примера) может варьироваться от $555B$ до $0.02B$. Этот факт делает невозможным разумную оценку выраженную в терминах глубины. Однако в классическом стерео принято оценивать ошибку по результатам вычисления диспаратности и сравнения с истинными значениями. Чтобы решить нашу первую задачу – оценку применимости эпиполярной модели к стереодвижениям – преобразуем схему, показанную на Рис. 1(б), к виду, представленному на Рис. 4. Если допустить, что смещение $dz(O_t) \ll z_{min\ eff}$ и виртуальная база B' оптической системы $O_t - O_{t+1}$ почти равны реальному вектору смещения B ,

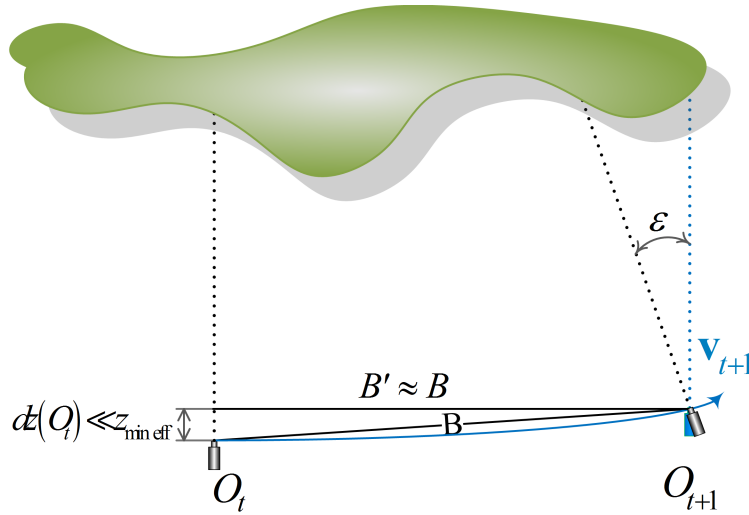


Рис. 4. Схема невязки стереодвижения относительно эпиполярной модели в зависимости от угла поворота.

то и искажения, связанные с неэпиполярностью системы, обусловлены только искажающим поворотом ε .

Для того чтобы оценить границы применимости эпиполярной модели определим два критерия точности восстановления карты диспаратности для ненулевого угла ε : теоретический и алгоритмический.

Теоретическая оценка точности является среднеквадратичной ошибкой от известных значений диспаратности, обусловленной действием искажающего оператора поворота, и выражается следующим образом:

$$ET(\varepsilon) = \sqrt{\frac{1}{|\Omega|} \sum_{x,y \in \Omega} (d^{GT}(x,y) - P^\varepsilon(d^{GT}(x,y)))^2}, \quad (6)$$

где Ω – область определения изображения, $d^{GT}(x,y)$ – известные значения диспаратностей относительно идеального (не повернутого) изображения, P^ε – искажающий оператор диспаратностей, смысл которого понятен из Рис. 5. Здесь $\mathbf{x}_L = [x, y, f]^T$ – координата проекции физической точки сцены \mathbf{p} на плоскость левого изображения, $\mathbf{x}_R = \mathbf{x}_L - d_L(\mathbf{x}_L)$ – координата проекции той же точки сцены на плоскость правого изображения, где $d^{GT}(\mathbf{x}_L)$ – известное значение диспаратности проекций точки \mathbf{p} на правое и левое изображения. Координаты проекции точки \mathbf{p} на повернутое правое изображение выражены:

$$\bar{\mathbf{x}}'_R = \mathbf{x}'_R \frac{f}{z'_R}; \quad \mathbf{x}'_R = \mathbf{R}(\varepsilon) \mathbf{x}_R; \quad (7)$$

где $\mathbf{R}(\varepsilon)$ – матрица поворота. Далее для простоты мы будем рассматривать только матрицы поворота вокруг одной из осей системы координат, связанной с движущейся камерой. Величина δd_L , по сути, определяет ошибку оценки диспаратности, связанную с искажающим оператором P^ε и вычисляется как:

$$\delta d_L(\mathbf{x}_L) = d_L(\mathbf{x}_L) - P^\varepsilon(d_L(\mathbf{x}_L)) = (\bar{\mathbf{x}}'_R - \mathbf{x}_R, \mathbf{X}) \quad (8)$$

В формуле (8) мы делаем допущение, что существует идеальный алгоритм, способный найти точное соответствие координат точки \mathbf{p} на обоих стереоизображениях. Поэтому этот критерий

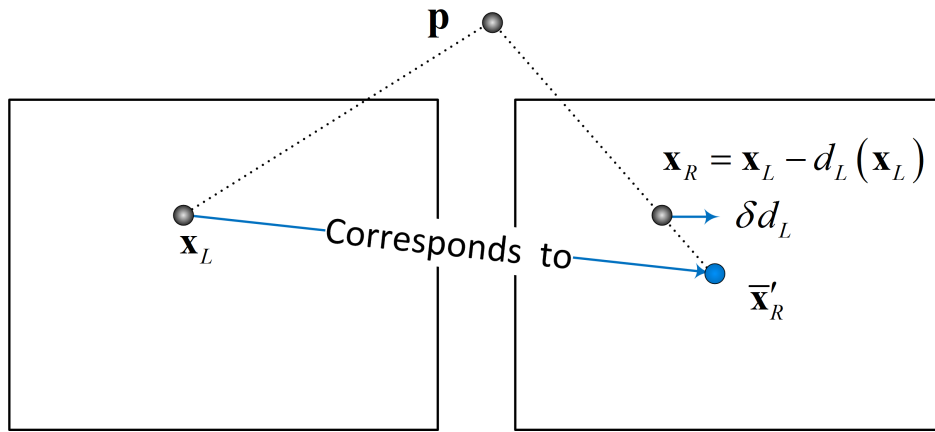


Рис. 5. Схема влияния поворота на ошибку вычисления диспаратности с использованием эпиполярной модели.

не требует алгоритмической верификации. Иными словами, если некоторое значение поворота приводит к неприемлемой точности даже для идеального алгоритма, значит, необходима коррекция изображения. Если допустить, что каждая точка на одном изображении соответствует какой-нибудь точке на другом (и это в целом верно), то теоретическая ошибка $ET(\varepsilon)$ не зависит от известных значений диспаратностей, а только от матрицы поворота:

$$ET(\varepsilon) = \sqrt{\frac{1}{|\Omega|} \sum_{x,y \in \Omega} (\mathbf{x} - \mathbf{R}(\varepsilon) \mathbf{x})^2} \tag{9}$$

Более того, для малых углов формула (9) может быть аппроксимирована линейной зависимостью от угла:

$$ET(\varepsilon) \approx \frac{X\varepsilon}{85}, \tag{10}$$

где X – горизонтальный размер изображения в пикселях, а угол измеряется в градусах. Так как оценка (10) исходит из некоего идеального алгоритма, разумно ввести ещё и критерий, способный определять точность работы алгоритма для искажённой (не эпиполярной) серии изображений. Этот критерий мы назовём ошибкой алгоритма:

$$EA(\varepsilon) = \sqrt{\frac{1}{|\Omega|} \sum_{x,y \in \Omega} (d^{EP}(x,y) - d^D(x,y))^2}, \tag{11}$$

где $d^{EP}(x,y)$ – результат работы предложенного алгоритма стереосопоставления для эпиполярного приближения и $d^D(x,y)$ – результат работы того же алгоритма при условии что правое изображение соответствует некоторому искажающему оператору поворота камеры. Другими словами, критерий показывает, в какой степени ухудшается результат работы предложенного алгоритма в случае поворота камеры. Этот критерий будет использоваться в экспериментальной части этой работы. Если угол искажающего поворота приводит к неприемлемой ошибке, предлагается использовать матрицу поворота обратную к искажающей $\mathbf{R}^{-1}(\varepsilon)$, чтобы скорректировать повернутое изображение и привести последовательность кадров к эпиполярной модели.

2. АЛГОРИТМ МОНОКУЛЯРНОГО СТЕРЕОЗРЕНИЯ

Прежде чем предложить алгоритм стереосопоставления для случая не эпиполярной модели необходимо отметить следующее. Во-первых, результаты лучших алгоритмов для эпиполярной модели, предложенных за последние пятнадцать лет, обобщены в таблице университета Миддлбёри на сайте <http://vision.middlebury.edu/stereo/eval/>. Во-вторых, результаты моделирования не эпиполярного стерео для лучших алгоритмов дают практически одинаковую ошибку от угла поворота по критерию (11). Более того, эта ошибка для малых углов практически совпадает с линейной зависимостью теоретической ошибки критерия (10) с тем же коэффициентом. Из выше сказанного можно сделать вывод, что лучший алгоритм для эпиполярного стерео будет лучшим и для не эпиполярной модели. Поэтому в статье за основу выбран алгоритм, который показывает лучший результат из опубликованных в таблице Миддлбёри [10]. В современном стереосопоставлении существуют два следующих основных подхода: с использованием локальных окон поддержки и глобальной оптимизации максимума апостериорной вероятности. Методы локальной фильтрации используют различные виды окрестностей и наиболее эффективные работают в так называемых окрестностях совмещённых цвет-изображение пространствах (билатеральный фильтр) [11–15]. Методы глобальной оптимизации [16–19] стали прорывом для алгоритмов стереозрения в начале двухтысячных годов, однако методы глобальной оптимизации плохо работают в областях с большой окклюзией. Выбранный метод [10] объединяет оба подхода эпиполярного стерео, что позволяет улучшить результат реконструкции в окклюзивных областях сцены. Тем не менее, проблема окклюзий не может быть решена полностью для стерео с двумя изображениями. Поэтому есть возможность использовать естественное преимущество стереодвижений – наличие трёх и более кадров последовательно, которые можно рассматривать, как тройное эпиполярное стерео [20]. Чтобы понять суть предложенного метода сформулируем сначала задачу реконструкции в терминах глобальной оптимизации. Чтобы получить 3D информацию реконструируемой сцены необходимо найти соответствие между пикселями на двух или более изображениях $\mathbf{f}^t(\mathbf{x}_p)$. Предполагается, что удовлетворяется условие оптического потока:

$$\mathbf{f}^t(\mathbf{x}_p) \stackrel{\Delta}{=} \mathbf{f}^{t+1}(\mathbf{x}_p + \mathbf{l}_p), \quad (12)$$

где \mathbf{f}^t – изображение, полученное в момент времени t , \mathbf{x}_p – координата пикселя p в области изображения, \mathbf{l}_p – 2D вектор диспаратности и знак $\stackrel{\Delta}{=}$ означает *соответствует*. Другими словами, значения интенсивностей в двух точках не только должны совпадать, но и координаты этих пикселей $\mathbf{x}_p + \mathbf{l}_p$ and \mathbf{x}_p в плоскостях двух разных изображений также должны соответствовать одной реальной точке трёхмерной сцены. Заметим, что функция $l(\mathbf{x}_p)$ является искомой в стереосопоставлении и называется картой диспаратностей. В рамках подхода глобальной оптимизации, задача стереосопоставления формулируется в терминах минимизации энергии (что соответствует максимизации апостериорной вероятности) с энергией выраженной в следующей форме:

$$E(l) = \sum_{p \in \mathcal{V}} u_p(l_p) + \sum_{(p,q) \in \mathcal{E}} B_{p,q}(l_p, l_q), \quad (13)$$

где множество $p \in \mathcal{V}$ соответствует пикселям изображения, а множество $(p, q) \in \mathcal{E}$ границам графа изображения $\mathcal{G} = (\mathcal{E}, \mathcal{V})$; l_p означает классификационную метку пикселя p , который, в свою очередь, принадлежит множеству значений функции диспаратности $l \in L$; $u_p(\cdot)$ определяет одинарный потенциал, соответствующий общепринятой цене несовпадения; $B_{p,q}(\cdot, \cdot)$ бинарный потенциал гладкости, который определяет влияние соседних пикселей через данную границу графа изображения (p, q) . Так как в этой статье предполагается, что два последовательных изображения соответствуют эпиполярной модели, то в (12) имеем $\mathbf{l}_p = \{l_p^x, l_p^y\} =$

$\{l_p^x, 0\} = l_p$. Выбор одинарного потенциала – важная задача стереосопоставления. Есть две разные категории цены: поточечная цена рассогласования и значение непараметрического преобразования [21] или значение кросскорреляции [22]. Используя комбинацию этих значений, можно улучшить результат реконструкции [23]. Метод оптимизации через билатеральный фильтр предложенного алгоритма подразумевает локальную окрестность в пространстве цветности-изображения, поэтому метод использует комбинацию двух поточечных потенциалов: $u_p^I(l)$ – потенциал между двумя кадрами последовательности $u_p^G(l)$ и потенциал между двумя изображениями градиентов этих кадров:

$$u_p(l) = \sum_{c \in C} |f_c^t(\mathbf{x}_p) - f_c^{t+1}(\mathbf{x}_p - l)| + \sum_{d \in \pm 1} |g_d^t(\mathbf{x}_p) - g_d^{t+1}(\mathbf{x}_p - l)|, \quad (14)$$

где изображение градиентов вычисляется как: $g_{\pm 1}^t(\mathbf{x}_p) = \sum_{c \in C} f_c^t(\mathbf{x}_p \pm 1) - \sum_{c \in C} f_c^t(\mathbf{x}_p)$. Данный выбор мотивирован статьями [24, 25], где такая комбинация поточечных потенциалов показала высокую робастность к изменениям освещения в кадрах. Отметим, что знак (-) перед значением диспаратности l в (14) предполагает, что движение камеры от кадра t к кадру $t+1$ такое же, как при смещении позиции камер левого и правого изображений в стереопаре. Предложенная в данной статье идея вычислять одинарный потенциал через выбор наиболее вероятного сопоставления была мотивирована статьями по синтезу промежуточных кадров. Если мы возьмём три последовательных кадра $\{\mathbf{f}^{t-1}, \mathbf{f}^t, \mathbf{f}^{t+1}\}$, тогда кадр \mathbf{f}^t можно рассматривать, как синтезированный между двумя известными. Известно [26], что в этом случае диспаратность некоторой точки центрального кадра обладает зеркальной симметрией относительно соседних:

$$l_p^{t \rightarrow t+1} = -l_p^{t \rightarrow t-1} \quad (15)$$

Таким образом, можно решить две задачи оптимизации (13) одновременно:

$$E(l) = \sum_{p \in V} u_p(l_p^{t \rightarrow t+1}) + \sum_{p \in V} u_p(-l_p^{t \rightarrow t-1}) + 2 \sum_{(p,q) \in E} B_{p,q}(l_p, l_q), \quad (16)$$

как, например, в [27].

Подобная техника улучшает реконструкцию в окклюзивных областях, однако всё ещё не решает проблему неоднозначности сопоставления на теоретическом уровне. Отметим, что в работе [26] было показано, что почти все точки 3D сцены, видимые на центральном синтезируемом кадре (например, \mathbf{x}_p^t), но невидимые на одном из соседних кадров (например, \mathbf{x}_p^{t+1}) тем не менее видимы на другом соседнем кадре \mathbf{x}_p^{t-1} . Таким образом можно предположить что один из одинарных потенциалов $\{u_p(l_p^{t \rightarrow t+1}), u_p(-l_p^{t \rightarrow t-1})\}$ в (16) отражает истинное сопоставление даже в окклюзивной области, и задача сделать правильный выбор. Вспомним, что одинарный потенциал обратно пропорционален логарифму вероятности верного соответствия $u_p(l) = -\log(P_p(l))$. Поэтому выбор меньшего значения потенциала из двух – вполне оправданная стратегия при вычислении комбинированного одинарного потенциала:

$$\hat{u}_p(l) = \min \{u_p(l_p^{t \rightarrow t+1}), u_p(-l_p^{t \rightarrow t-1})\}, \quad (17)$$

где $\hat{u}_p(l)$ обозначает комбинированный одинарный потенциал. Следовательно, функционал в (13) можно записать как:

$$E(l) = \sum_{p \in V} \hat{u}_p(l_p) + \sum_{(p,q) \in E} B_{p,q}(l_p, l_q) \quad (18)$$

Данный функционал можно минимизировать как в работе [10], где было показано, что при определённых ограничениях задача минимизации может быть решена с помощью билатеральной фильтрации. Действительно, для Марковской модели, применяемой в базовом методе, любая комбинация пикселей (узлов графа) для бинарного потенциала есть граница (p, q) данного графа \mathcal{G} принадлежащая множеству \mathcal{E} . В этом случае бинарный потенциал в (18) выражается как:

$$B_{p,q}(l_p, l_q) = \omega(p, q) \varphi(l_p, l_q), \quad (19)$$

где $\varphi(\cdot, \cdot)$ – бинарная цена гладкости, $\omega(p, q)$ – бинарный коэффициент, отражающий взаимное попарное пространственное влияние точек (p, q) в цвет – изображение пространстве. Тогда классическая форма билатерального фильтра выражается как:

$$\omega(p, q) = e^{-\frac{|\mathbf{x}_p - \mathbf{x}_q|^2}{2\sigma_x^2} - \frac{|\mathbf{f}_p - \mathbf{f}_q|^2}{2\sigma_f^2}}, \quad (20)$$

где \mathbf{x}_p – координаты точки на изображении p , \mathbf{f}_p – значение изображения в этой точке, как вектор в пространстве RGB, σ_x и σ_f – некоторые параметры ядра свёртки $\omega(p, q)$.

Задача минимизации функционала (18) может быть решена в рамках известной концепции алгоритма распространения доверия (АРД). Конечный результат – расчёт маргинала для каждой точки $\bar{u}_p(\cdot)$ и последующая оценка диспаратности в данной точке p по простой формуле “победитель забирает всё” (ПЗВ):

$$l(\mathbf{x}_p) = \arg \min_{l \in L} \bar{u}_p(l). \quad (21)$$

Основная операция АРД – рассылка сообщений [28]. В работе [10] доказано, что искомый маргинал при операции рассылки сообщений вычисляется как:

$$\bar{u}_p(l) = \sum_{q \in V} \omega(p, q) \hat{u}_q(l). \quad (22)$$

Таким образом исходная задача оптимизации функционала (18) переформулируется в задачу фильтрации (22). В данной статье мы используем быстрый билатеральный фильтр предложенный в [29]. Для данного шага алгоритма параметры σ_x и σ_f в (22) и (20) равны 10 и 5 соответственно. Значительные одноцветные области фона сцены (безоблачное небо, например) вносят дополнительную неопределённость в задачу сопоставления. Подобная проблема может быть преодолена усреднением одинарного потенциала \hat{u}_q внутри области. Такая процедура по сути должна предшествовать билатеральной фильтрации в (22). Сначала изображение \mathbf{f}_q должно быть сегментировано. Множество узлов \mathcal{V} четырёх-связанного графа $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ разделяется на $k \in K$ сегментов:

$$\bigcup_{k \in K} V_k = V; \quad \bigcap_{k \in K} V_k = \emptyset; \quad (p, q) \in \bar{E}_k \Rightarrow |\mathbf{f}_p - \mathbf{f}_q| \leq \Delta \quad (23)$$

Такая сегментация выполняется с помощью алгоритма анализа связанных компонент [30]. В нашем алгоритме порог Δ равен 7. Обычно все алгоритмы стереосопоставления включают дополнительную обработку полученной карты диспаратности, состоящую как минимум из двух шагов: проверка соответствия правой и левой карты диспаратности, медианная фильтрация результата. В нашем случае проверка соответствия избыточна, в силу использования трёх кадров и отсутствия формальных окклюзий в этом случае. Тем более, что подобная процедура увеличивает время работы алгоритма в два раза. Поэтому в предложенном алгоритме

используется только медианная фильтрация, при этом предложенная схема отличается от традиционной схемы. Для медианной фильтрации используется базовый функционал (18), однако вместо одинарного потенциала подставляются значения 0 или 1, в зависимости от полученной на предыдущем этапе карты диспаратности:

$$\hat{u}_p(l_p) = \begin{cases} 0 & \text{если } l_p = l'_p, \\ 1 & \text{иначе} \end{cases}, \quad (24)$$

где l'_p карта диспаратности полученная на предыдущем шаге алгоритма. Затем алгоритм повторяет билатеральную фильтрацию (22) и получаем новую карту диспаратностей (21). Параметры фильтра в этом случае σ_x и σ_f в (22) и (20) равны 10 и 30 соответственно. Полностью предложенный алгоритм состоит из следующих шагов:

- вычисление функции одинарного потенциала $\hat{u}_p(l)$ по формулам (14) и (17);
- пре-фильтрация $\hat{u}_p(l)$ по формуле (23);
- билатеральная фильтрация (22) с использованием алгоритма [29];
- вычисление карты диспаратности $l(\mathbf{x}_p)$ с использованием (21);
- коррекция карты диспаратности $l(\mathbf{x}_p)$ с помощью (24).

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Для оценки работы предложенного алгоритма мы использовали три серии изображений, полученных в результате стереодвижений с известными значениями карты диспаратности: первая TrainView на основе реальной видео съёмки нашей лаборатории и две синтетические последовательности Teddy и Cones, скомпонованные на основе базы данных университета Миддлбери. Все значения карты диспаратности для этих серий лежат в пределах 0-29. Для визуальной оценки результатов также использовались несколько других видео серий (в том числе записных с бортовой камерой робота) без известных значений карты диспаратности. В этой секции представлено три группы экспериментов. Первая группа экспериментов призвана показать преимущество предложенного метода в сравнении с аналогами на основе таблицы результатов. Вторая группа экспериментов показывает преимущество метода на основе визуальной оценки. Третья группа экспериментов посвящена оценке ухудшения работы предложенного алгоритма для не эпиполярной последовательности зависимости от угла поворота вокруг оси Y (наиболее часто встречающееся искажение при движении камеры бокового обзора).

Для оценки результатов мы используем стандартный критерий плохого значения в пикселе (BPP) с порогом 1:

$$BPP(Th) = \frac{100}{|V|} \sum_{p \in V} [|l(\mathbf{x}_p) - l^{GT}(\mathbf{x}_p)| > Th], \quad (25)$$

где $l^{GT}(\mathbf{x}_p)$ – известные значения карты диспаратности, Th – порог ошибки BPP, символ $[\bullet]$ означает иверсоновские скобки. Данные в Таблице 1 представляют результаты экспериментов, полученные для различных способов вычисления одинарного потенциала. В этой таблице предложенный алгоритм вычисления на основе вероятностного выбора обозначен как PCCF, накопительный алгоритм вычисления потенциала предложенный в [27] обозначен ADCF, STCF и STCF+CCN означают стандартные алгоритмы вычисления потенциала с использованием только двух изображений и суффикс +CCN означает, что используется пост-обработка карты диспаратности с проверкой соответствия правой и левой карты. Дополнительная колонка PPS показывает влияние медианной фильтрации. Из Таблицы 1 можно видеть, что предложенный алгоритм вычисления одинарного потенциала (PCCF) значительно улучшает результат реконструкции по сравнению с другими алгоритмами.

Вычислительная сложность алгоритма – $O(NL)$, где N – число пикселей в изображении и L – максимальная диспаратность. Время работы алгоритма в секундах с серией Cones указано в дополнительной колонке Таблицы 1. Используемый компьютер – an IntelCore i5-4300U 1.9-GHz и 6-GB RAM. Рис. 6 даёт возможность оценить работу различных алгоритмов визуально, то есть получить визуальную оценку. Интересно отметить на Рис. 6, что для серии TrainView предобработка с подавлением больших одноцветных областей работает только в случае расчёта одинарного потенциала с помощью предложенного алгоритма.

Как было показано в разделе 1, отход от эпиполярного приближения ведёт к ошибке вычисления диспаратности, линейно-зависящей от угла поворота, даже для идеального алгоритма. Примерно такая же зависимость ошибки от угла поворота наблюдается и для реальных алгоритмов. Чтобы проиллюстрировать этот тезис мы подготовили Рис. 7 и Рис. 8. Рис. 7 отражает функциональную зависимость ошибки от угла поворота ε . На Рис. 8 показаны четыре карты диспаратности, полученные с помощью предложенного алгоритма при угле поворота ε , равном 0, 0.5, 1, и 3 градусам соответственно. Понятно, что для поворотов более одного градуса использование не модифицированных алгоритмов стереосопоставления, разработанных для эпиполярного приближения, приведёт к неточной реконструкции.

Таблица 1. Оценка результатов с использованием стандартного критерия плохого значения в пикселе (BPP) с порогом 1 (BPP(1) в (25)) для различных алгоритмов вычисления одинарного потенциала. Правая колонка показывает время работы алгоритма для последовательности кадров Cones.

Алгоритм	TrainView		Teddy		Cones		Time Cones	
	Base	+PPS	Base	+PPS	Base	+PPS	Base	+PPS
PCCF	9.90	7.55	5.15	4.44	3.85	3.18	0.5	0.9
ADCF	43.5	42.8	8.33	6.31	7.85	6.67	0.5	0.9
STCF	45.8	45.0	9.88	8.42	7.91	6.11	0.5	0.9
STCF+CCH	44.8	44.1	7.86	6.29	6.39	5.56	0.8	1.45

ЗАКЛЮЧЕНИЕ

В статье предложены два критерия оценки карты диспаратности для случая латерального движения монокулярной камеры: теоретический и алгоритмический. Используя комбинацию данных методов оценки, возможно определить применимость классических методов стереосопоставления в контексте конкретно поставленной задачи. Предложен новый алгоритм стереосопоставления для латерального движения, улучшающий качество карты диспаратности на основе использования трех кадров. Постановка задачи построения карты диспаратности для движущейся монокулярной камеры дает возможность повысить точность работы, позволяя уменьшить ошибку вычисления диспаратности, устранить или по крайней мере уменьшить области окклюзии, а так же адаптивно выбирать базу для каждой тройки кадров. Так же проведен ряд экспериментов на основе предложенных методов для различных относительных углов поворота между системами координат камер.

СПИСОК ЛИТЕРАТУРЫ

1. Forsyth D., Ponce J. Computer Vision: A Modern Approach. NJ: Prentice-Hall : Englewood Cliffs, 2002.
2. Obinata G., Dutta A. Vision Systems: Segmentation and Pattern Recognition. Austria : Advanced Robotic Systems Publications, 2007.
3. Park M., Mozerov M., Kim D. et al. Object shape recovery in lateral navigation system using motion stereo technique // Proc. Multisensor Fusion and Integration for Intelligent Systems. 1999. P. 273–278.

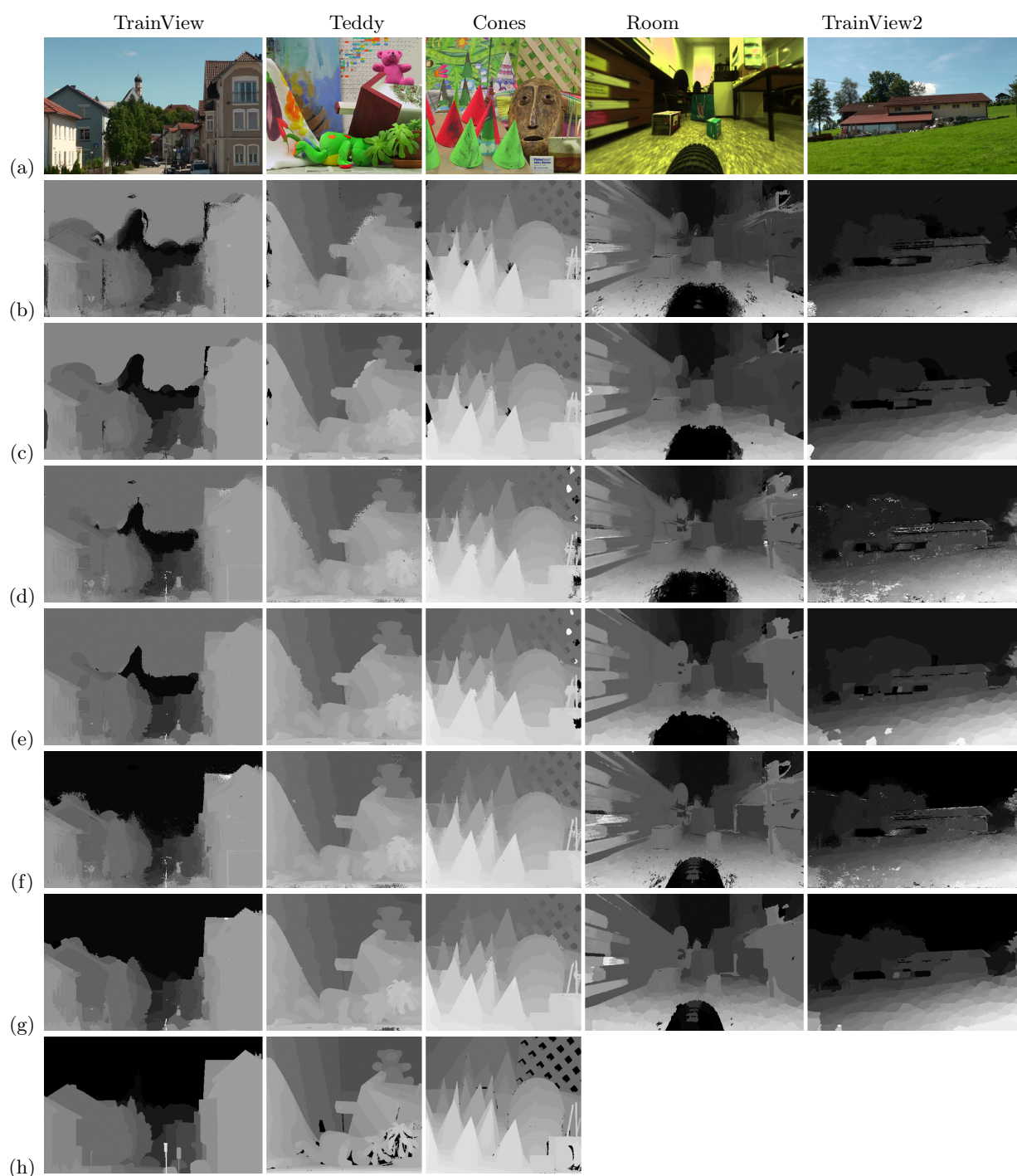


Рис. 6. (a) – Центральный кадр f^t последовательности; (b) – результат для STCF+CCH; (c) – результат для STCF+CCH +PPS; (d) – результат для ADCF; (e) – результат для ADCF+PPS; (f) – результат для PCCF; (g) – результат для PCCF+PPS; (h) – Известные значения карты диспаратности.

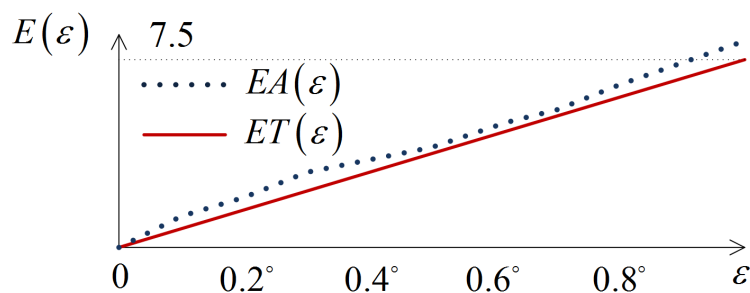


Рис. 7. Функциональная зависимость ошибки работы алгоритма от угла поворота ε : $EA(\cdot)$ и $ET(\cdot)$ – алгоритмическая и теоретическая оценки среднеквадратичной ошибки диспаратности (в пикселях) соответственно.

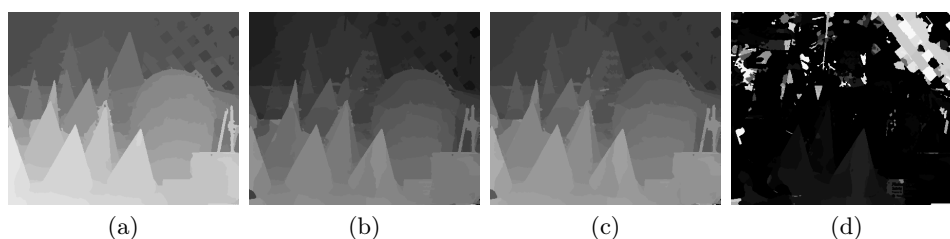


Рис. 8. Четыре карты диспаратности, полученные с помощью предложенного алгоритма при угле поворота ε : (a) – 0 градусов; (b) – 0.5 градуса; (c) – 1 градус; (d) – 3 градуса.

4. Mozerov M. G., Kober V. I., Ovseyevich I. A., Choi T. S. Motion stereo matching using a modified dynamic programming // Pattern Recognition And Image Analysis. 2000. Vol. 10, no. 1. P. 90–96.
5. Shvets E. Stochastic multi-agent patrolling using social potential fields // Proceedings 29th European Conference on Modelling and Simulation (ECMS 2015). 2015. P. 42–49.
6. I. Konovalenko A. Miller B. M. D. N. Uav navigation on the basis of the feature points detection on underlying surface // Proceedings 29th European Conference on Modelling and Simulation (ECMS 2015). 2015. P. 499–505.
7. A. Grigoryev D. Bocharov A. T. D. N. Vision-based vehicle wheel detector and axle counter // Proceedings 29th European Conference on Modelling and Simulation (ECMS 2015). 2015. P. 521–526.
8. Scharstein D., Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // International Journal of Computer Vision. 2002. Vol. 47, no. 1. P. 7–42.
9. Brown M., Burschka D., Hager G. Advances in computational stereo // IEEE Trans. Pattern Analysis and Machine Intelligence. 2003. Vol. 25, no. 8. P. 993–1008.
10. Mozerov M., van de Weijer J. Accurate stereo matching by two-step energy minimization // IEEE Transactions on Image Processing. 2015. Vol. 24, no. 3. P. 1153–1163.
11. Yoon K.-J., Kweon I.-S. Adaptive support-weight approach for correspondence search // IEEE Trans. Pattern Analysis and Machine Intelligence. 2006. Vol. 28, no. 4. P. 650–656.
12. Tombari F., Mattoccia S., L.di-Stefano, Addimanda E. Near real-time stereo based on effective cost aggregation // ICPR. 2008.
13. Yang Q. A non-local cost aggregation method for stereo matching // CVPR. 2012.
14. Min D., Lu J., Do M. A revisit to cost aggregation in stereo matching: how far can we reduce its computational redundancy? // ICCV. 2011.
15. Min D., Lu J., Do M. Joint histogram-based cost aggregation for stereo matching // IEEE Trans. Pattern Analysis and Machine Intelligence. 2013. Vol. 35, no. 10. P. 2539–2545.

16. Ihler A., Fisher J., Willsky A. Loopy belief propagation: Convergence and effects of message errors // *J. Machine Learning Research*. 2005. Vol. 6. P. 905–936.
17. Boykov Y., Veksler O., Zabih R. Fast approximate energy minimization via graph cuts // *IEEE Trans. Pattern Analysis and Machine Intelligence*. 2001. Vol. 23, no. 11. P. 1222–1239.
18. Kolmogorov V., Zabih R. Computing visual correspondence with occlusions using graph cuts // *ICCV*. 2001. P. 508–515.
19. Sun J., Zheng N.-N., Shum H.-Y. Stereo matching using belief propagation // *IEEE Trans. Pattern Analysis and Machine Intelligence*. 2003. Vol. 25, no. 7. P. 787–800.
20. Mozerov M., Gonzalez J., Roca X., Villanueva J. J. Trinocular stereo matching with composite disparity space image // *ICIP*. 2009. P. 2089–2092.
21. Zabih R., Woodfill J. Non-parametric local transforms for computing visual correspondence // *ECCV*. 1994. P. 151–158.
22. Lewis J. P. Fast template matching // *Vision Interface*. 1995. P. 120–123.
23. Mei X., Sun X., Zhou M. et al. On building an accurate stereo matching system on graphics hardware // *GPUCV*. 2011.
24. Xu L., Jiaand J., Matsushita Y. Motion detail preserving optical flow estimation // *CVPR*. 2010. P. 18–23.
25. Mozerov M. Constrained optical flow estimation as a matching problem // *IEEE Transactions on Image Processing*. 2013. Vol. 22, no. 5. P. 2044–2055.
26. Chen S. E., Williams L. View interpolation for image synthesis // *Proceedings of the 20th annual conference on Computer graphics and interactive techniques / ACM*. 1993. P. 279–288.
27. Mozerov M., Kober V., Choi T. S. Improved motion stereo matching based on a modified dynamic programming // *Opt. Eng.* 2001. Vol. 40(10). P. 2234–2239.
28. Szeliski R., Zabih R., Scharstein D. et al. A comparative study of energy minimization methods for markov random fields with smoothness-based priors // *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2008. Vol. 30, no. 6. P. 1068–1080.
29. Adams A., Baek J., Davis M. Fast high-dimensional filtering using the permutohedral lattice // *Computer Graphics Forum*. 2010. Vol. 29, no. 2. P. 753–762.
30. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. *Introduction to Algorithms, Second Edition*. The MIT Press and McGraw-Hill Book Company, 2001.

Stereo vision algorithms based on motion parallax of a monocular lateral camera

Ershov E., Karnaukhov V., Mozerov M.

In this paper we propose a way to estimate stereo matching algorithms applicability for moving monocular camera. We propose two ways to estimate degradation of a disparity map due to non-epipolarity: theoretical and empirical. The theoretical approach let us to compute epipolar distortion estimation as a function of rotation angles. The algorithmic method of disparity quality evaluation relies on comparison between computed disparities for different epipolar distortions. For the second method new stereo-matching algorithm for the case of moving monocular camera is developed.

KEYWORDS: stereo vision, stereo matching, lateral stereo, motion stereo.