

УДК 004.932

МАСШТАБИРОВАНИЕ ЭТАЛОННОЙ КАРТЫ ГЛУБИНЫ ДЛЯ ЗАДАЧ СТЕРЕОЗРЕНИЯ

Е. Р. Муратов, к.т.н., доцент кафедры ЭВМ РГПТУ, Рязань, Россия;
orcid.org/0000-0002-1664-3954, e-mail: myratov_er@mail.ru

Предлагается алгоритм уменьшения разрешения эталонной карты глубины. Представляется методика оценки, а также сравнение значений ошибок, возникающих в результате уменьшения размера эталонной карты глубины и изображений различными алгоритмами масштабирования. Целью работы является оценка точности масштабирования карты глубины предлагаемым алгоритмом. Необходимость масштабирования возникает в задачах приведения существующих разных тестовых наборов данных к единым параметрам стереосистемы, одним из которых является разрешение сенсоров. Обобщенный набор данных с одинаковыми параметрами требуется для проектирования более точных, гибридных алгоритмов стереозрения с элементами искусственного интеллекта.

Ключевые слова: масштабирование эталонной карты глубины, уменьшение размера массива значений, плотная карта глубины, точность карт глубины, показатель оценки $BAD(X)$, масштабирование изображений, сравнение алгоритмов масштабирования.

DOI: 10.21667/1995-4565-2019-70-127-135

Введение

Использование стереосистем для оценки расстояний до объектов является актуальным для систем технического зрения [1-4]. В настоящее время ведется активное внедрение алгоритмов стереозрения на программно-аппаратном уровне в смартфоны с целью повышения качества снимков с камеры в портретном режиме. Предъявляемые требования к точности результата работы стереоалгоритма высоки. Современные классические алгоритмы им не соответствуют. А следовательно, требуется разрабатывать гибридные алгоритмы, использующие как классические подходы вычисления карт глубин, так и вычисления, выполняемые с использованием алгоритмов искусственного интеллекта (Artificial Intelligence – AI). Стоит отметить, что доля алгоритмов AI в задачах обработки изображений стремительно растет. Качество вычислений существенно зависит от того, как и чем обучили нейронную сеть. Для обучения нейронных сетей, например методом обучения «с учителем», требуются размеченные данные, из которых формируют множество данных для тренировки и тестирования (training и testing dataset). Для обучения нейронной сети для вычисления высокоточной карты глубины требуется набор данных, содержащий огромное количество разнообразных сюжетов. Количество данных должно исчисляться сотнями тысяч изображений от двух камер вместе с эталонной картой глубины (Ground Truth, далее сокращенно GT). Получение набора данных с высокоточным GT является одной из актуальных и достаточно трудоемких задач. В интернете можно найти различные наборы данных для тестирования стереоалгоритмов [5], но их не очень много. Анализ состава этих данных показал, что они почти всегда не будут соответствовать требуемым параметрам стереосистемы, так как получены с использованием различных сенсоров. Одним из актуальных параметров является разрешение изображений и GT, следующими по актуальности являются угол обзора сенсора, а третьим является расстояние между сенсорами [6]. Трансформация данных из различных наборов в единый набор с требуемыми параметрами стереосистемы позволит решать задачи разработки и тестирования гибридных стереоалгоритмов на более качественном уровне. В работе [5] приводится сравнение некоторых популярных наборов данных между собой (таблица 1).

Таблица 1 – Сравнение некоторых datasets

Table 1 – Comparison of some datasets

Название Dataset	Количество сцен	Количество значений в GT $\times 10^6$	Плотность GT %	Точность GT	Разрешение изображений (MPix)
Middlebury	33	~179	100	0,2 пикс	5,5
Make3D	257	~9*	0,4	55 мм	0,8
Ladicky	1175	–	–	–	0,1
EISATS	464	~11*	7	20 мм	0,3
KITTI	389	~90*	50	20 мм	0,5
CATS	343	~51*	50	2 мм	1

Анализируя значения в таблице 1, можно прийти к выводу, что разрешение изображений и GT, состав сцен и точность GT удовлетворяют текущие запросы большинства задач обработки изображений. Однако каждого отдельного dataset не достаточно, например, для обучения AI алгоритмов. Существует метод обучения нейронной сети, использующий только изображения от сенсоров стереосистемы (без эталонного GT) – алгоритм «обучения без учителя» [7]. Однако, как показывает практика, этого не достаточно для обучения AI высокоточного стереоалгоритма. Следовательно потребуются все методы для обучения AI, а также набор данных, сформированный из множества существующих и откорректированных наборов, а также вновь созданных. Так как создать новый высокоточный набор – задача трудоемкая и длительная по времени, то более простым решением является трансформация существующих наборов под требуемые параметры стереосистемы.

Методика оценки алгоритмов масштабирования

Для масштабирования изображений существует множество алгоритмов [9, 10, 11]. Масштабирование, как правило, выполняется сверткой ядра масштабирования с оригинальным изображением с последующей обработкой, например алгоритмом антиалиасинга. Некоторые алгоритмы при этом стараются повысить контрастность и четкость мелких деталей [12]. Однако для масштабирования эталонных стереоданных такая обработка носит только негативный характер, так как снижает правдоподобность оригинальных данных, в особенности GT. Требование к масштабированию GT – максимальная правдоподобность результата, другими словами, минимальная абсолютная разность между эталоном и результатом масштабирования. Кроме того, значения масштабированного GT нуждаются в коррекции. Это видно из формулы (1) [8] вычисления дальности по стереокадрам соосных и откалиброванных стереосенсоров.

$$h = \frac{W \cdot b}{d \cdot 2 \operatorname{tg}(\alpha / 2)}, \quad (1)$$

где h – расстояние до объекта в метрах, W – ширина изображения в пикселях, b – расстояние между камерами в метрах, d – диспаратет в пикселях, α – угол зрения камеры по ширине изображения. Формула (1) показывает зависимость диспаратета от всех параметров стереосистемы. Например, изменение масштаба по ширине вызывает линейное изменение значений каждого диспаратета, т.е. $d = Wf$, где $f = b/(2\operatorname{tg}(\alpha/2))$. Эта зависимость позволяет синтезировать наборы данных с новыми требуемыми параметрами (разрешение изображения, угол зрения и расстояние между сенсорами). При этом коррекция одного из параметров стереосистемы потребует пересчета значений GT. Также потребуется выполнить масштабирование изображений и GT при коррекции разрешения. При коррекции расстояния между сенсорами потребуются синтез изображений (не линейная коррекция): правого по левому и левого по правому, учитывая новые значения GT. Следует отметить, что вследствие коррекции значения расстояния между сенсорами – вновь синтезированные изображения становятся менее реалистичными, так как они могут получить артефакты, связанные с наличием зон окклюзий [6] в оригинальном изображении. Оценка алгоритма выполняется сравнением результата с эталоном. Отсутствие эталонов в различных разрешениях становится существенным ограничи-

вающим фактором для точной оценки работы алгоритмов. Для экспериментов и последующей оценки работы алгоритмов масштабирования возьмем в качестве эталонов данные с различающимися GT из «Middlebury dataset 2014» [10].

Каждый набор данных содержит изображения от левого и правого сенсоров в нескольких разрешениях, полученные в разных режимах экспозиции и освещения. А также эталонные карты глубин GT для левого и правого сенсоров и файл описание параметров системы. Наборы данных содержат GT, различающиеся как по разрешению, так и по диапазону диспаритетов. Можно выделить множество наборов с различающимися GT – это наборы с названиями Adirondack, ArtL, Jadeplant, Motorcycle, Piano, Pipes, Playroom, Playtable, PlaytableP, Recycle, Shelves, Teddy, Vintage [10]. Каждый набор представлен в трех различных масштабах: F (полный размер), H (половина), Q (четверть), где $F=2H=4Q$.

Для оценки ошибки в результате масштабирования будем использовать показатель $BAD(X)$, вычисляемый по формуле (2):

$$BAD(X) = \frac{\sum_{j=0}^{h-1} \sum_{i=0}^{w-1} \begin{cases} 1, |D_{j,i} - GT_{j,i}| > X \\ 0, |D_{j,i} - GT_{j,i}| \leq X \end{cases}}{h \cdot w}, \quad (2)$$

где X – дистанция отклонения в пикселях, D – вычисленный массив диспаритетов (масштабированная карта глубины), GT – эталонный массив диспаритетов, h и w – соответственно высота и ширина этих массивов. А также среднее значение модуля ошибки и отклонение, вычисляемые по формуле (3):

$$Mx = \frac{\sum_{j=0}^{h-1} \sum_{i=0}^{w-1} |D_{j,i} - GT_{j,i}|}{h \cdot w}, \quad (3)$$

$$\sigma^2 = \frac{\sum_{j=0}^{h-1} \sum_{i=0}^{w-1} (D_{j,i} - GT_{j,i})^2}{h \cdot w} - \left(\frac{Mx}{h \cdot w} \right)^2.$$

Оценку масштабирования изображений будем измерять показателями $BAD(X)$, где $X \in \{0, 5; 1; 2; 4; 8\}$ (например, показателем с $X=0,5$ можно оценить минимальное отличие в один уровень яркости), а для GT будем использовать большее количество показателей, так как требуется высокая точность, $X \in \{0, 1; 0, 2; 0, 5; 1; 2; 4; 8\}$.

Следует отметить, что уменьшение разрешения вносит меньшую ошибку, чем увеличение, поэтому приводить к общему разрешению следует по возможности уменьшением масштаба.

Для определения тенденций среди алгоритмов и выбора оптимального по качеству достаточно выполнить оценку результата масштабирования (пропорционального уменьшения размера в два раза) из разрешения H в Q изображений и GT левого сенсора с эталоном.

Для получения полного представления о точности алгоритмов необходимо выполнять масштабирование на различные масштабные коэффициенты. Отсутствие эталонов не позволяют это сделать напрямую. Используем масштабирование в промежуточное разрешение $Q_I < H$, из которого затем выполним масштабирование в разрешение Q . Для ряда задач требуется увеличивать разрешение. Процедура увеличения масштаба вносит большую погрешность в результат, чем процедура уменьшения. Для комплексной оценки точности будем масштабировать изображение до разрешения Q_I , а затем обратно до размера H .

Выбор оптимального существующего алгоритма

Результаты оценки сокращения размера изображений и GT в два раза были усреднены по всем наборам данных. Оценки уменьшения размера изображений приведены в таблицах 2 и 3.

Таблица 2 – Оценки уменьшения разрешения изображений

Table 2 – Image Resolution Reduction Estimates

Метод	BAD(X) %					Mx	σ
	X=0,5	X=1	X=2	X=4	X=8		
Nearest	36,04	21,37	15,55	9,94	5,09	1,65	4,47
Bilinear	33,21	14,35	9,23	4,51	1,24	0,8	1,8
Bicubic	22,22	5,17	1,87	0,331	0,012	0,31	0,69
Box	48,41	14,03	6,95	2,25	0,347	0,8	1,26
Lanczos2	21,27	4,68	1,59	0,25	0,007	0,29	0,65
Lanczos3	11,74	0,015	<0,01	<0,01	0	0,12	0,32

Таблица 3 – Оценки уменьшения разрешения GT

Table 3 – GT Resolution Decrease Estimates

Метод	BAD(X) %							Mx	σ
	X=0,1	X=0,2	X=0,5	X=1	X=2	X=4	X=8		
Nearest	5,27	3,73	1,37	1,26	1,16	1,06	0,93	0,34	3,35
Bilinear	10,22	8,22	7,03	6,01	4,73	3,19	1,75	0,45	2,30
Bicubic	14,28	11,25	8,39	6,18	4,05	2,61	1,56	0,43	2,20
Box	5,29	3,49	3,16	2,93	2,64	2,22	1,47	0,33	2,15
Lanczos2	14,02	10,98	8,32	6,24	4,12	2,63	1,56	0,43	2,20
Lanczos3	17,47	13,8	10,03	7,24	4,77	2,85	1,59	0,47	2,22

Из таблицы 2 видно, что для масштабирования изображения можно выбрать приемлемый алгоритм, например «Lanczos3». Масштабирование GT этими алгоритмами дает существенные ошибки (таблица 3). Величина ошибки не позволяет точно обучить нейросеть или оценить конечную работу современного алгоритма стереозрения. Однако среди алгоритмов следует выделить лучший по показателю ошибки BAD(X), это «Nearest», а также лучший по значениям Mx и σ – «Box».

В настоящее время к современным стереоалгоритмам предъявляются требования по точности показателя BAD(0,5) порядка 15-25 % и BAD(1) порядка 12-15 %. Например, по показателю BAD(2) передовые алгоритмы позволяют получить точность 11-12 % [13, 14]. Несмотря на достаточно низкую ошибку, допускаемую современными алгоритмами стереозрения (в 11-12 %), этих значений недостаточно, например когда необходимо оценить карту глубины для тонких, периодических объектов, особенно расположенных на близкой дистанции от камер. Достичь более низких показателей могут только гибридные алгоритмы с элементами AI. Опыт показывает, что достаточно сложно снизить значение показателей даже на 1 %. На это влияют разные факторы, например сложная сцена, величина зон окклюзий и т.п. Поэтому к масштабированному GT как основному объекту, участвующему в обучении AI, должны предъявляться максимальные требования по достоверности и точности. Иначе ошибки масштабирования будут оказывать влияния на ошибки работы обученных AI алгоритмов. Следовательно, ошибка масштабирования, измеренная показателем BAD(1), должна быть сопоставима примерно с 1 %, а при $X > 1$ значение показателя должно быть существенно меньше 1 %, а лучше стремиться к 0 % при увеличении значения X.

Алгоритм уменьшения размера GT

Рассмотрим специфику масштабирования GT.

Во-первых, при масштабировании основная ошибка приходится на значения представляющие границы объектов на разных дальностях. При этом следует учитывать, что граница разделяет два объекта переднего и заднего плана, а следовательно, сглаживание этих частей GT при масштабировании недопустимо.

Во-вторых, при формировании изображения на фотоматрице яркость пикселя изображения складывается из энергий светового потока от объектов, которые проецируются линзой на этот пиксель. Как правило, световой поток от ближнего объекта имеет большую энергию. В

большинстве случаев участки объектов имеют гладкие поверхности, а значит, и гладкое распределение значений диспаритетов соседних элементов в карте глубины, за исключением краев объектов.

Учитывая сказанное выше, предлагается следующий алгоритм масштабирования GT эффективного для уменьшения размера (downsampling).

Шаг 1 – в исходном массиве размером $h \times w$ линейно масштабируется система координат до разрешения $H \times W$.

Поясним на примере: пусть j, i индексы значения в оригинальном массиве M . Каждое значение $M_{j,i}$ занимает квадрат с площадью 1. Пусть j', i' индексы некоторого значения массива M' в масштабированной системе координат. Если наложить эти координаты друг на друга так, чтобы совпали точки $(0,0)$ с $(0',0')$ и $(h-1, w-1)$ с $(h'-1, w'-1)$, то прямоугольник с диагональю $A'B'$ накроет прямоугольную область в оригинальной системе координат (рисунок 1). На рисунке 1 w обозначает весовой коэффициент.

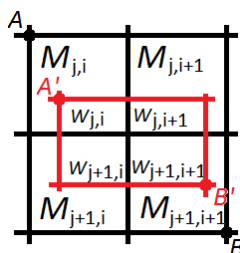


Рисунок 1 – Графическая интерпретация процесса масштабирования массива
Figure 1 – Graphical interpretation of array scaling process

Эта область в зависимости от случая масштабирования будет состоять из частей площадей, занимаемых соседними значениями в массиве M . Сумма этих значений, каждое из которых умножено на весовой коэффициент (эквивалент перекрытой площади прямоугольником $A'B'$), будет являться значением M' с индексом j', i' .

Далее алгоритм описывает действия, совершаемые для каждого масштабируемого элемента массива M' .

Шаг 2 – Вычисление значений массива M и весовых коэффициентов, входящих в сумму элемента массива M' . Каждое вычисленное значение составляет элемент кластера S , элемент является записью из двух значений: $S[i].v$ – числовое значение из массива M и $S[i].w$ – весовой коэффициент (на рисунке 1 обозначен как w).

Шаг 3 – Составление нового кластера C . Изначально кластер C пуст, а структура элемента аналогична элементу кластера S . Каждый элемент S сравнивается с элементами кластера C . Если среди элементов в составе C найдется элемент с индексом n , для которого выполняется условие $|S[m].v - C[n].v| < \varepsilon$, где m – индекс значения из GT в образованной сумме, ε – некоторая погрешность, то значения кластера обновляется по формулам (4):

$$C[n].v := \frac{S[m].v \cdot S[m].w + C[n].v \cdot C[n].w}{S[m].w + C[n].w}, \quad (4)$$

$$C[n].w := S[m].w + C[n].w,$$

иначе создается новый элемент кластера с значениями $C[z] = S[m]$, где z – увеличенный на 1 текущий размер кластера C . Порог ε определяется исходя из точности оригинального GT и должен быть меньше этого значения.

Шаг 4 – определяется элемент $C[x]$ с максимальным весом.

Шаг 5 – значение для массива M' определяется как $C[x].v * k$, где k – коэффициент масштаба изображения $k = WH/wh$.

Оценка предложенного алгоритма

Результаты усредненной оценки по всем наборам данных для предложенного алгоритма уменьшения размера GT (назовем «My») из разрешения H в Q , а также результаты уменьшения размера изображения представлены в таблице 4.

Таблица 4 – Оценки уменьшения размера GT

Table 4 – GT and Image Reduction Estimates

Показатель	Уменьшение размера GT			Уменьшение размера изображения	
	<i>My</i> ($\varepsilon=0,15$)	<i>Nearest</i>	<i>Box</i>	<i>Lanczos3</i>	<i>My</i> ($\varepsilon=128$)
<i>BAD</i> (0,1)	3,62	5,27	5,29	---	---
<i>BAD</i> (0,2)	2,6	3,73	3,49	---	---
<i>BAD</i> (0,5)	0,85	1,37	3,16	11,74	48,82
<i>BAD</i> (1)	0,78	1,26	2,93	0,015	19,11
<i>BAD</i> (2)	0,72	1,16	2,64	0,0033	10,37
<i>BAD</i> (4)	0,65	1,06	2,22	0,00016	3,48
<i>BAD</i> (8)	0,58	0,93	1,47	0	0,56
<i>Mx</i>	0,20	0,34	0,33	0,1176	0,94
σ	2,57	3,35	2,15	0,32	1,51

Уменьшение размера в k раз, а затем уменьшение до размера Q увеличивает ошибку. Вместе с этим можно отметить неэффективность существующих алгоритмов при масштабировании до не кратного 1/2 разрешения. На рисунке 2 представлены зависимости показателей *BAD*(X) от масштабного коэффициента k промежуточного разрешения (масштабирование в два шага).

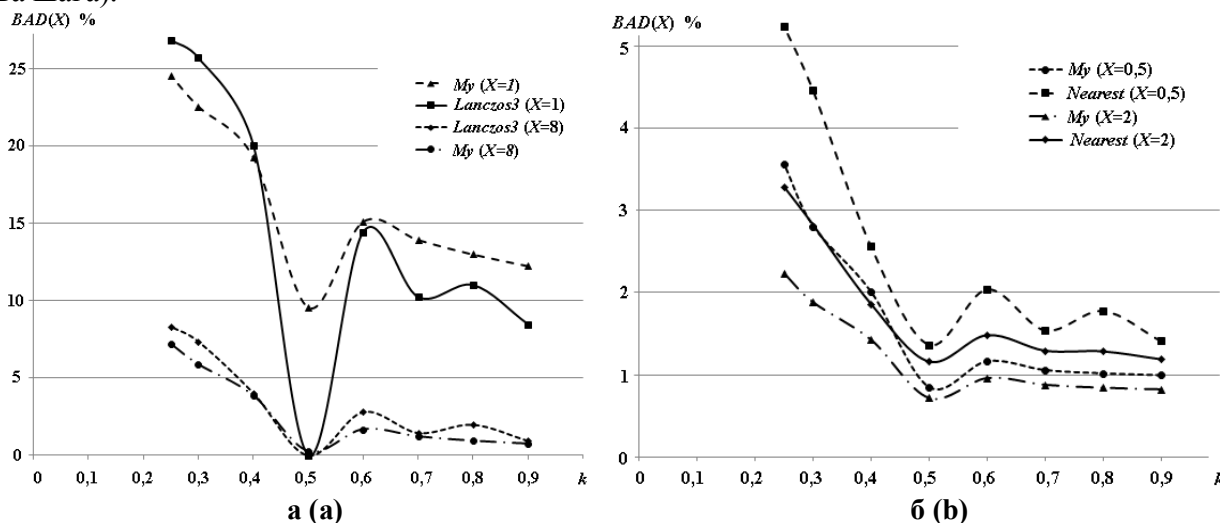


Рисунок 2 – Зависимость *BAD*(X) от k : а – для масштабирования изображений ($\varepsilon=128$), б – для масштабирования GT ($\varepsilon=0,15$)

Figure 2 – Dependence of *BAD* (X) on k : a – for scaling images ($\varepsilon = 128$), b – for scaling GT ($\varepsilon = 0,15$)

Зависимость на рисунке 2, б отражает эффективность предлагаемого алгоритма для задачи уменьшения разрешения GT. Однако для масштабирования изображений (рисунок 2, а) эффективность предложенного алгоритма проявляется только при сильном уменьшении разрешения ($k < 0,4$).

Оценки комплексного масштабирования, с применением как уменьшения, так и увеличения разрешения, представлены на рисунках 3 и 4. Масштабирование выполнено в два шага. Сначала разрешение уменьшено с применением масштабного коэффициента k , затем увеличено до первоначального размера. Для упрощения понимания тенденции зависимостей на рисунке 2 изображены зависимости только для метрик *BAD* 0,5, *BAD* 1 и *BAD* 8.

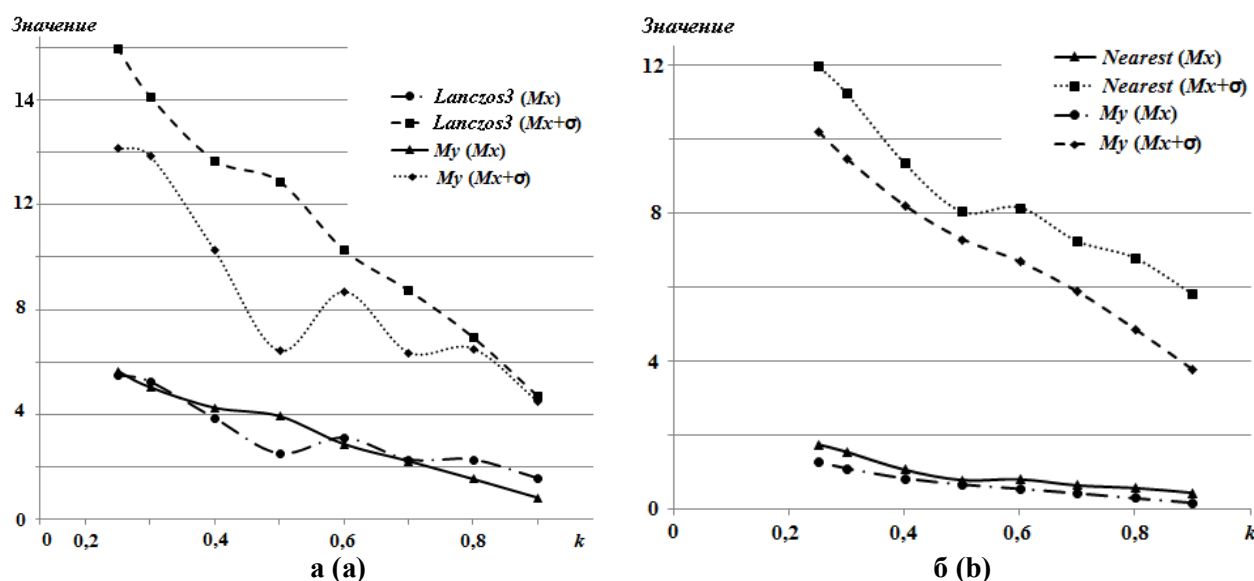


Рисунок 4 – Зависимость усредненной ошибки и отклонения от коэффициента масштабирования: а – для масштабирования изображений, б – для масштабирования GT

Figure 4 –Dependence of average error and deviations from scaling factor:

a – for image scaling, b – for GT scaling

Следует заметить, что другая последовательность действий, сначала увеличение размера, а затем уменьшение, не оценивалась, так как не имеет смысла. При такой последовательности действий все показатели $BAD(X)$ будут нулевыми.

Заключение

Как видно из таблицы 4, ошибка при уменьшении разрешения GT в два раза, приблизительно в 1,55 раза меньше по всем показателям по сравнению с известными алгоритмами масштабирования. Однако ориентированность алгоритма на сохранение отношений значений на границах объектов не дает выигрыша при уменьшении размера изображений. Минимальная ошибка при уменьшении размера изображений достигается только для больших значений ε ($\varepsilon = 128$ для 8 битных значениях яркостных каналов). Следовательно, для достижения минимальной погрешности при сжатии эталонных изображений с GT необходимо применять предложенный алгоритм только для масштабирования GT и алгоритм «Lanczos3» для масштабирования изображений. Ошибка масштабирования зависит от состава сцены (точнее, от количества значений, обозначающих границы объектов). При уменьшении разрешения изображения менее чем на половину предлагаемым алгоритмом можно получить преимущество перед алгоритмом «lanczos3» по точности метрик $BAD(X)$ и по значению усредненной ошибки. При уменьшении разрешения GT преимущество перед рассмотренными алгоритмами масштабирования наблюдается для всех коэффициентов масштаба в метриках $BAD(X)$, а также для усредненной ошибки и отклонения.

Уменьшение значения ошибки при масштабировании GT обеспечивает часть алгоритма, реализующая кластеризацию с последующим выбором элемента с наибольшим весом. Исключения этого этапа из алгоритма делает предложенный алгоритм подобным «nearest» по метрикам $BAD(X)$, но при этом алгоритм не будет в точности повторять «nearest» алгоритм.

Применение предложенного алгоритма позволило синтезировать один унифицированный dataset, содержащий ~2500 эталонных стереоизображений и GT с единым разрешением.

Библиографический список

1. Jiho Chang, Jae-chan Jeong, Dae-hwan Hwang. Real-time Hybrid Stereo Vision System for HD resolution disparity map. Proceedings of the British Machine Vision Conference, BMVA Press, 2014, pp. 1-11. DOI: <http://dx.doi.org/10.5244/C.28.89>.
2. Szelinski R. Computer Vision: Algorithms and Applications. Springer, 2011, 812 p.

3. **Muratov Y.R., Nikiforov M.B., Ustukov D. I., Gurov V.S.** Implementing one of Stereovision Algorithms on FPGA. 5th Mediterranean Conference on Embedded Computing (MECO) Proceedings. Montenegro, Budva, 2016, pp.64-67.
4. **Muratov Y.R., Nikiforov M.B., Ustukov D. I., Melnik O.V.** Analysis of The Efficiency of Dense Stereovision Algorithm Implementation on Different Computer Architectures. The 2nd International Conference on Computer and Communication Systems ICCCS 2017, Krakow , Poland, 2017, pp.78-81.
5. **Wayne Treible, Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, Michael O'Neal, Brian Phelan, Kelly Sherbondy, Chandra Kambhamettu.** CATS: A Color and Thermal Stereo Benchmark. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2961-2969. DOI: 10.1109/CVPR.2017.22.
6. **Muratov Y.R., Nikiforov M.B., Rusakiov A.B., Gurov V.S.** Estimation of distance to objects by stereovision. 4rd Mediterranean conference on embedded computing (MECO), Budva, 2015, pp. 155-158.
7. **Goodfellow et al.** Deep Learning. MIT Press, 2016. 778 p.
8. **Forsyth D.A., Ponce J.** Computer Vision a modern approach. Upper Saddle River, NJ 07458, 2003, 926 p.
9. **Gonzalez R.C. and Woods R.E.** Digital Image Processing, 2nd edition, Prentice Hall, 2002, 954 p.
10. **2014 Stereo datasets with ground truth.** [Электронный ресурс] middlebury.edu. Режим доступа: <http://vision.middlebury.edu/stereo/data/scenes2014/> , свободный.
11. **Roy R., Pal M. and Gulati T.** Zooming Digital Images using Interpolation Techniques. International Journal of Application or Innovation in Engineering & Management (IJAIEEM), Vol 2, Issue 4, 2013, pp. 34-45.
12. **Ватолин Д., Путилин С.** Оценка качества методов масштабирования изображений и результаты сравнений разных методов. [Электронный ресурс] GraphiCon 2003 Proceedings. Режим доступа: <http://graphicon.ru/oldgr/ru/publications/text/gc2003vatput.pdf>, свободный.
13. **D. Scharstein and R. Szeliski.** Middlebury stereo evaluation. [Электронный ресурс] Режим доступа: <http://vision.middlebury.edu/stereo/eval/>, свободный.
14. **D. Scharstein and R. Szeliski.** A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithm. International Journal of Computer Vision 47(1-3), 2002, pp. 7-42.

UDC 004.932

SCALING OF GROUND TRUTH FOR STEREO TASKS

E. R. Muratov, Ph.D. (Tech.), assistant professor, RSREU, Ryazan, Russia;
orcid.org/0000-0002-1664-3954, e-mail: myratov_er@mail.ru

*The article proposes an algorithm for reducing the resolution of depth maps and ground truth, estimating and comparing error metrics resulting from a reduction size of ground truth and images for various scaling algorithms. **The aim of the work** is to estimate the accuracy of scaling ground truth by the proposed algorithm. The need for scaling stereo dataset arises in the task of obtaining data sets with uniform parameters of a stereo system, one of which is sensor resolution, from existing different of stereo datasets. A generalized data set with the same parameters of stereo system is required for designing more accurate hybrid stereo vision algorithms with the elements of artificial intelligence.*

Key words: depth maps scaling, downsampling ground truth, dense depth map, depth map accuracy, BAD error estimation metric, image scaling, scaling algorithms comparison.

DOI: 10.21667/1995-4565-2019-70-127-135

References

1. **Jiho Chang, Jae-chan Jeong, Dae-hwan Hwang.** Real-time Hybrid Stereo Vision System for HD resolution disparity map. *Proceedings of the British Machine Vision Conference, BMVA Press.* 2014, pp. 1-11. DOI: <http://dx.doi.org/10.5244/C.28.89>.
2. **Szelinski R.** Computer Vision: Algorithms and Applications. *Springer.* 2011, 812 p.

3. **Muratov Y. R., Nikiforov M. B., Ustukov D. I., Gurov V. S.** Implementing one of Stereovision Algorithms on FPGA. *5th Mediterranean Conference on Embedded Computing (MECO) Proceedings*. Montenegro, Budva. 2016, pp.64-67.
4. **Muratov Y. R., Nikiforov M. B., Ustukov D. I., Melnik O. V.** Analysis of The Efficiency of Dense Stereovision Algorithm Implementation on Different Computer Architectures. *The 2nd International Conference on Computer and Communication Systems ICCCS 2017*. Krakow, Poland. 2017, pp.78-81.
5. **Wayne Treible, Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, Michael O'Neal, Brian Phelan, Kelly Sherbondy, Chandra Kambhamettu.** CATS: A Color and Thermal Stereo Benchmark. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA. 2017, pp. 2961-2969. DOI: 10.1109/CVPR.2017.22.
6. **Muratov Y. R., Nikiforov M. B., Rusakiov A. B., Gurov V. S.** Estimation of distance to objects by stereovision. *4rd Mediterranean conference on embedded computing (MECO)*. Budva. 2015, pp. 155-158.
7. **Goodfellow et al.** *Deep Learning*. MIT Press. 2016, 778 p.
8. **Forsyth D.A., Ponce J.** Computer Vision a modern approach. *Upper Saddle River, NJ 07458*. 2003, 926 p.
9. **Gonzalez R. C. and Woods R. E.** *Digital Image Processing, 2nd edition, Prentice Hall*. 2002, 954 p.
10. **2014 Stereo datasets with ground truth.** [Elektronnyj resurs] middlebury.edu. Rezhim dostupa: <http://vision.middlebury.edu/stereo/data/scenes2014/>.
11. **Roy R., Pal M. and Gulati T.** Zooming Digital Images using Interpolation Techniques. *International Journal of Application or Innovation in Engineering & Management (IIAEM)*. 2013, vol 2, issue 4, pp. 34-45.
12. **Vatolin D., Putilin S.** *Ocenka kachestva metodov masshtabirovanija izobrazhenij i rezul'taty sravnenij raznyh metodov* (Quality assessment of image scaling methods and results of comparisons of different methods). GraphiCon. 2003. Proceedings. Rezhim dostupa: URL <http://graphicon.ru/oldgr/ru/publications/text/gc2003vatput.pdf>.
13. **D. Scharstein and R. Szeliski.** *Middlebury stereo evaluation*. [Elektronnyj resurs] Rezhim dostupa: <http://vision.middlebury.edu/stereo/eval/>.
14. **D. Scharstein and R. Szeliski.** A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithm. *International Journal of Computer Vision*. 2002, 47 (1-3), pp. 7-42.