УДК 778.534.19
ББК 37.95

Мухамедов А.М., Ватолин Д.С., Долганов С.В.

АВТОМАТИЧЕСКИЙ ВЫБОР АЛГОРИТМА ПОСТРОЕНИЯ КАРТ ДИСПАРАТНОСТИ ДЛЯ СТЕРЕОПАР С АРТЕФАКТАМИ СЪЁМКИ ИЛИ КОНВЕРТАЦИИ, УСТОЙЧИВОГО К ДАННЫМ АРТЕФАКТАМ

Мухамедов Артур Мансурович
E-mail: arthur.mukhamedov@graphics.cs.msu.ru
Московский государственный университет им. М.В. Ломоносова

Ватолин Дмитрий Сергеевич, кандидат физико-математических наук
E-mail: dmitriy@graphics.cs.msu.ru
Московский государственный университет им. М.В. Ломоносова

Долганов Станислав Викторович
E-mail: sdolganov@graphics.cs.msu.ru
Московский государственный университет им. М.В. Ломоносова

Ежегодно выпускается большое количество фильмов в стереоформате. В процессе создания стереоконтента возникают специфические проблемы, отражающиеся на качестве получаемого продукта. Автоматические системы анализа фильмов во многих случаях опираются на карты диспаратности, алгоритмы построения которых работают хуже при наличии определённых артефактов, что может привести к снижению точности анализа. В данной работе предлагается по стереопаре выбирать алгоритм, устойчивый к имеющимся на ней артефактам. При рассмотрении 28 модификаций пяти алгоритмов точность выбора на вручную размеченном наборе данных составила 0,907.

Ключевые слова: карта диспаратности, оценка движения, стерео, видео, оценка качества.

ВВЕДЕНИЕ

Ежегодно выпускаются десятки стереофильмов, однако их популярность постепенно снижается, а зрители предпочитают смотреть в кинотеатрах их 2D-версии. Одной из возможных причин является головная боль, часто возникающая у зрителей. Её причиной могут быть появляющиеся в стереовидео искажения. В подверженных им стереокадрах правый и левый ракурсы различаются нехарактерным для человеческого зрения образом, что и вызывает дискомфорт. При этом даже высокобюджетные фильмы могут содержать, например, геометрические или цветовые несоответствия между ракурсами. Один из способов борьбы с искажениями — контролировать качество контента на этапе его производства. Для этого создаются специальные системы автоматического анализа [9], которые для обнаружения искажений нередко используют карты диспаратности. Однако, в некоторых случаях, эти искажения не соответствуют предположениям, используемым алгоритмами построения карт диспаратности, например, об отсутствии вертикального параллакса и несоответствий цвета между ракурсами. Для повышения устойчивости к таким нарушениям требуется добавлять новые компоненты в функции затрат для методов как локального, так и глобального сопоставления. С другой стороны, учёт слишком большого количества артефактов негативно отражается на скорости вычислений в тех случаях, когда этих артефактов нет. В данной работе предлагается альтернативный подход — использовать несколько различных алгоритмов, каждый из которых устойчив к своему набору артефактов, выбирая оптимальный для каждой сцены.

ПРЕДЛОЖЕННЫЙ АЛГОРИТМ

Поставим задачу формально. Даны две видеопоследовательности $\{L_t\}_{t=0}^{N}$ и $\{R_t\}_{t=0}^{N}$ длиной N кадров для левого и правого ракурсов стереоскопического видео. Необходимо каждой стереопаре поставить в соответствие алгоритм, строящий карту диспаратности с наименьшей ошибкой. Поскольку при отсутствии искажений разница между результатами работы различных алгоритмов значительно меньше, чем при их наличии, стереокадры с ошибкой

следует пометить только во втором случае. Способ выбора метки для каждой стереопары описывается ниже.

Сначала строится карта диспаратности одним из имеющихся методов. Для этого используется блочный алгоритм оценки движения, описанный в [8], так как он применяется при анализе некоторых типов искажений [10], после чего полученная карта фильтруется [11]. По полученной карте диспаратности строится карта доверия, которая вычисляется на основе оценки LRC (Left-Right Consistency check) [1] и дисперсии яркости внутри каждого блока. Таким образом, пиксель имеет низкое значение доверия в одном из трёх случаев:

— пиксель находится в низкотекстурированной области (в этом случае ему нельзя однозначно сопоставить соответствующий ему пиксель из другого ракурса);

— пиксель находится в области открытия, т. е. в другом ракурсе не существует соответствующего ему пикселя;

— пиксель не находится в низкотекстурированной области или области открытия (в таком случае он сопоставлен неточно).

Обычно области открытия занимают небольшую часть кадра. В противном случае структура рассматриваемого кадра может быть сложной для сопоставления, т. е. среднее низкое значение доверия может является одним из признаков неправильного сопоставления.

Кроме карты доверия для оценки точности сопоставления используется карта ошибки компенсации. Для этого вычисляется скомпенсированный левый ракурс смещением правого ракурса по векторам сопоставления левого. Затем из скомпенсированного кадра поэлементно вычитаются значения пикселей левого ракурса для получения искомой карты. При отсутствии цветовых искажений в рассматриваемой стереопаре, чем ближе каждый из элементов к нулю, тем точнее сопоставление. При их наличии даже при точном сопоставлении полученная карта будет состоять в основном из ненулевых элементов, однако если искажение линейное (т. е. вызвано неправильными настройками камер, а не особенностями сцены), то её элементы будут пропорциональны соответствующим элементам матрицы левого ракурса с одинаковыми коэффициентами. Чтобы это учесть, вводится ещё один признак, получаемый следующим образом: карта ошибки компенсации левого ракурса поэлементно делится на матрицу, состав-

ляющую левый ракурс, после чего по каждой из трёх цветовых компонент вычисляется дисперсия по всему кадру. Если она мала, то левый ракурс и карта ошибки компенсации действительно пропорциональны, а кадр сопоставлен правильно, но содержит цветовые искажения.

Суммарно для проверки сопоставимости стереокадра используются среднее значение доверия и число пикселей с низким доверием, получаемые из карты доверия, а также среднее значение ошибки компенсации и дисперсия относительной ошибки, получаемые из карты ошибки компенсации. Они сравниваются с пороговыми значениями, при превышении которых считается, что кадр сопоставлен неточно. В противном случае предполагается, что исходная карта диспаратности построена точно, и стереокадру назначается метка использованного алгоритма.

Далее из всех проблемных стереокадров извлекаются следующие характеристики:

— GIST [7] для обоих ракурсов;

— число углов, обнаруженных методом Харриса [2];

— гистограмма векторов диспаратности, разбитых на 3 корзины по вертикальной компоненте и 7 корзин по горизонтальной;

— среднее значение горизонтального параллакса;

— число пикселей, в которых ошибка LRC составляет более 1 пикселя;

— гистограммы интенсивности цвета в обоих изображениях, а также их абсолютная разность.

Эти характеристики объединяются в итоговый дескриптор стереокадра.

Выбор алгоритма построения карты диспаратности происходит на основе бинарной классификации. Для каждого из методов классы определяются независимо, исходя из точности сопоставления этим методом, т. е. в каждом классе находятся стереокадры, точность сопоставления которых такая же, как у стереокадров без артефактов. Таким образом, для рассматриваемого стереокадра по его дескриптору предсказывается класс для каждого из методов. Затем из набора методов, предсказанных как точно сопоставляющие данный кадр, выбирается имеющий наибольшую вероятность выбора этого класса, он и является выходным значением предложенного алгоритма.

ИСПОЛЬЗУЕМЫЕ АЛГОРИТМЫ СОПОСТАВЛЕНИЯ

В случае, если при анализе стереокадра с некоторым искажением среди рассматриваемого набора алгоритмов не будет устойчивого к нему, сопоставление будет неточным вне зависимости от выбранного алгоритма. Таким образом, очень важно, чтобы в рассматриваемом наборе встречались алгоритмы, которые по своей архитектуре устойчивы к наиболее распространённым искажениям. Рассмотрим такие искажения.

Несовпадение цвета в ракурсах. При данном искажении блоки в разных ракурсах, соответствующие одному объекту, отличаются по значениям цветовых компонент, нарушая предположение о постоянстве цвета. При равномерном искажении устойчивость можно обеспечить, учитывая разность интенсивностей между блоками, например, используя функцию затрат ZSAD [3]. В более сложных случаях, таких как разная засветка камер источником света, возникает нелинейное цветовое искажение между пикселями блока. В этом случае попытка точного выравнивания цветов при сопоставлении сильно замедляет алгоритм. Устойчивость в этом случае можно обеспечить использованием функций стоимости, не зависящих от абсолютных значений цветов. Такими, например, являются преобразования Census и Rank, строящие дескриптор на основе относительных значений цветов внутри блока, а также нормализованная взаимнокорреляционная функция (NCC).

Несовпадение резкости между ракурсами. Данное искажение, как и предыдущее, снижает точность сопоставления блоков, играя роль шума при сравнении малых окрестностей пикселя. Если же сопоставляемая окрестность большая, то зависимость от данного шума снижается, однако захватывается больше пикселей, принадлежащих другим объектам, величина диспаратности которых может отличаться от величины у сопоставляемого объекта. Устойчивость к шуму имеется, например, в NeedleMatch [6], использующего окрестности нескольких различных размеров с одинаковыми весами.

Наличие вертикального параллакса, несовпадение угла поворота и несоответствие масштаба камер. При наличии артефактов данного типа нарушается предположение об отсутствии вертикального параллакса между двумя ракурсами, которое часто

используется при построении карт диспаратности через уменьшение множества допустимых смещений, ограничивая их одной осью. Учёт возможности таких искажений приближает эту задачу к построению карты оптического потока, которая успешно решается многими алгоритмами. При этом скорость работы можно увеличить, ограничив диапазон возможных значений смещения по вертикали.

Большие значения параллакса между двумя ракурсами для объекта. Хотя эта ситуация не является искажением, она может привести к снижению качества работы методов, использующих ограничения на длину максимального значения сдвига. В этом случае истинное значение сдвига может превысить используемый порог. Кроме этого, объекты с большим параллаксом часто находятся на переднем плане и занимают значительную часть кадра. Из-за этого часть некоторого объекта в одном ракурсе может быть сопоставлена с другой, похожей частью этого же объекта в другом ракурсе, поскольку она находится ближе к исходной. В этом случае, как и при несовпадении резкости между ракурсами, точность сопоставления можно повысить, анализируя пирамиду изображений либо уменьшив исходное изображение. Во втором случае потери информации будут не очень велики из-за того, что проблемный объект занимает большую часть кадра.

С учётом описанного, выбор оптимального алгоритма происходит из следующих методов:

— с функциями затрат ZSAD и NCC [8];

— иерархически работающий с пирамидой изображений, с функцией затрат ZSAD [8];

— Coarse-to-Fine PatchMatch [4];

— DISflow [5] с и без сглаживающей постобработки;

— NeedleMatch с использованием функций стоимости ZSAD и SAD при сопоставлении блоков со значениями масштаба 0,5 и 0,75 и 2 и 4 уровнями.

Кроме того, для каждого из перечисленных методов рассматривался аналогичный, но с предобработкой в виде уменьшения входной пары изображений в 4 раза при сохранении размеров сравниваемых блоков. Таким образом, всего рассматривались 28 подходов, основанных на 5 различных алгоритмах.

СОСТАВЛЕНИЕ ВЫБОРКИ

Используемая выборка была получена в результате анализа фрагментов стереофильмов с проблемными кадрами. Далее она расширялась в процессе поиска подобных проблемных кадров в других стереофильмах. Суммарно была получена выборка из 1471 стереопары из 35 стереофильмов, каждое изображение имеет разрешение 960×400. Из них 210 стереопар имеют значительные артефакты, не позволяющие построить точную карту диспаратности ни одним из рассматриваемых методов. Для использования при обучении бинарная классификация всей выборки для каждого из методов производилась вручную.

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

Для тестирования детектора неправильно сопоставленных кадров проводился анализ полнометражных стереофильмов. Было проанализировано 14 фильмов, в которых 194 сцены были классифицированы как содержащие неправильно сопоставленные кадры. В двух фильмах ни в одной сцене не были обнаружены подобные кадры. Результаты анализа представлены на рис. 1.

Заметим, что большое число ложноположительных срабатываний возникало в сценах, содержащих большие области открытия.

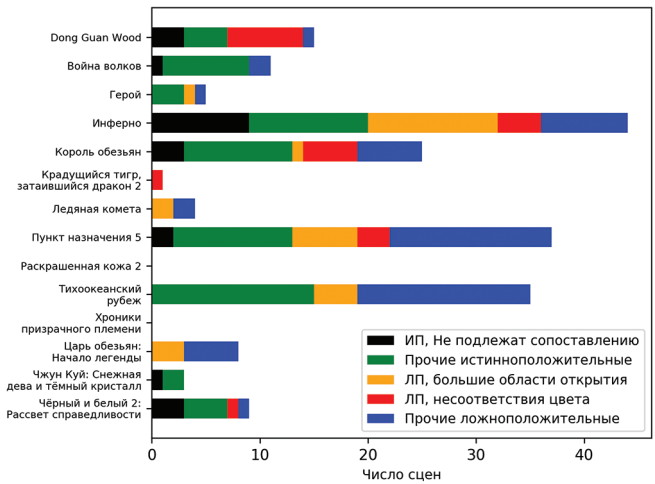

Рис. 1. Результаты работы детектора проблемных кадров

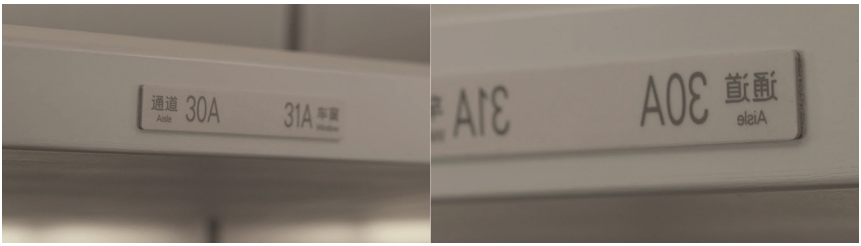

Рис. 2. Пример кадра, не подлежащего сопоставлению. Кадр 120 901 из фильма «Dong Guan Wood»

Ещё одним источником ложноположительных срабатываний являются цветовые искажения, в основном нелинейного характера или затрагивающие только часть кадра. Часть истинноположительных срабатываний была помечена как не подлежащая сопоставлению. В таких случаях левый и правый ракурсы различаются настолько, что они не содержат общих объектов либо невозможны с точки зрения расположения камер, как показано на рис. 2.

Точность предсказаний алгоритмов оценивалась с помощью кросс-валидации. Были оценены три различных метода классификации: метод опорных векторов, случайные леса, градиентный бустинг. Наибольшая точность (0,907±0,0161) была достигнута градиентным бустингом, как показано на рис. 3.

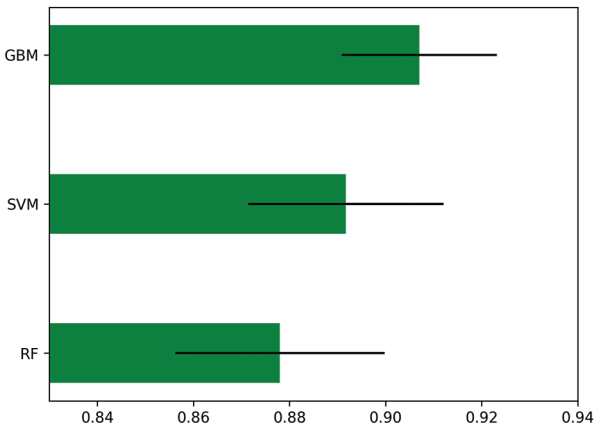

Рис. 3. Точность кросс-валидации для предсказаний алгоритмов

ЗАКЛЮЧЕНИЕ

Предложен подход для поиска кадров, для которых используемый алгоритм построения карт диспаратности обеспечивает низкую точность сопоставления. Также предложен алгоритм, выбирающий метод сопоставления на основе обученной модели, обеспечивающий бóльшую точность. Алгоритм был применён для анализа стереофильмов. По результатам работы были выявлены сложные случаи для детектора. Планируемое развитие алгоритма включает в себя увеличение точности работы в сложных случаях, увеличение выборки для обучения классификаторов, классификация невосстанавливаемых кадров, а также поиск набора характеристик, позволяющих выбирать алгоритм без предварительного построения карты диспаратности.

ЛИТЕРАТУРА

1. Egnal G., Wildes R.P. Detecting binocular half-occlusions: Empirical comparisons of five approaches // IEEE Transactions on pattern analysis and machine intelligence. 2002. Vol. 24. No 8. P. 1127–1133.

2. Harris C., Stephens M. A combined corner and edge detector // Alvey vision conference. 1988. Vol. 15. No 50. P. 10.5244.

3. Hermann S., Klette R. The naked truth about cost functions for stereo matching. Department of Computer Science, University of Auckland, 2008.

4. Hu Y., Song R., Li Y. Efficient coarse-to-fine patchmatch for large displacement optical flow // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 5704–5712.

5. Kroeger T. et al. Fast optical flow using dense inverse search // European Conference on Computer Vision. Springer, Cham, 2016. P. 471–488.

6. Lotan O., Irani M. Needle-match: Reliable patch matching under high uncertainty // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 439–448.

7. Oliva A., Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope // International journal of computer vision. 2001. Vol. 42. No 3. P. 145–175.

8. Simonyan K. et al. Fast video super-resolution via classification // Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE, 2008. P. 349–352.

9. Video Quality Measurement Tool 3D Project. http://compression.ru/video/vqmt3d/. Дата обращения: 10.04.2018.

10. Voronov A., Borisov A., Vatolin D. System for automatic detection of distorted scenes in stereo video // Proceedings of Sixth International Workshop on Video Processing and Quality Metrics (VPQM). 2012.

11. Zhang Q., Xu L., Jia J. 100+ times faster weighted median filter (WMF) // Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014. P. 2830–2837.

Arthur M. Mukhamedov, Dmitriy S. Vatolin, Stanislav V. Dolganov

AUTOMATIC CHOICE OF DISPARITY ESTIMATION ALGORITHM FOR STEREO MATES RESISTANT TO FILMING AND CONVERSION ARTIFACTS

Arthur M. Mukhamedov
E-mail: arthur.mukhamedov@graphics.cs.msu.ru
Lomonosov Moscow State University

Dmitriy S. Vatolin, PhD (Physics and Mathematics)
E-mail: dmitriy@graphics.cs.msu.ru
Lomonosov Moscow State University

Stanislav V. Dolganov
E-mail: sdolganov@graphics.cs.msu.ru
Lomonosov Moscow State University

A lot of new 3D movies are released each year. Several specific problems arise during stereo content production that affect the quality of the final product. Automatic systems for movie analysis often rely on disparity maps that sometimes degrade due to specific artifacts decreasing matching quality. We propose a way to choose an algorithm most resistant to the artifacts on a given stereo mate. On a hand-marked dataset precision of 0,907 was reached when choosing between 28 modifications of five algorithms.

Key words: disparity map, motion estimation, stereo video, quality assessment.

REFERENCES

1. Egnal G., Wildes R.P. Detecting binocular half-occlusions: Empirical comparisons of five approaches // IEEE Transactions on pattern analysis and machine intelligence. 2002. Vol. 24. No 8. P. 1127–1133.

2. Harris C., Stephens M. A combined corner and edge detector // Alvey vision conference. 1988. Vol. 15. No 50. P. 10.5244.

3. Hermann S., Klette R. The naked truth about cost functions for stereo matching. Department of Computer Science, University of Auckland, 2008.

4. Hu Y., Song R., Li Y. Efficient coarse-to-fine patchmatch for large displacement optical flow // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 5704–5712.

5. Kroeger T. et al. Fast optical flow using dense inverse search // European Conference on Computer Vision. Springer, Cham, 2016. P. 471–488.

6. Lotan O., Irani M. Needle-match: Reliable patch matching under high uncertainty // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 439–448.

7. Oliva A., Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope // International journal of computer vision. 2001. Vol. 42. No 3. P. 145–175.

8. Simonyan K. et al. Fast video super-resolution via classification // Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE, 2008. P. 349–352.

9. Video Quality Measurement Tool 3D Project. http://compression.ru/video/vqmt3d/. Data obrashcheniya: 10.04.2018.

10. Voronov A., Borisov A., Vatolin D. System for automatic detection of distorted scenes in stereo video // Proceedings of Sixth International Workshop on Video Processing and Quality Metrics (VPQM). 2012.

11. Zhang Q., Xu L., Jia J. 100+ times faster weighted median filter (WMF) // Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014. P. 2830–2837.