Информационные процессы, Том 21, № 2, 2021, стр. 117–124
© 2021 Карнаухов, Кобер, Мозеров, Зимина.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

Робастный стерео мэтчинг с использованием фазовых признаков на основе преобразования Уолша-Адамара

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

** Институт проблем передачи информации, Российская академия наук, Москва, 127051, Россия*
*** Московский политехнический университет, Москва, 107023, Россия*

Поступила в редколлегию 11.06.2021

Аннотация—Основное предположение классического стерео мэтчинга заключается в утверждении, что совпадающие пиксели на стереоизображениях имеют одинаковые значения яркости. Однако, это предположение в целом неверное, если учесть наличие шума на изображениях и различия в освещенности левого и правого изображений стереопары. Кроме того, попиксельное сопоставление не работает в областях, где нет реальной текстуры. Поэтому важно разработать и использовать робастные локальные признаки, основанные на некоторой окрестности сопоставляемых пикселей. В данной работе мы разработали и предложили использовать абсолютно новый локальный признак, основанный на преобразовании Уолша-Адамара. Преобразование в локальном окне 8×8 легко кодируется в 64-битную величину, сравнимую по расстоянию Хэмминга. Точность подобного мэтчинга превосходит использование близкого по принципу робастного сенсус-признака. Более того, комбинирование обоих признаков позволяет достигнуть наилучших результатов стерео мэтчинга для методов, не основанных на схемах глубокого обучения.

КЛЮЧЕВЫЕ СЛОВА: Стерео мэтчинг, преобразование Уолша-Адамара, робастные признаки.

1. ВВЕДЕНИЕ

Стерео мэтчинг используются во многих приложениях и остается одной из самых сложных открытых проблем в компьютерном зрении [1–3]. Методы стерео играют ключевую роль в дополненной реальности, в управлении транспортными средствами, 3D-реконструкции и обработке медицинских изображений. Тем не менее, стерео согласование в практических приложениях часто зависит от многих факторов, таких как шум изображения, низкая текстура, окклюзия, различное освещение и экспозиция. Поэтому возникает необходимость разработки робастных локальных признаков.

Самый распространенный способ решения проблемы стерео согласования: минимизация энергии. Решение должно сходиться к максимальной апостериорной вероятности (MAB) или к минимуму энергии функционала условного случайного поля, определенного над графом изображения, где пиксели или участки изображения (суперпиксели) являются вершинами, а попарные отношения между пикселями обычно кодируется ребрами графа [4]. Локально связанная модель рассматривалась в [5–8]. Позднее стали применять глобальную модель связанности [9].

Модель минимизации энергии требует вычисления функции стоимости сопоставления или функции рассогласования. Формирование функции рассогласования — важный компонент любого алгоритма стерео мэтчинга. Более того, функция рассогласования является основным элементом любого мэтчинга, включая задачу определения оптического потока. В последнем случае — это вычисление и минимизация энергии в многомерном пространстве 3D области

(для стерео это 2D). В этом случае вычислительная сложность расчета функции рассогласования является чрезвычайно высокой [10] и кодировка признака окна окрестности в битовое слово является крайне желательной. Для этой цели мы разработали такой признак, который уменьшает вычислительную сложность формирования функции рассогласования.

Традиционно функции рассогласования можно разделить на две группы. Первая группа основана на мерах несходства сопоставления пикселей [11]. Вторая группа основана на непараметрических преобразованиях, таких как ранг и сенсус-признак [12] или нормализованная взаимная корреляция [13]. Недавний прогресс в области свёрточных нейронных сетей (CNN) привел к вычислению более надежных функций рассогласования для стерео [14]. Такая функция рассогласования позволила достигнуть результатов, существенно превосходящих точность алгоритмов, использующих классические функции рассогласования. Однако алгоритмы, использующие глубокое обучение, требуют достаточно серьезных вычислительных ресурсов, включая карты видеопроцессора и значительную оперативную память. Кроме того, сетевые алгоритмы зависят от базы данных, на которых производилось обучение. Таким образом эти алгоритмы менее гибкие и не предполагают использование устройств с ограниченными ресурсами уровня мобильного телефона, например.

В этой работе мы мотивированны именно улучшением классических методов стерео мэтчинга. Мы разработали и предложили использовать абсолютно новый локальный признак, основанный на преобразовании Уолша-Адамара. Для представления идеи предлагаемого признака сопоставим корреляцию преобразования Уолша-Адамара на бинарном гиперкубе со стандартной корреляцией на основе преобразования Фурье. Обе корреляции позволяют обнаружить максимум согласования между двумя коррелирующими признаками на левом и на правом стереоизображениях. Более того, чисто фазовая корреляция Фурье спектров окон дает более точное значение пика, поэтому по аналогии мы предлагаем использовать в преобразовании Уолша-Адамара чистую фазу, точнее знак спектра. Таким образом преобразование в локальном окне 8×8 легко кодируется в 64-битную величину, сравнимую по расстоянию Хэмминга. Как будет показано в экспериментальной части точность определения диспаратности нашего подхода превосходит использование близкого по принципу и популярного робастного сенсус-признака [12]. Более того, если при формировании функции рассогласования использовать сразу оба признака, то это позволяет достигнуть наилучших результатов стерео мэтчинга для методов, не основанных на схемах глубокого обучения.

Статья организована следующим образом: в разделе 2 дается описание робастного признака для предложенного алгоритм стерео мэтчинга, основанного на преобразовании Уолша-Адамара, в разделе 3 приведены экспериментальные результаты и наконец, раздел заключение суммирует наши выводы.

2. ПРИНЦИП ФОРМИРОВАНИЯ И КОДИРОВАНИЯ ПРИЗНАКА УОЛША-АДАМАРА

Для описания идеи нашего метода, необходимо рассмотреть один из самых популярных локальных признаков мэтчинга на изображениях: сенсус-признак [12].

Пусть $I_{i_n \in N_{i_0}} = [i_0, i_1, .., i_n, i_{N-1}]$ — значение сигнала на изображении в некоторой окрестности N пиксела p_0 (для сенсус-признака это обычно окно 8×8 пикселей) со значением яркости на изображении в этом пикселе i_0. Тогда сенсус-признаком окрестности N на изображении называется бинарный вектор $B^N = [b_0, b_1, .., b_{N-1}]$ такой что:

$$b_n = \begin{cases} 1 \text{ if } i_n > i_0, \\ 0 \text{ elsewhere.} \end{cases} \tag{1}$$

Вычислительные преимущества такого признака очевидны: вместо сравнения двух окрестностей на разных изображениях достаточно сравнить два битовых слова по метрике Хемминга:

$$D\left(B_p, B_q\right) = \sum_{k=0}^{N-1} b_k^p \oplus b_k^q, \tag{2}$$

где $\oplus$ — исключающее или. Преимущества такого признака в плане точности мэтчинга показал опыт работы с этим признаком многих стерео алгоритмов и методов. По сравнению с прямым мэтчингом двух окрестностей сенсус-признак не зависит от изменения освещенности на правом и левом изображениях стереопары. Также сенсус-признак более робастен по сравнению с полным значением сигнала и на границах областей с разными диспаратностями. В самом деле, наличие окклюзий на границе двух областей с разными диспаратностями, приводит к тому, что ода половина окрестности полностью не совпадает с подобной на другом изображении, в то время как сенсус-признак задает только полукачественную характеристику: одна половина окрестности ярче другой. И на границе областей с разными диспаратностями такое соотношение сохраняется как на левом, так и на правом стереоизображениях. Попытка видоизменить формулу (1), например, введением различных порогов сравнения с центральным элементом окрестности, не приводила к успеху. Поэтому мы искали принципиально другой подход в формировании робастного признака, при этом мы хотели сохранить идею трансформации окрестности пиксела в кодовое бинарное слово. Известно, что в распознавании образов хорошо зарекомендовали себя корреляционные фильтры на основе преобразования Фурье. Более того, только-фазовые фильтры часто оказываются предпочтительней обычных. Фаза в преобразовании Фурье такая же полноценная переменная, требующая плавающей переменной, как и амплитуда. Поэтому мы выбрали преобразование Уолша-Адамара, как базовое для получения локального признака. В самом деле, для пространственного бинарного гиперкуба, преобразование Фурье превращается в преобразование Уолша-Адамара, а корреляционный фильтр в частотной области Фурье-спектра является полным аналогом корреляционного фильтра Уолша-Адамара в том же самом пространстве. Напоминаем, что преобразование Уолша-Адамара (ПУА) [15] — это обобщенная форма преобразований Фурье. ПУА — ортогональное несинусоидальное преобразование, которое используется при сглаживании изображений, обработке речи и анализе медицинских сигналов [16]. Матрица Адамара второго порядка имеет вид

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3}$$

Поэтому если рассматривать некоторую окрестность, как элементы гиперкуба порядка $N = 2^L$, то быстрое преобразование может быть естественно реализовано по ребрам такого гиперкуба и состоять исключительно из операций сложения и вычитания. Более того, фаза в частотной области такого преобразования имеет лишь знак плюс или минус.

Итак, получение значений признака Уолша-Адамара (УАП) состоит из следующих шагов:
1. Преобразование значений некоторой окрестности изображения $I_{i_n \in N_{i_0}}$ в коэффициенты преобразования:

$$\mathrm{I}_N = [\iota_0, \iota_1, .., \iota_{N-1}] = \mathbf{H}^L I_{i_n \in N_{i_0}}. \tag{4}$$

2. Формирование бинарного слова-признака размерности 2^L (в нашем случае 64-битное слово, что соответствует окрестности 8×8) по следующему правилу:

$$b_n = \begin{cases} 1 \text{ if } \iota_n > 0, \\ 0 \text{ elsewhere.} \end{cases} \tag{5}$$

Здесь бит b_n в слове B^N имеет тот же смысл, что и в уравнении (1).

Чтобы добавить интуитивного понимания, почему УАП имеет сходство с сенсус-признаком и одновременно не коррелирует с ним, рассмотрим, что означает знак во втором коэффициенте разложения Уолша-Адамара. Если разбить окно локального признака на две симметричные части, то сумма значений сигнала на изображении в одной половине окна может быть больше или меньше аналогичной суммы во второй половине окна. Если больше, то знак плюс, если меньше, то знак минус. Иными словами, если в битовом слове сенсус-признака каждый бит отвечает за соотношение больше–меньше каждого пиксела окрестности, то в битовом слове УАП это соотношение различных полусумм той же окрестности. Именно поэтому два рассматриваемых признака не коррелируют друг с другом, что позволяет увеличить точность мэтчинга при одновременном использовании обоих признаков. Это и будет показано в следующем разделе статьи.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Экспериментальная часть была задумана таким образом, чтобы показать основные достоинства предложенного алгоритма. Она поделена на два раздела где:

- мы анализируем численные результаты и преимущества использования предложенного комбинированного признака, а часть результатов экспериментов представлена в иллюстрациях статьи чтобы читатель мог сравнить качество работы метода визуально;
- мы анализируем численные результаты использования предложенного комбинированного признака для стереоизображений с добавленным аддитивным шумом, где часть результатов так же представлена в иллюстрациях.

3.1. Сравнительные эксперименты с численным и визуальным анализом результатов без аддитивного шума

В этой части экспериментальной секции мы сравниваем результаты, полученные с помощью четырех различных признаков: IG — комбинированная попиксельная функция рассогласования яркости и градиента изображения, CNS — функция рассогласования на основе локального сенсус-признака, WH — функция рассогласования на основе локального признака Уолша-Адамара, CNS+WH — комбинированная функция рассогласования на основе локальных признаков Уолша-Адамара и сенсус-признака. Эксперимент организован следующим образом: в вычислительной схеме известного алгоритма [9] мы заменяем изначальную функцию рассогласования на четыре различные функции, описанные выше, и применяем к стерео изображениям из базы данных Мидлберри [17]. Затем, полученные результаты, усредненные по всем 15 стерео изображениям из базы данных Мидлберри [17], сравниваются в Таблице 1 по четырем критериям стерео: 2px nocc, 2px occ — процент превышения ошибки диспаратности только в видимых регионах стерео изображения и для всего изображения соответственно; avrg nocc, avrg occ — средняя ошибка только в видимых регионах стерео изображения и для всего изображения соответственно. В эту таблицу мы также добавили результаты, полученные с помощью CNNs — функции рассогласования на основе обученных сетей и метода MeshStero — одного из лучших алгоритмов стерео мэтчинга [18], не использующих функцию рассогласования на основе обучающих нейросетей.

Итогом эксперимента можно считать тот факт, что предложенный локальный признак на основе ПУА позволяет получить более точные результаты по сравнению с популярным сенсус-признаком. Кроме того, комбинированная функция рассогласования превосходит точность как WH, так и CNS признаков, используемых отдельно и этот результат превосходит точность метода [18] по критерию средней ошибки. Визуально результаты можно сравнить на Рис. 1.

Таблица 1. Результаты сравнения (усредненные по всем 15 стерео изображениям из базы данных Мидлберри [17]) по четырем критериям стерео: 2px nocc, 2px all — процент превышения ошибки диспаратности только в видимых регионах стерео изображения и для всего изображения соответственно; avrg nocc, avrg all — средняя ошибка только в видимых регионах стерео изображения и для всего изображения соответственно. Здесь IG — комбинированная попиксельная функция рассогласования яркости и градиента изображения, CNS — функция рассогласования на основе локального сенсус-признака, WH — функция рассогласования на основе локального признака Уолша-Адамара, CNS+WH — комбинированная функция рассогласования на основе локальных признаков Уолша-Адамара и сенсус-признака, CNNs — функция рассогласования на основе обученных сетей. MeshStero — результат одного из лучших алгоритмов стерео мэтчинга [18], не использующих функцию рассогласования на основе обучающих нейросетей.

Признаки:	IG	CNS	WH	CNS+WH	MeshStereo	CNNs
2px nocc	30.12	18.92	18.65	16.15	13.22	8.84
2px all	36.16	25.86	25.46	23.34	19.80	15.05
avrg nocc	9.08	8.21	6.11	4.35	4.69	2.23
avrg all	11.84	10.41	8.19	6.67	8.12	5.19

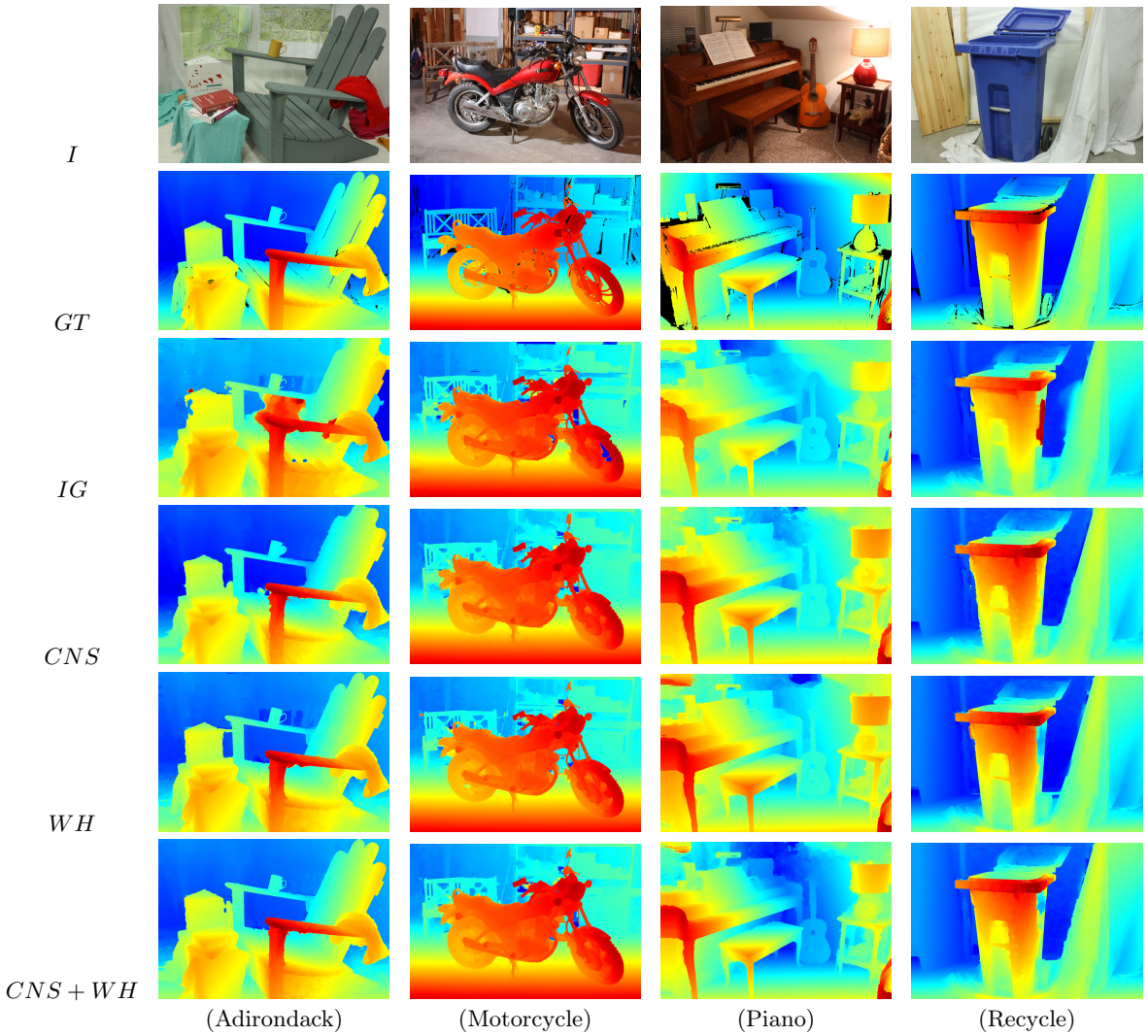

Рис. 1. Результат стерео мэтчинга для визуального сравнения. С использованием четырех стерео изображений из базы данных Мидлберри [17].

3.2. Сравнительные эксперименты с численным и визуальным анализом результатов с наложением аддитивного шума

В этой части экспериментальной секции мы сравниваем результаты, полученные с помощью тех же четырех различных признаков, что и в предыдущей секции: IG — комбинированная попиксельная функция рассогласования яркости и градиента изображения, CNS — функция рассогласования на основе локального сенсус-признака, WH — функция рассогласования на основе локального признака Уолша-Адамара, CNS+WH — комбинированная функция рассогласования на основе локальных признаков Уолша-Адамара и сенсус-признака. Эксперимент организован следующим образом: в вычислительной схеме известного алгоритма [9] мы заменяем изначальную функцию рассогласования на четыре различные функции, описанные выше, и применяем к стерео изображениям из базы данных Мидлберри [17]. Только теперь мы добавляем к одному из стерео изображений аддитивный шум 5%. Затем, полученные результаты, усредненные по всем 15 стерео изображениям из базы данных Мидлберри [17] сравниваются в Таблице 2 по четырем критериям стерео. В эту таблицу мы не включили результаты, полученные с помощью CNNs — функции рассогласования на основе обученных сетей и метода MeshStereo как нерелевантные этой экспериментальной части. Итогом эксперимента можно

Таблица 2. Результаты сравнения при наличии аддитивного шума 5%(усредненные по всем 15 стерео изображениям из базы данных Мидлберри [17]) по четырем критериям стерео: 2px nocc, 2px al — процент превышения ошибки диспаратности только в видимых регионах стерео изображения и для всего изображения соответственно; avrg nocc, avrg all — средняя ошибка только в видимых регионах стерео изображения и для всего изображения соответственно. Здесь IG — комбинированная попиксельная функция рассогласования яркости и градиента изображения, CNS — функция рассогласования на основе локального сенсус-признака, WH — функция рассогласования на основе локального признака Уолша-Адамара, CNS+WH — комбинированная функция рассогласования на основе локальных признаков Уолша-Адамара и сенсус-признака.

Признаки:	IG	CNS	WH	CNS+WH
2px nocc	62.12	34.92	36.65	32.9
2px all	66.16	40.86	42.46	38.7
avrg nocc	25.08	13.21	16.11	19.17
avrg all	29.84	17.41	20.19	21.67

считать тот факт, что предложенный локальный признак на основе ПУА хоть и более робастен к аддитивному шуму по сравнению с попиксельной функцией рассогласования IG, тем не менее, уступает функции рассогласования с сенсус-признаком. Однако, комбинированная функция рассогласования превосходит как WH, так и CNS признаков, используемых отдельно даже при наличии аддитивного шума. Визуально результаты можно сравнить на Рис. 2.

ЗАКЛЮЧЕНИЕ

В данной работе мы предложили использовать абсолютно новый локальный признак для стерео мэтчинга, основанный на преобразовании Уолша-Адамара. Как итог, использование предложенного локального признака позволяет получить более точные результаты стерео мэтчинга по сравнению с популярным сенсус-признаком. Более того, комбинирование обоих признаков позволяет достигнуть наилучших результатов стерео мэтчинга для методов, не основанных на схемах глубокого обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Scharstein D., Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // International journal of computer vision. 2002. Vol. 47, no. 1-3. P. 7–42.

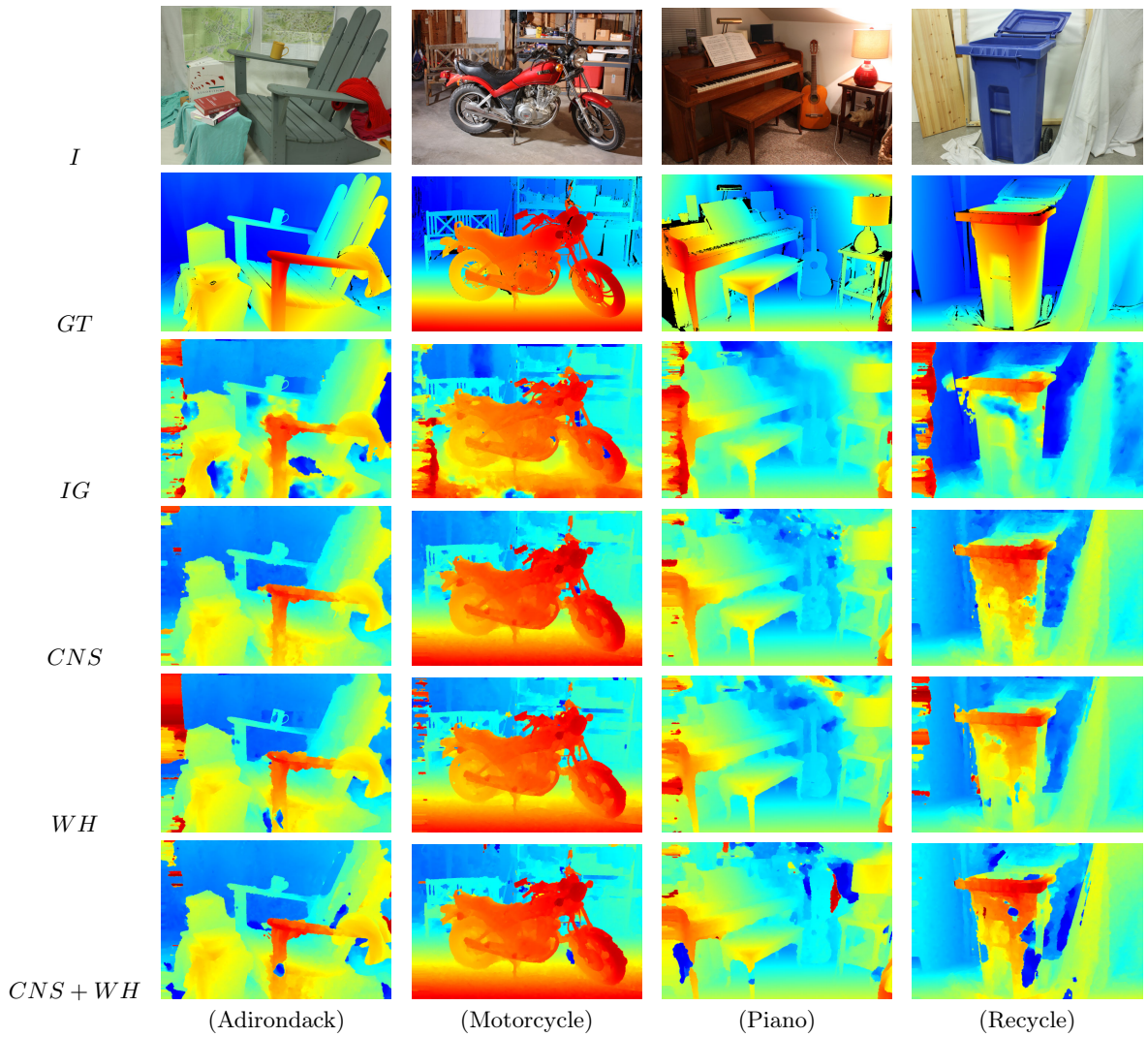

Рис. 2. Результат стерео мэтчинга для визуального сравнения при наличии аддитивного шума 5%. С использованием четырех стерео изображений из базы данных Мидлберри [17].

2. 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.

3. 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.

4. Kolmogorov V., Zabin R. What energy functions can be minimized via graph cuts? // IEEE transactions on pattern analysis and machine intelligence. 2004. Vol. 26, no. 2. P. 147–159.

5. 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.

6. Kolmogorov V., Zabih R. Computing visual correspondence with occlusions using graph cuts // ICCV. 2001. P. 508–515.

7. 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.

8. 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.

9. 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.

10. Mozerov M. Constrained optical flow estimation as a matching problem // IEEE Transactions on Image Processing. 2013. Vol. 22, no. 5. P. 2044–2055.

11. 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.

12. Zabih R., Woodfill J. Non-parametric local transforms for computing visual correspondence // ECCV. 1994. P. 151–158.

13. Lewis J. P. Fast template matching // Vision Interface. 1995. P. 120–123.

14. Zbontar J., LeCun Y. Stereo matching by training a convolutional neural network to compare image patches // Journal of Machine Learning Research. 2016. Vol. 17, no. 1-32. P. 2.

15. Fino B. J., Algazi V. R. Unified matrix treatment of the fast walsh-hadamard transform // IEEE Transactions on Computers. 1976. Vol. 25, no. 11. P. 1142–1146.

16. Andrushia A. D., Thangarjan R. Saliency-based image compression using walsh–hadamard transform (wht) // Biologically rationalized computing techniques for image processing applications. Springer, 2018. P. 21–42.

17. Scharstein D., Hirschmüller H., Kitajima Y. et al. High-resolution stereo datasets with subpixel-accurate ground truth // German conference on pattern recognition / Springer. 2014. P. 31–42.

18. Zhang C., Li Z., Cheng Y. et al. Meshstereo: A global stereo model with mesh alignment regularization for view interpolation // Proceedings of the IEEE International Conference on Computer Vision. 2015. P. 2057–2065.

Robust Stereo Matching Using Phase Features Based on the Walsh-Hadamard Transform

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

The basic assumption of classic stereo matching is 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. Furthermore, pixel-by-pixel mapping does not work in areas where there is no real texture. Therefore, it is important to develop and use robust local features based on a certain neighborhood of the pixels being matched. In this work, we have developed and proposed to use a completely new local feature based on the Walsh-Hadamard transformation. The transformation in a local 8X8 window is easily encoded into a 64-bit value comparable in Hamming distance. The accuracy of such a matching is superior to the use of a robust sensus feature that is close to the principle. Moreover, combining both features allows to achieve the best stereo matching results for methods that are not based on deep learning schemes. .

KEYWORDS: Stereo matching, Walsh-Hadamard transform, robust features. english