

Решение проблемы окклюзий при оценке глубины сцены по заданному световому полю с использованием геодезического расстояния и принципа симметрии обзора

М.Г. Мозеров*, В.Н. Карнаухов*, В.И. Кобер*, Л.В. Зимина**

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

**Московский политехнический университет, Москва, 107023, Россия

Поступила в редколлегию 19.02.2024 г. Принята 10.05.2024 г.

Аннотация—Задача оценки глубины сцены по заданному световому полю можно свести к проблеме классического стереомэтчинга с утверждением, что совпадающие пиксели на стереоизображениях имеют одинаковые значения яркости. Однако это предположение в целом неверное, если учесть наличие шума на изображениях и различную освещенность левого и правого изображения в стереопаре, а также существование окклюзий. В этом плане световой поток — представляющий собой 80 изображений попарно эпиполярных центральному — дает ряд преимуществ, особенно в плане решения задачи окклюзий. В этой статье мы предлагаем принцип симметрии обзора: если пиксель центрального изображения лежит в зоне окклюзии относительно одного из периферийных изображений светового поля, то этот пиксель не принадлежит зоне окклюзии для осесимметричного изображения периферийного поля. Таким образом можно сформировать робастный относительно окклюзии объем весов несоответствия. Как результат — предложенный в статье алгоритм значительно улучшает результат реконструкции глубины сцены. Эффективность нашего подхода продемонстрирована с использованием основной базы данных светового поля и сравнением с лучшими алгоритмами реконструкции.

КЛЮЧЕВЫЕ СЛОВА: Световое поле, оценка глубины сцены, стереомэтчинг, фильтр на основе геодезического расстояния.

DOI: 10.53921/18195822_2024_24_1_93

1. ВВЕДЕНИЕ

Камеры светового поля собирают и записывают свет, поступающий с разных направлений. Таким образом световое поле фиксирует небольшие изменения вида сцены: обычно это 80 периферийных изображений попарно эпиполярных центральному, таким образом позволяя использовать такие изображения для определения глубины сцены. Определение глубины светового поля (LF) [1,2,3] является фундаментальной задачей компьютерного зрения со многими дальнейшими приложениями, такими как сверхразрешение [4], перефокусировка [5], 3D реконструкция [6] и виртуальная реальность [7]. Световое поле включает в себя обширную пространственную и угловую информацию о реконструируемой сцене, что позволяет оценить глубину этой сцены лучше, чем с помощью классического стереомэтчинга. Основываясь на характеристиках изображений светового поля, для оценки глубины изображения светового поля были предложены традиционные методы [2,3,8,9] и методы, основанные на обучении [10,11,12,13,14].

В последнее время методы, основанные на глубоком обучении [13, 14, 15, 16, 17], показывают хорошие результаты, особенно на тестовых базах данных. Многие методы, основанные на глубоком обучении, обычно включают четыре ключевых этапа: анализ и конструирование

признаков [18], построение объема затрат (или весов несоответствия) [19], агрегирование затрат и регрессия по глубине. Характерно, что точно такие же этапы являются ключевыми для классических подходов, основанных на математических моделях стереомэтчинга. Таким образом понятно, что значение новых методов, основанных на математических моделях, важно и для развития алгоритмов оценки глубины с помощью нейро-сетевых технологий. Более того, методы глубокого обучения могут давать худшие результаты в случае, если база данных для обучающей выборки не соответствует тестовой.

В этой статье мы представляем новый подход в вычислении значений весов несоответствия (cost), робастный к окклюзиям и основанный на математическом моделировании. Главная идея формулируется как принцип симметрии визуального обзора светового поля: если пиксель центрального изображения лежит в зоне окклюзии относительно одного из периферийных изображений светового поля, то этот пиксель не принадлежит зоне окклюзии для осесимметричного изображения периферийного поля. Такой подход имеет некоторое сходство с методом, предложенным в [9], однако цитируемый алгоритм двухшаговый: на первом этапе производится грубая оценка глубины, а затем уже формируется объем весов несоответствия с учетом обнаруженных зон окклюзий. Проблема подобных алгоритмов заключается в том, что грубая оценка как раз плохо выделяет зоны окклюзий, и восстановление реальной глубины в таких зонах не является достаточно точным. Предложенный алгоритм одношаговый и не требует предварительных оценок. Наш подход апеллирует к работе по стерео с тремя изображениями [20], а также к методу, используемому в навигации с камерой бокового обзора [21]. Предложенный метод разработан в рамках классического стереомэтчинга [22, 23], включающего построение объема затрат и регрессию глубины на основании глобальной минимизации энергии этого объема и постобработки полученной карты глубин. Кроме того, мы расширили рамки подхода, включив в него геодезическое расстояние и фильтрацию, основанную на этой метрике [24]. При таком подходе минимизация энергии включает в себя подобную свертку с ядром на основе геодезического расстояния, подобно тому, как это было реализовано в работе [25].

Статья организована следующим образом: в разделе 2 предложен новый принцип формирования объема затрат для светового поля, в разделе 3 предложен алгоритм стереомэтчинга, основанный на свертке с ядром геодезического расстояния, в разделе 4 приведены экспериментальные результаты и наконец, раздел заключение суммирует наши выводы.

2. ПРИНЦИП СИММЕТРИИ ОБЗОРА ПРИ ФОРМИРОВАНИИ ОБЪЕМА ЗАТРАТ

Формирование веса несоответствия (затрат) является важной частью любого алгоритма стереомэтчинга. Существует два вида традиционных затрат: мера несхожести, соответствующая пикселю, и непараметрические преобразования с некоторой опорной областью [26, 27]. Последние достижения в области глубокого обучения обеспечивают более надежную меру несхожести сопоставления для стерео, однако это стратегия не подходит для светового поля с 81 изображением. В этой работе мы будем использовать простейшую меру несоответствия — попиксельную. Для этого существуют две предпосылки: а) формирование функции затрат на основе окрестности пикселя противоречит предложенному принципу симметрии обзора — часть пикселей окрестности может не попадать в область окклюзий и наоборот; б) предлагаемый в статье метод минимизации энергии на основе ядра свертки геодезического расстояния по сути включает в себя адаптацию к окрестности рассматриваемого пикселя. Концепция эпиполярного стереомэтчинга предполагает применительно к световому полю, что если физическая точка на поверхности объекта отображена на центральном $I^{0,0}$ и одном из периферийных изображений $I^{n,m}$, как два пикселя с разными координатами $\mathbf{p}^{0,0}$ и $\mathbf{p}^{n,m}$, то соотношение координат

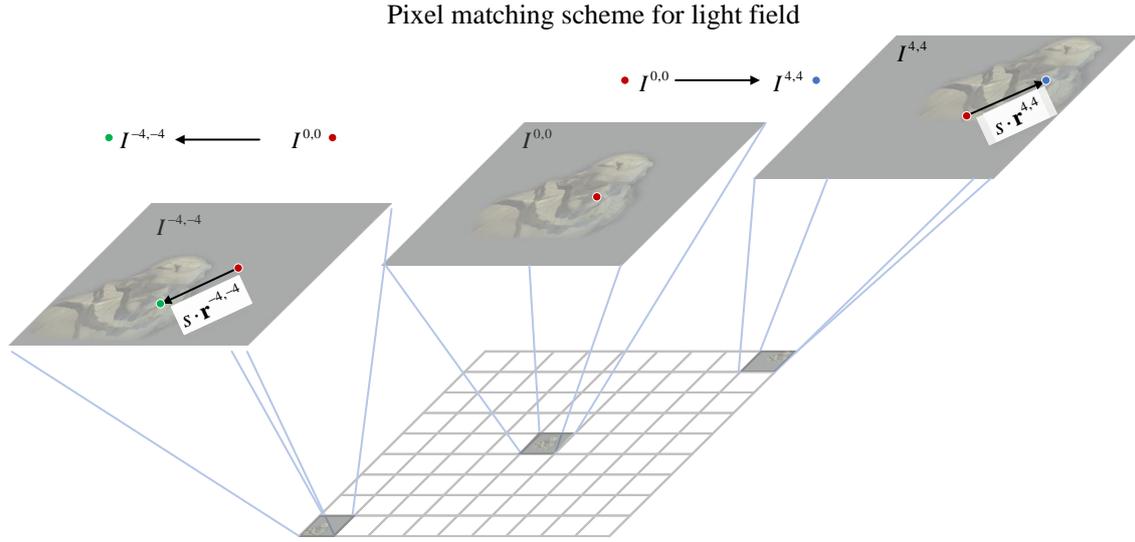


Рис. 1. Схема попиксельного мэтчинга для изображений светового поля.

этих пикселей представимо как:

$$\mathbf{p}^{n,m} = \mathbf{p}^{0,0} + s \cdot \mathbf{r}^{n,m}, \quad (1)$$

где $\mathbf{r}^{n,m}$ — направляющий вектор соответствующих периферийных изображений $I^{n,m}$, s — скалярная величина, называемая диспаратностью (обратно пропорциональная глубине рассматриваемой физической точки сцены). Схема стереомэтчинга для светового поля иллюстрируется на Рис. 1.

Важно отметить, что для светового поля $n, m \in \{-N, -M\}, \dots, \{-n, -m\}, \dots, \{n, m\}, \{N, M\}$, а для используемых в этой статье изображений светового поля $N = M = 4$. Одним из параметров изображений светового поля является максимальный вектор диспаратности $\delta_{\max}^{n,m} = \{|\mu_{\max}|n, |\mu_{\max}|m\}$. Направляющий вектор для размерности измеряемой в пикселях в свою очередь определен как: $\mathbf{r}^{n,m} = \{n, m\}$. Откуда следует, что переменная диспаратности принадлежит интервалу целых чисел $s \in \{-\mu_{\max}, \dots, \mu_{\max}\}$. Однако если необходимо повысить точность измерения диспаратности до субпиксельной размерности, то интервал этой переменной можно сделать произвольным: $s \in \{-S_{\max}, \dots, S_{\max}\}$, хотя в этом случае требуется нормировка направляющего вектора:

$$\mathbf{r}^{n,m} = \left\{ \left\lfloor \frac{\mu_{\max}}{S_{\max}} \right\rfloor n, \left\lfloor \frac{\mu_{\max}}{S_{\max}} \right\rfloor m \right\}.$$

Так как целью стереомэтчинга является определение глубины всех точек в апертуре центрального изображения, то необходимо вычислить функцию $s_{\mathbf{p}}$. В первом приближении $s_{\mathbf{p}} = \arg \min_{s \in \mathcal{S}} \tilde{c}_{\mathbf{p}}(s)$, где $\tilde{c}_{\mathbf{p}}(s)$ — классическая функция затрат, определенная для случая светового потока как:

$$\tilde{c}_{\mathbf{p}}(s) = \sum_{n,m \in \{M,N\} \setminus \{0,0\}} c_{\mathbf{p}}^{n,m}(s), \quad (2)$$

где

$$c_{\mathbf{p}}^{n,m}(s) = |I_{\mathbf{p}}^{0,0} - I_{\mathbf{p}+s \cdot \mathbf{r}^{n,m}}^{n,m}|^2. \quad (3)$$

В (3) представлена функция затрат в классическом понимании мэтчинга. Однако такая функция затрат не отражает реального совпадения пикселей на изображении, соответствующих одной физической точке на поверхности объекта в зонах окклюзии. Поэтому в этой статье мы предлагаем робастную к окклюзии функцию затрат:

$$c_{\mathbf{p}}(s) = \sum_{n,m \in \{M,N\} \setminus \{0,0\}} c_{\mathbf{p}}^{|n|,|m|}(s), \quad (4)$$

где интегральная робастная функция затрат $c_{\mathbf{p}}(s)$ является суммой симметричных функций затрат:

$$c_{\mathbf{p}}^{|n|,|m|}(s) = \min \left(\left| I_{\mathbf{p}}^{0,0} - I_{\mathbf{p}+s \cdot \mathbf{r}^{n,m}}^{n,m} \right|^2, \left| I_{\mathbf{p}}^{0,0} - I_{\mathbf{p}+s \cdot \mathbf{r}^{-n,-m}}^{-n,-m} \right|^2 \right). \quad (5)$$

Принцип вычисления робастной функции затрат проиллюстрирован на Рис. 2. Чтобы понять, почему такая функция затрат реально робастна к окклюзии, заметим, что если пиксель $\mathbf{p}^{0,0}$ не имеет соответствия с пикселем $\mathbf{p}^{n,m}$ из-за окклюзии, но соответствует пикселю $\mathbf{p}^{-n,-m}$, то в идеальном случае $c_{\mathbf{p}}^{n,m}(s) > c_{\mathbf{p}}^{-n,-m}(s)$ если диспаратность s совпадает с истинным значением.

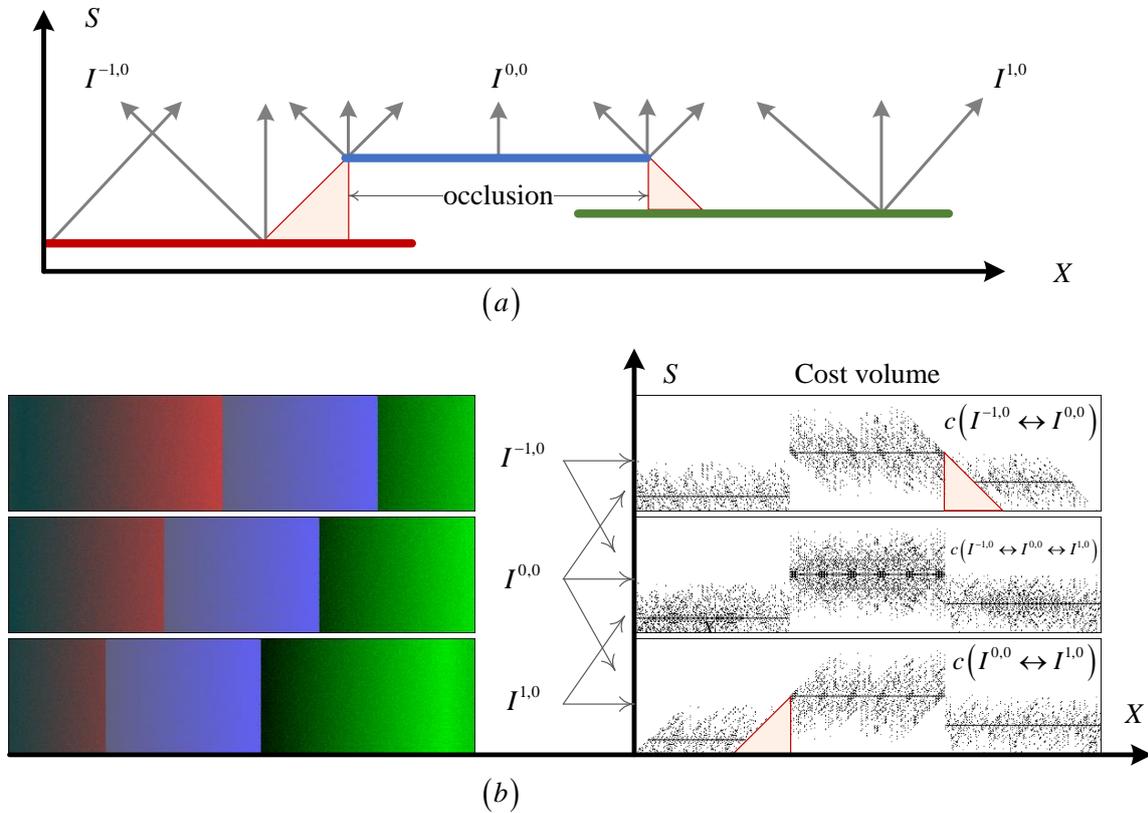


Рис. 2. Принцип формирования робастной функции затрат.

3. МИНИМИЗАЦИЯ ЭНЕРГИИ С ПОМОЩЬЮ ПОЛНОСТЬЮ СВЯЗАННОЙ МОДЕЛИ УСЛОВНОГО СЛУЧАЙНОГО ПОЛЯ

В этом разделе мы рассмотрим минимизацию энергии с помощью полностью связанной модели условного случайного поля (CRF). Такая модель является модификацией локально

связанной моделью, предложенной в [28] и дополненной до полностью связанной. Первый раз такая модель была предложена в [23], а затем обобщена в одном итеративном процессе минимизации энергии полностью связанной модели с привязкой к геодезическому расстоянию в [25].

Принципиальное различие между полностью связанными моделями и локально связанными моделями заключается в количестве пикселей, которые включены в рассматриваемую окрестность пикселя. В полностью связанных моделях все пиксели изображения потенциально связаны друг с другом. Однако сила каждого соединения зависит от выбранного показателя сходства между пикселями. Эта сила связности формализуется весами взаимного сходства изображений. Следовательно, эти веса становятся частью бинарных потенциалов в терминах гладкости энергетического функционала. Сначала давайте рассмотрим энергетическую функцию следующего вида:

$$E(s) = \sum_{p \in \mathcal{V}} c_p(s_p) + \sum_{(p,q) \in \mathcal{E}} B_{p,q}(s_p, s_q), \quad (6)$$

где (и далее в этой статье) $p, q, k, l \in \mathcal{V}$ соответствуют пикселям или вершинам и задает $\varepsilon = (p, q) \in \mathcal{E}$ ребрам графа изображения $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, s_p обозначает метку пикселя p , который принадлежит некоторому дискретному набору различий $s \in S$; $c_p(\cdot)$ определяет унарный потенциал, который соответствует обычной стоимости штрафа или минус логарифм вероятности правдоподобия. $B_{p,q}(\cdot, \cdot)$ — двоичный потенциал, который определяет взаимодействие краев между пикселями (p, q) . В случае полностью связанной модели CRF каждая комбинация пикселей (p, q) графа изображения \mathcal{G} принадлежит набору ребер \mathcal{E} . В этом случае бинарный потенциал для модели должен быть выражен как:

$$B_{p,q}(s_p, s_q) = b_q w_{p,q} \varphi(s_p, s_q), \quad (7)$$

где $\varphi(\cdot, \cdot)$ — попарная функция затрат, которая обеспечивает плавность оцениваемой функции меток, а $w_{p,q}$ — попарный вес, который определяет некоторое взаимное влияние между каждой вершиной графа. Переменная b_q является коэффициентом нормализации, чтобы сделать уравнение (6) выполнимым при суммировании по всем границам изображения \mathcal{E} в пикселях q , таким образом $b_q = \sum_{p \in \mathcal{V}}^{-1} w_{p,q}$. Проблема минимизации энергии в (6) может быть решена в рамках подхода “belief propagation” (BP). Цель состоит в том, чтобы найти предельную функцию $\bar{c}_q(\cdot)$, а затем оценить метки в каждом пикселе q по простой формуле “winner takes it all” (WTA), которая определяет максимальную вероятность:

$$s_q = \arg \min_{s \in S} \bar{c}_q(s). \quad (8)$$

Основной операцией BP является передача сообщения $m_{p \rightarrow q}$ от вершины p к вершине q для направленного ребра $(p \rightarrow q) \in \mathcal{E}$. Тогда предельная функция выражается следующим образом:

$$\bar{c}_q^t(s) = c_q(s) + \sum_{p,q \in \mathcal{E}} m_{p \rightarrow q}^{t-1}(s). \quad (9)$$

Здесь t — номер итерации, и предполагается, что соответствующая итеративная процедура передачи сообщения задана [29]:

$$m_{p \rightarrow q}^t(i) = \min_{j \in S} (\beta_{p,q} \bar{c}_p^t(j) - m_{p \rightarrow q}^{t-1}(j) + B_{p,q}(i, j)), \quad (10)$$

где индексы $(i, j) \in S$ принадлежат области дискретного несоответствия S , а $\beta_{p,q}$ является коэффициентом нормализации, который может быть установлен равным $\beta_{p,q} = b_q w_{p,q}$. Термином $m_{p \rightarrow q}^{t-1}(j)$ можно пренебречь для полностью подключенной модели, поскольку в целом это значение равно нулю относительно суммы всех других сообщений, поступающих в пиксель q . Таким образом, мы можем переписать (10) следующим образом:

$$\begin{aligned} m_{p \rightarrow q}^t(i) &= b_q w_{p,q} m_p^t(i), \\ m_p^t(i) &= \min_{j \in S} (\bar{c}_p^t(j) + \varphi_{p,q}(i, j)) \end{aligned} \quad (11)$$

Формула (9) также может быть переписана:

$$\bar{c}_q^t(s) = c_q(s) + \frac{1}{W_q - 1} \sum_{p \in \mathcal{V}} w_{p,q} m_p^{t-1}(s) - m_q^{t-1}(s), \quad (12)$$

где W_q — нормализующий множитель:

$$W_q = \sum_{p \in \mathcal{V}} w_{p,q}.$$

Можно видеть, что суммарный член $\sum_{p \in \mathcal{V}} w_{p,q} m_p^{t-1}$ в уравнении (12) совпадает с определением фильтрации для дискретного изображения с помощью большого класса ядер фильтров $w_{p,q}$, которые включают традиционные ядра, основанные на гауссовом, билатеральном или геодезическом расстоянии.

В этой статье мы предлагаем использовать быстрый фильтр на основе геодезических расстояний, описанный в следующем разделе, поскольку его рекурсивное вычисление на основе дерева графа позволяет последовательно обновлять сообщения с использованием (11). Таким образом, решение задачи минимизации энергопотребления с помощью полностью подключенной модели CRF может быть получено с помощью итерационного процесса, использующего уравнения (12), (11) и (8).

3.1. Рекурсивное дерево вычислений на основе геодезических расстояний

Фильтр, основанный на геодезическом расстоянии, обычно выбирается в следующем виде, что делает фильтр рекурсивным:

$$\begin{aligned} \tilde{f}_q &= \frac{1}{W_q} \sum_{p \in \mathcal{V}} e^{-ad_{p,q}} f_p \\ W_q &= \sum_{p \in \mathcal{V}} e^{-ad_{p,q}}, \end{aligned} \quad (13)$$

где вес $e^{-ad_{p,q}}$ определяет близость на основе геодезического расстояния между любыми двумя пикселями изображения. Переменная $d_{p,q}$ в уравнении (13) — геодезическое расстояние между пикселями изображения (p, q) , которое для изображения I_p может быть определено на дискретной сетке в виде:

$$\begin{aligned} d_{q,p} &= \min_{P_{p,q}} \sum_{\varepsilon \in P_{p,q}} u_\varepsilon, \\ u_{\varepsilon=(p,p)} &= \|I_q - I_p\| + \delta, \end{aligned} \quad (14)$$

где $P_{q,p}$ — путь между двумя вершинами графа (p, q) и δ — пространственный член.

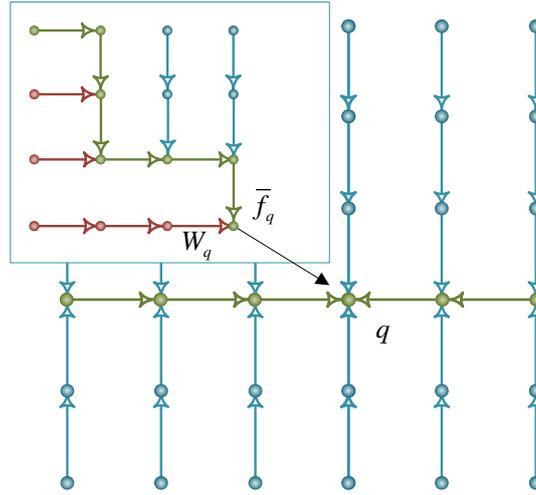


Рис. 3. Одно из четырех рекурсивных деревьев вычислений [24].

Обратите внимание, что внутренние параметры фильтра a и δ в уравнении (14) приблизительно соответствуют параметрам классического двустороннего фильтра с гауссовым ядром следующим образом:

$$a = \frac{2}{\sigma_r^2}, \delta = \frac{\sigma_r^2}{\sigma_s^2}, \quad (15)$$

где σ_r^2 и σ_s^2 — диапазон и пространственная дисперсия соответственно. Также из уравнения (14) можно вывести следующее полезное соотношение:

$$w_{p,q} = \min_{P_{p,q}} \sum_{p \in \mathcal{V}} e^{-ad_{p,q}} = \max_{P_{q,p}} \prod_{\varepsilon \in P_{p,q}} e^{-a u_\varepsilon}. \quad (16)$$

В [24] показано, что сумма фильтра в уравнении (12) может быть вычислена быстро, и эта схема вычисления допускает последовательное обновление сообщений.

Полное дерево, которое соответствует этому быстрому алгоритму [24], состоит из четырех квадрантов-доменов (или ветвей дерева). Одна из четырех ветвей проиллюстрирована на Рис. 3, и W_q представляет сумму весов $w_{p,q}$ для первого квадранта, когда \tilde{f}_q выдает результаты фильтрации также для этого квадранта. Обратите внимание, что оптимальное дерево и все веса в уравнении (16) могут быть вычислены для всех других процедур.

3.2. Последовательное обновление сообщения

Разница между последовательным и итеративным обновлением сообщений ВР определена в [28]. В соответствии с этим определением можно переписать маргинальное уравнение (12) как:

$$\bar{c}_q^t(s) = c_q(s) + \frac{1}{W_q - 1} \sum_{p \in \mathcal{V}} w_{p,q} m_p^t(s) - m_q^t(s) \quad (17)$$

Принимая во внимание уравнение (9), можно сказать, что сообщение, поступающее от пиксела p , зависит от всех входящих сообщений одной и той же итерации. Однако, поскольку последовательный процесс предполагает, что вычисление сообщения происходит рекурсивно по ветви

дерева (квадрант-домен) от листьев (все предыдущие пиксели ветви) до корня p , то имеет значение только одно обновление ветви, и полная итерация состоит из всех четырех проходов ветви.

Чтобы лучше понять идею последовательного обновления сообщений ВР, давайте рассмотрим одномерный случай (например, строку на изображении) конвергенции ВР. Известно, что энергия сходится после конечного числа итераций и что это число пропорционально или равно числу узлов, из которых состоит граф этих одномерных данных. С другой стороны, подход динамического программирования (DP) позволяет достичь решения сразу, что эквивалентно одной итерации. Обратите внимание, что рекурсивная функция перехода от вершины к вершине DP в этом случае в точности эквивалентна сообщению ВР, поступающему из тех же вершин. Таким образом, последовательный подход является реализацией принципа DP в методе ВР.

3.3. Параметры полностью связанной модели

Чтобы решить функционал в уравнении (6) и, таким образом, решить основную часть нашего алгоритма сопоставления стерео, необходимо выполнить процесс последовательного обновления сообщений, определенный в уравнении (17). Чтобы сделать это, нам нужно определить внутренние метрические веса $w_{p,q}$ нашей полностью связанной модели. Эти веса $w_{p,q}$ вычисляются с помощью уравнений (14-15) и в наших экспериментах параметры фильтра равны: $\sigma_r = 30$ и $\sigma_s = 8$.

3.4. Постобработка карты диспаратностей

Для получения результатов, сравнимых с лучшими современными алгоритмами по реконструкции глубины с помощью светового поля, требуется несколько этапов постобработки. В нашей работе мы полностью повторяем такие шаги, основываясь на технике предложенной в статье [25].

4. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Экспериментальная часть была задумана таким образом, чтобы показать основные достоинства предлагаемого алгоритма. Экспериментальная секция нашей работы поделена на две части где:

- мы анализируем численные результаты и преимущества использования предложенного метода формирования функции затрат, а также минимизации энергии полностью связанной модели с привязкой к геодезическому расстоянию. Анализ численных результатов производится на основе открытой базы данных изображений светового поля [30].
- часть результатов экспериментов представлена в иллюстрациях статьи, чтобы читатель мог сравнить качество реконструкции поля диспаратностей визуально.

4.1. Эксперименты с численным анализом результатов

В этой части экспериментальной секции результат, полученный с помощью предложенного алгоритма ORC-FCM (occlusion robust cost and fully connected model), сравнивается с двумя методами Omg-occl [9] и OccUnsup [31], в которых рассматривается проблема окклюзии. Второй из этих методов, опубликованный в конце 2023 года, по сути, является имплементацией первого для нейро-сетевых технологий. Кроме этого, мы сравниваем наш метод с тремя

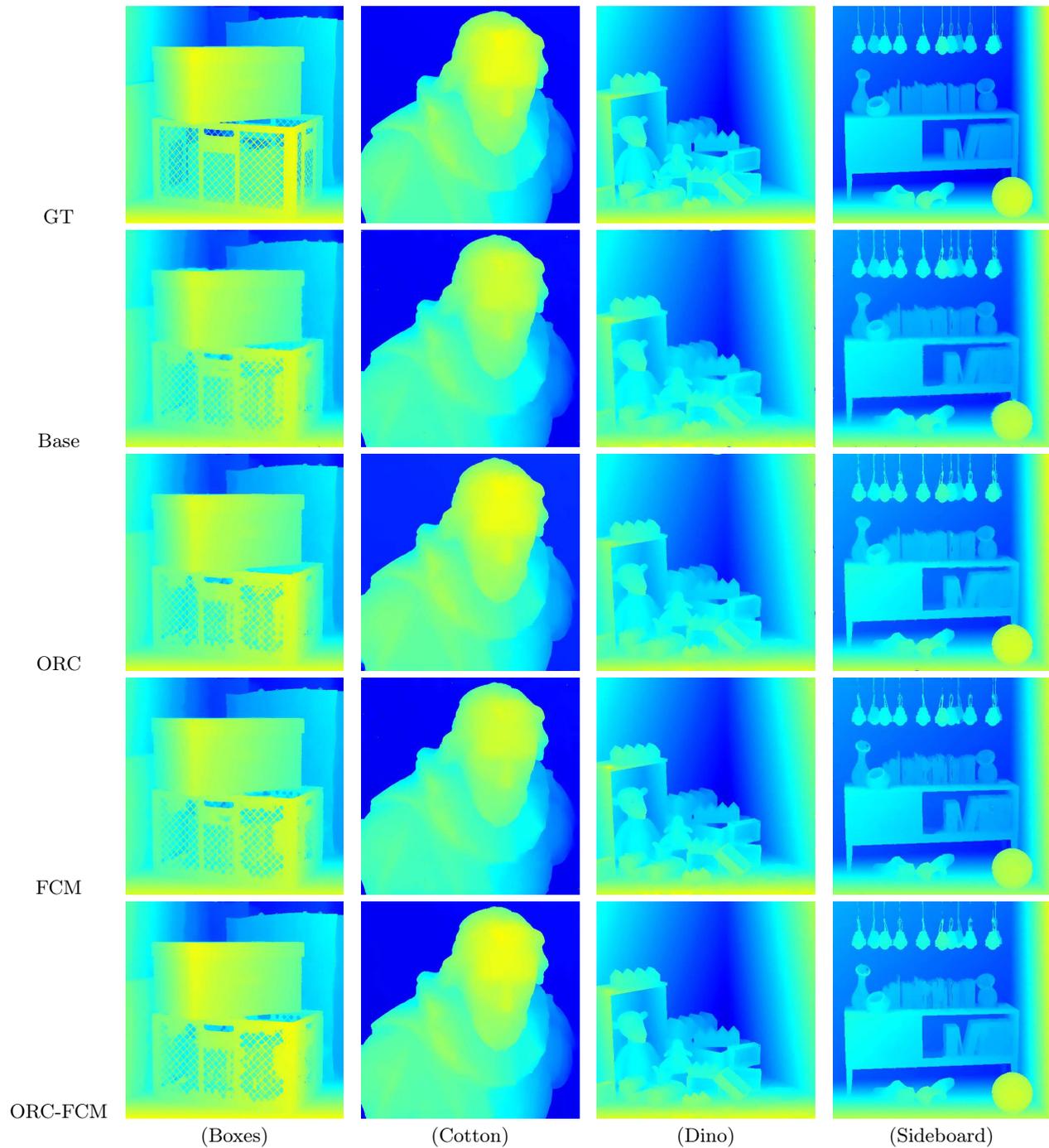


Рис. 4. Результат реконструкции карты диспаратности для визуального сравнения. Мы сравниваем предложенный алгоритм ORC-FCM с тремя модификациями предложенного метода: Base, ORC и FCM.

Таблица 1. Численные результаты сравнения реконструкции карты диспаратности различных методов с истинными значениями по критерию MSE (верхняя часть таблицы) и критерию превышения порога неточности 0.07 пикселей в процентах (нижняя часть таблицы).

Images:	Boxes	Cotton	Dino	Sideboard
Process step:				
Omg-Occl [9]	7.494	3.341	0.885	1.318
OccUnsup [31]	6.232	0.436	0.403	1.156
Base	6,764	2,097	0,652	1,465
ORC	5,471	0,651	0,316	1,219
FCM	6,393	0,653	0,456	1,197
ORC-FCM	5,279	0,428	0,272	1,166
Omg-Occl [9]	24.66	6.212	6.648	11.92
OccUnsup [31]	16.82	1.524	3.424	9.391
Base	15,44	2,541	5,223	9,935
ORC	13,31	1,71	3,74	8,91
FCM	13,22	1,461	4,142	8,601
ORC-FCM	11,51	1,430	3,712	8,891

модификациями предложенного метода: а) Base — реализован на основе классического подхода стереомэтчинга, когда функция затрат вычисляется по формуле (2)-(3), а объем затрат обрабатывается с помощью фильтра, чувствительного к границам (edge preserved filter) для каждой плоскости этого объема; б) ORC — вариант предложенного метода, где используется предложенная функция затрат, робастная к окклюзии, но не используется предложенная минимизация энергии; в) FCM — вариант, когда функция затрат вычисляется по формуле (2)-(3), как и в случае Base, но применяется метод минимизации энергии полностью связанной модели. В Таблице 1 представлены результаты этого сравнения.

Анализируя результат в Таблице 1, мы можем сказать что предложенный в статье алгоритм значительно превосходит не только классические методы реконструкции карты диспаратности, но и нейро-сетевую имплементацию известного алгоритма по созданию робастной к окклюзии функции затрат [31].

4.2. Результаты экспериментов для визуального анализа

В этой части экспериментальной секции результат реконструкции карты диспаратности, полученный с помощью предложенного метода ORC-FCM, визуально сравнивается с тремя модификациями предложенного метода: Base, ORC и FCM. Сравнимые карты диспаратностей представлены в Рис. 4.

ЗАКЛЮЧЕНИЕ

В этой статье мы представили новый подход в вычислении функции затрат, робастный к окклюзиям, основанный на математическом моделировании как принцип симметрии визуального обзора светового поля: если пиксель центрального изображения лежит в зоне окклюзии относительно одного из периферийных изображений светового поля, то этот пиксель не принадлежит зоне окклюзии для осесимметричного изображения периферийного поля. Ключевой алгоритм основан на глобальной минимизации энергии объема предложенной функции затрат.

Кроме того, мы расширили рамки подхода, включив туда геодезическое расстояние и фильтрацию, основанную на этой метрике.

СПИСОК ЛИТЕРАТУРЫ

1. Wilburn B., Joshi N., Vaish V. et al. High performance imaging using large camera arrays // ACM SIGGRAPH 2005 Papers. 2005. P. 765–776.
2. Jeon H.-G., Park J., Choe G. et al. Accurate depth map estimation from a lenslet light field camera // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. P. 1547–1555.
3. Williem W., Park I. K. Robust light field depth estimation for noisy scene with occlusion // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 4396–4404.
4. Zhang S., Lin Y., Sheng H. Residual networks for light field image super-resolution // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. P. 11046–11055.
5. Wang Y., Yang J., Guo Y. et al. Selective light field refocusing for camera arrays using bokeh rendering and superresolution // IEEE Signal Processing Letters. 2018. Vol. 26, no. 1. P. 204–208.
6. Kim C., Zimmer H., Pritch Y. et al. Scene reconstruction from high spatio-angular resolution light fields. // ACM Trans. Graph. 2013. Vol. 32, no. 4. P. 73–1.
7. Yu J. A light-field journey to virtual reality // IEEE MultiMedia. 2017. Vol. 24, no. 2. P. 104–112.
8. Tao M. W., Hadap S., Malik J., Ramamoorthi R. Depth from combining defocus and correspondence using light-field cameras // Proceedings of the IEEE International Conference on Computer Vision. 2013. P. 673–680.
9. Zhu H., Wang Q., Yu J. Occlusion-model guided antiocclusion depth estimation in light field // IEEE Journal of Selected Topics in Signal Processing. 2017. Vol. 11, no. 7. P. 965–978.
10. Heber S., Pock T. Convolutional networks for shape from light field // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 3746–3754.
11. Tsai Y.-J., Liu Y.-L., Ouhyoung M., Chuang Y.-Y. Attention-based view selection networks for light-field disparity estimation // Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. 2020. P. 12095–12103.
12. Chen J., Zhang S., Lin Y. Attention-based multi-level fusion network for light field depth estimation // Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. 2021. P. 1009–1017.
13. Wang Y., Wang L., Liang Z. et al. Occlusion-aware cost constructor for light field depth estimation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 19809–19818.
14. Sheng H., Liu Y., Yu J. et al. Lfnat 2023 challenge on light field depth estimation: Methods and results // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. P. 3472–3484.
15. Zhou S., Hu L., Wang Y. et al. Aif-1fnet: All-in-focus light field super-resolution method considering the depth-varying defocus // IEEE Transactions on Circuits and Systems for Video Technology. 2023.
16. Liu J., Song N., Xia Z. et al. A dense light field reconstruction algorithm for four-dimensional optical flow constraint equation // Pattern Recognition. 2023. Vol. 134. P. 109101.
17. Chao W., Wang X., Wang Y. et al. Learning sub-pixel disparity distribution for light field depth estimation // IEEE Transactions on Computational Imaging. 2023. Vol. 9. P. 1126–1138.
18. Liang Z., Wang Y., Wang L. et al. Learning non-local spatial-angular correlation for light field image super-resolution // arXiv preprint arXiv:2302.08058. 2023.
19. Wood D. N., Azuma D. I., Aldinger K. et al. Surface light fields for 3d photography // Seminal Graphics Papers: Pushing the Boundaries, Volume 2. 2023. P. 487–496.

20. Mozerov M., Gonzalez J., Roca X., Villanueva J. J. Trinocular stereo matching with composite disparity space image // ICIP. 2009. P. 2089–2092.
21. Ershov E., Karnaukhov V., Mozerov M. Probabilistic choice between symmetric disparities in motion stereo matching for a lateral navigation system // Optical Engineering. 2016. Vol. 55, no. 2. P. 023101–023101.
22. Scharstein D., Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // International journal of computer vision. 2002. Vol. 47. P. 7–42.
23. Mozerov M. G., 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.
24. Mozerov M., van de Weijer J. Improved recursive geodesic distance computation for edge preserving filter // IEEE Transactions on Image Processing. 2017. Vol. 26, no. 8. P. 3696–3706.
25. Mozerov M. G., Van De Weijer J. One-view occlusion detection for stereo matching with a fully connected crf model // IEEE Transactions on Image Processing. 2019. Vol. 28, no. 6. P. 2936–2947.
26. Zabih R., Woodfill J. Non-parametric local transforms for computing visual correspondence // ECCV. 1994. P. 151–158.
27. Lewis J. P. Fast template matching // Vision Interface. 1995. P. 120–123.
28. Kolmogorov V. Convergent tree-reweighted message passing for energy minimization // International Workshop on Artificial Intelligence and Statistics / PMLR. 2005. P. 182–189.
29. Szeliski R., Zabih R., Scharstein D. et al. A comparative study of energy minimization methods for markov random fields // ECCV. 2006. P. 16–29.
30. Honauer K., Johannsen O., Kondermann D., Goldluecke B. A dataset and evaluation methodology for depth estimation on 4d light fields // Asian Conference on Computer Vision / Springer. 2016.
31. Zhou W., Lin L., Hong Y. et al. Beyond photometric consistency: Geometry-based occlusion-aware unsupervised light field disparity estimation // IEEE Transactions on Neural Networks and Learning Systems. 2023.

Occlusion handling in depth estimation of a scene from a given light field using geodesic distance and the principle of symmetry of the view

M.G. Mozerov, V.N. Karnaukhov, V.I. Kober, L.V. Zimina

The problem of estimating the depth of a scene from a given light field can be reduced to the problem of classical stereo matching with the statement that matching pixels in stereo images have the same brightness values. However, this assumption is generally incorrect if we take into account the presence of noise in the images and the different illumination of the left and right images in the stereo pair, as well as the existence of occlusions. In this regard, the luminous flux — representing 80 images in pairs epipolar to the central one — provides a number of advantages, especially in terms of solving the problem of occlusion. In this article, we propose the principle of viewing symmetry: if a pixel of the central image lies in the occlusion zone relative to one of the peripheral images of the light field, then this pixel does not belong to the occlusion zone for an axisymmetric image of the peripheral field. In this way, it is possible to form a robust volume of the discrepancy weights relative to occlusion. As a result, the algorithm proposed in the article significantly improves the result of reconstruction of the depth of the scene. The effectiveness of our approach is demonstrated using the main test database of the light field and comparing it with the best reconstruction algorithms in the efficiency of border recognition and in the speed of calculation.

KEYWORDS: Light field, scene depth estimation, stereomatching, filter based on geodetic distance.