

Матирование видеопоследовательностей с использованием восстановленного фона

Ерофеев Михаил Викторович¹, Ватолин Дмитрий Сергеевич²
merofeev@graphics.cs.msu.ru¹, dmitriy@graphics.cs.msu.ru²

¹ — аспирант

² — к. ф.-м. н., с. н. с.

Московский Государственный Университет им. М. В. Ломоносова,
Факультет вычислительной математики и кибернетики,
Лаборатория компьютерной графики
г. Москва

Аннотация

Задача матирования состоит в разделении данного изображения или видео на передний план, задний план и карту прозрачности переднего плана. Построение такого разделения важно для решения многих задач редактирования и обработки видео, таких как: замена фона, применение спецэффекта только к фону либо только к переднему плану, построение стереоизображения.

В работе предлагается метод матирования видеопоследовательностей, основанный на методе матирования изображений Learning Based Matting. Описывается способ модификации базового метода, позволивший использовать информацию о фрагментах фона, скрытых объектом переднего плана, в качестве дополнительных данных; а также способ итеративной фильтрации последовательности карт прозрачности. В заключение приводятся результаты сравнения предложенного алгоритма с 11 аналогами, иллюстрирующие преимущества предложенного подхода.

УДК: 004.932.4

Ключевые слова: матирование, обработка видео, восстановление фона, матирующий лапласиан

Video matting with aid of reconstructed background

Formally, matting is a problem of image decomposition into foreground image, background image and foreground transparency map. This problem is extremely important for such video and image editing problems as: background replacement, applying transform to background or foreground only, stereoscopic image generation. In this paper we propose video matting method based on Learning Based Matting method. We describe modification to base method which enabled us to use reconstructed background sequence as additional input. We also propose iterative method for spatio-temporal transparency map smoothing. Finally, we show that proposed approach outperforms 11 image and video matting methods in objective comparison.

Keywords: matting, video processing, background reconstruction, background inpainting, matting Laplacian

1 Введение

Одной из ключевых задач, возникающих при редактировании изображений и видеопоследовательностей, является построение карты прозрачности (матирования) объекта перед-

него плана для последующей замены фона или элементов фона, или изменения положения объекта относительно фона.

Формальная постановка задачи матирования имеет следующий вид: разделить данное изображение I на изображение объекта переднего плана F , изображение заднего плана B и карту прозрачности переднего плана α таким образом, чтобы было верно следующее уравнение:

$$I = \alpha F + (1 - \alpha)B \quad (1)$$

Задача матирования видео является обобщением задачи матирования изображений на видеопоследовательности. Такое обобщение порождает дополнительные требования к алгоритмам матирования видео:

- Результирующая карта прозрачности должна быть стабильна во времени, ввиду высокой чувствительности зрительной системы человека к различиям между соседними кадрами
- Алгоритм должен иметь адекватную объемам видеоданных вычислительную сложность

В данной работе предлагается алгоритм матирования видеопоследовательностей, в основу которого положен алгоритм матирования изображений на основе обучения (Learning Based Matting [1]). Отличительной особенностью предложенного алгоритма является использование изображений фона, содержащих фрагменты скрытые объектом переднего плана на исходной видеопоследовательности, в качестве дополнительных входных данных. Для построения таких изображений фона применяется алгоритм, описанный в [2], выполняющий восстановление фрагментов фона, скрытых на данном кадре, путем копирования этих же фрагментов с других кадров видеопоследовательности, где они являлись видимыми. Предлагаемый подход также выполняет два последовательных прохода по видеопоследовательности с целью обеспечения большей стабильности во времени результата в сравнении с результатом независимого применения алгоритма [1] к кадрам видеопоследовательности. Подробное описание предложенного алгоритма содержится в разделе 3.

Предложенный алгоритм был протестирован на наборе из 10 видеопоследовательностей с известными эталонными картами прозрачности (данный набор был подготовлен в рамках работы [3]) и сравнен с 11 другими методами матирования изображений и видео. Описание проведенной экспериментальной оценки содержится в разделе 4.

2 Обзор области

Исследования в области матирования видео, в сравнении с областью матирования изображений, затруднены большим объемом входных данных, необходимых для работы алгоритмов матирования видео, и необходимостью обеспечить стабильность во времени результирующей карты прозрачности. Ввиду этого исследования в области матирования видео всегда отставали от прогресса в области матирования изображений. Несмотря на это, на сегодняшний день существуют 12 оригинальных методов матирования видео, а также объективная методика сравнения алгоритмов видео, описанная в [3].

Первый из известных алгоритмов матирования видео был предложен в 2001 году [4]. В основе метода лежит алгоритм матирования изображений, описанный в работе [5]. В качестве дополнительных входных данных алгоритм использует набор тернарных масок (карт, задающих три области: гарантированный передний план, гарантированный задний план и неизвестная область, для которой требуется решить задачу матирования) для ключевых кадров, выбранных пользователем. Алгоритм выполняет распространение тернар-

ной разметки ключевых кадров на остальные кадры видеопоследовательности с использованием метода оценки оптического потока. Затем алгоритм матирования изображений [5] независимо применяется к каждому кадру видеопоследовательности. С целью повышения качества итоговой карты прозрачности авторы предложили применять простой алгоритм восстановления неизвестных фрагментов фона и в дальнейшем использовать эти фрагменты в качестве источника дополнительных образцов пикселей фона для алгоритма матирования изображений. Серьезным недостатком метода восстановления фона, использованного в работе, является требование принадлежности всех объектов фона одной плоскости, что существенно ограничивает его применимость на практике. Также к недостаткам работы следует отнести независимое применение алгоритма матирования к кадрам видеопоследовательности, что неизбежно приводит к нестабильности результата во времени, и отсутствие данных об оценке результатов работы алгоритма с помощью объективных методов.

Работа 2004 года [6] развивает идею предыдущей работы. Авторы применили методы машинного обучения к эталонным данным для построения модели функции плотности совместного распределения градиента исходной видеопоследовательности и последовательности карт прозрачности. Полученная модель легла в основу нового слагаемого в функции стоимости, оптимизируемой алгоритмом матирования изображений [5]. Ввиду того, что вектор градиента вычислялся в пространственно-временном домене, новое слагаемое повысило стабильность результата во времени. Несмотря на элегантность идеи, следует отметить, что новое слагаемое никак не учитывает движение объектов, что ограничивает применимость метода видеопоследовательностями с медленным равномерным движением. Работа также не содержит результатов объективной оценки качества.

Авторы работы 2008 года [7] предложили обобщение алгоритма сегментации изображений GrabCut [8] для обработки видеопоследовательностей. В моделях переднего плана и фона, которые в исходном методе определяли вероятность принадлежности пиксела заданного цвета к переднему плану или фону соответственно, цвет пиксела был заменен следующим вектором признаков: цвета пиксела, вектора оптического потока между текущим и предыдущим кадром, вектора оптического потока между текущим и следующим кадром. Алгоритм последовательно производит сегментацию видеопоследовательности с использованием текущих моделей переднего плана и фона, а затем уточняет модели с помощью полученной сегментации на протяжении 10 итераций. Для инициализации моделей используется грубая маска переднего плана, предоставляемая пользователем, или же данная маска строится автоматически путем анализа поля векторов движения. Очевидным достоинством метода является возможность полностью автономной работы без дополнительного ввода со стороны пользователя. К недостаткам метода следует отнести низкое визуальное качество примеров работы метода, продемонстрированных в статье, и отсутствие данных об объективном сравнении метода с аналогами.

Алгоритм матирования видеопоследовательностей [9], предложенный в 2009 году, благодаря высокому визуальному качеству результатов его работы был включен в профессиональный программный продукт для обработки видеопоследовательностей Adobe After Effects в виде инструмента Roto Brush. На первом шаге алгоритм выполняет бинарную сегментацию исходной видеопоследовательности с использованием множества локальных классификаторов, независимо примененных к небольшим перекрывающимся окнам вдоль границы объекта. Параметры обученных классификаторов переносятся на следующий кадр с помощью векторов оптического потока для обеспечения стабильности во времени результата сегментации. Затем для каждого кадра вычисляется карта прозрачности путем применения модифицированного алгоритма матирования изображений, предложенного в [10]. Предложенная модификация состоит во внедрении дополнительного слагаемого в функцию стоимости, требующего близости карты прозрачности текущего кадра к

карте прозрачности предыдущего кадра с поправкой на вектора движения. Отсутствие результатов объективной оценки качества компенсируется масштабной апробацией данного алгоритма в составе популярного коммерческого продукта.

В 2010 году был предложен алгоритм матирования видеопоследовательностей [11], обобщающий алгоритм матирования изображений [12]. Дополнительными входными данными алгоритма являются тернарные маски для ключевых кадров, которые затем переносятся на остальные кадры видеопоследовательности с использованием алгоритма вычисления оптического потока. Затем к каждому кадру применяется алгоритм матирования изображений [12], на последнем этапе которого матирующий лапласиан с окнами размером 3×3 пиксела, принадлежащими текущему кадру, заменен пространственно-временным матирующим лапласианом с окнами $3 \times 3 \times 3$ пиксела, захватывающими два соседних кадра. Тем не менее, алгоритм формирования данных способен учесть лишь малое движение объектов в видеопоследовательности. Также в работе были представлены результаты объективного сравнения предложенного метода с результатами работы алгоритма матирования изображений, примененного независимо к каждому кадру видеопоследовательности. В качестве набора тестовых данных авторы использовали одну синтетическую видеопоследовательность, полученную методами компьютерной графики, с известной эталонной картой прозрачности. Критериями сравнения являлись: L_2 расстояние результата до эталона и отношение модуля производной карты прозрачности по времени к модулю производной видеопоследовательности

$$\text{AdtIdt}(\alpha, I) = \frac{\left| \frac{d\alpha}{dt} \right|}{\left| \frac{dI}{dt} \right|}, \quad (2)$$

которое использовалось для оценки временной стабильности результата. По обоим критериям алгоритм превзошел свою базовую версию.

Алгоритм [13], который лег в основу инструмента Refine Edge программного продукта Adobe After Effects, был предложен в 2011 году. Как и алгоритм, рассмотренный выше, он использует тернарные маски для ключевых кадров в качестве дополнительных входных данных, на основании которых строит тернарные маски остальных кадров. Затем независимо к каждому кадру применяется алгоритм матирования изображений [12]. После чего к полученной последовательности карт прозрачности применяется процедура временной фильтрации, основанная на дистантном преобразовании [14]. Результаты объективной оценки качества результатов работы алгоритма не были предоставлены авторами.

Алгоритм 2012 года [15] в качестве дополнительных входных данных использует карту прозрачности первого кадра (авторы предполагают, что пользователь использует алгоритм матирования изображений для ее построения). Для каждого следующего кадра алгоритм вычисляет грубую карту прозрачности с использованием пространственно-временного матирующего лапласиана и преобразует ее с помощью алгоритма разреза графов в бинарную маску. Затем полученная маска преобразуется в тернарную с помощью модели цвета переднего плана и фона для текущего и предыдущего кадров. Для вычисления итоговой карты прозрачности кадра пространственно-временной матирующий лапласиан применяется повторно. Авторы проводят объективное сравнение результатов работы предложенного алгоритма с версией своего алгоритма, примененной независимо к каждому кадру. Критериями сравнения являются: L_1 расстояние до эталона и производная этой величины по времени. По обоим критериям было продемонстрировано превосходство предложенного метода.

В 2012 году был предложен еще один метод матирования видеопоследовательностей [16]. Ключевой особенностью этого метода является интерактивность — пользователь наносит на кадры видеопоследовательности мазки двух типов (передний план и фон), после каждого мазка алгоритм обновляет карту прозрачности. Такой подход очевидно требовал

быстрого ее пересчета, что было обеспечено использованием быстрого метода фильтрации изображений Guided Filter [17] в основе алгоритма. Алгоритм строит модели цвета переднего плана и фона и вычисляет вероятность принадлежности каждого пиксела переднему плану, затем сглаживает эти карты с помощью Guided Filter и бинаризирует результат по порогу 0,5. Полученная маска сглаживается с помощью пространственно-временной модификации алгоритма Guided Filter. Результат принимается за карту прозрачности кадра. Авторы сравнивают предложенный алгоритм с его версией, примененной независимо к каждому кадру, с использованием критерия, описанного в уравнении (2).

Также в 2012 году был предложен алгоритм [18]. Авторы работы предлагают для обеспечения стабильности результата во времени сглаживать карту прозрачности не вдоль векторов оптического потока исходной видеопоследовательности (как это делали авторы многих ранее рассмотренных работ), а вдоль векторов оптического потока карт прозрачности. В качестве первого приближения оптического потока карт прозрачности используется оптический поток исходной видеопоследовательности. Затем алгоритм вычисляет карту прозрачности каждого кадра с использованием метода [10] с дополнительным слагаемым в оптимизируемой функции стоимости, требующим постоянства значений прозрачности вдоль векторов оптического потока. Полученные карты прозрачности используются в качестве дополнительных данных для алгоритма вычисления оптического потока. Затем обновленный оптический поток используется для повторного вычисления карт прозрачности.

Заключает список методов матирования видео, предложенных в 2012 году, алгоритм [19]. Авторы вычисляют карту прозрачности с помощью нелокального матирующего лапласиана (другими словами, лапласиана, задающего связи между пикселями, не обязательно находящимися по соседству), действующего во всем видеообъеме. Очевидно, что такая постановка задачи приводит к колоссальной системе линейных уравнений. Например, для видео разрешения 1920×1080 из 300 кадров при 10% пикселей, отмеченных как неизвестные, система будет состоять из более чем 62 миллионов уравнений. Для уменьшения вычислительной сложности задачи авторы предлагают разделять исходный видеообъем на пространственно-временные параллелепипеды с использованием k-d дерева. К сожалению, авторы не приводят результатов объективного сравнения своего алгоритма с аналогами.

В 2014 году был предложен алгоритм матирования видеопоследовательностей [20], основанный на идее пространственно-временного сэмплирования. Многие методы матирования изображений (например, [21, 12]) для вычисления прозрачности пиксела строят набор образцов пикселей переднего плана и фона. Авторы предлагают строить набор таких образцов на основании сразу нескольких кадров, а также для обрабатываемого в данный момент пиксела выбирать образцы таким образом, чтобы его прозрачность была близка к прозрачности соответствующего ему пиксела прошлого кадра. Тем самым, по мнению авторов, обеспечивается стабильность карты прозрачности во времени еще на начальном этапе ее построения. Затем полученная последовательность карт прозрачности дополнительно фильтруется с помощью пространственно-временного матирующего лапласиана, построенного с учетом оптического потока, дополненного требованием близости карты прозрачности текущего кадра к карте прозрачности предыдущего с поправкой на оптический поток. Авторы сравнивают результаты работы своего метода с результатами алгоритмов [9] и [13] на наборе тестовых видеопоследовательностей с известной эталонной картой прозрачности. Критериями сравнения являются: L_2 расстояние до эталона и L_2 расстояние между производными по времени эталонной и вычисленной последовательностями карт прозрачности. По второму критерию предложенный алгоритм занял первое место в трех из четырех тестах, по первому критерию — во всех.

Алгоритм матирования видеопоследовательностей [22], основанный на идее разрежен-

ного низкорангового кодирования, был предложен в 2015 году. На первом шаге алгоритм строит словарь из двух разделов: пиксели переднего плана и фона. Каждое слово образовано цветом пиксела в системе RGB и его координатами. Словарь формируется таким образом, чтобы разреженные линейные комбинации пикселей переднего плана, включенных в словарь, хорошо описывали остальные пиксели переднего плана и плохо описывали пиксели фона (аналогично для пикселей фона). Затем каждый пиксель неизвестной области кодируется в виде линейной комбинации пикселей словаря таким образом, чтобы матрица, образованная коэффициентами этих линейных комбинаций всех пикселей неизвестной области тернарной маски, была разреженной и имела низкий ранг. Затем линейные комбинации преобразуются в значения прозрачности путем суммирования коэффициентов, стоящих при пикселях переднего плана. Наконец, карты прозрачности фильтруются способом, аналогичным предыдущей работе. Авторы сравнивают предложенный подход с методами [10], [23], [9] с использованием критерия, описанного уравнением (2), на пяти тестовых видеопоследовательностях, подготовленных в рамках работы [3]. Алгоритм занимает первое место во всех пяти тестах, тем не менее, следует заметить, что набор тестовых видеопоследовательностей, предложенный в рамках работы [3], состоит из 10 последовательностей, таким образом, результаты для других пяти тестов авторами приведены не были.

В заключение обзора области следует также упомянуть работу [3], в рамках которой было подготовлено 10 видеопоследовательностей с эталонными картами прозрачности и тернарными масками, были предложены методы объективной оценки качества карт прозрачности, а также было проведено объективное сравнение ряда существующих алгоритмов матирования изображений и видео.

3 Предложенный метод

Как уже отмечалось во Введении, отличительной особенностью предлагаемого алгоритма является использование восстановленного фона. Очевидно, что знание цвета пикселей фона B в неизвестной области тернарной маски уменьшает множество возможных решений уравнения (1). Однако в случае задачи матирования изображения можно лишь примерно оценить их цвет (что и делают многие алгоритмы матирования изображений). Ситуация кардинально меняется в случае матирования видеопоследовательности: фрагменты фона, скрытые объектом переднего плана на одном кадре, зачастую видимы на других кадрах этой видеопоследовательности. Этот факт позволяет вычислить цвет пикселей фона неизвестной области путем копирования данных с соседних кадров.

Несмотря на то, что различными авторами ранее предлагались алгоритмы матирования видеопоследовательностей, использующие восстановленный фон, данная работа имеет новизну, благодаря:

- Использованию алгоритма восстановления фона, применимого к более широкому классу видеопоследовательностей, чем алгоритмы ранних работ.
- Предлагаемому обобщению алгоритма матирования, основанного на матирующем лапласиане, использующему восстановленный фон в качестве дополнительных входных данных. Ранее в литературе восстановленный фон использовался только в методах, основанных на сэмплировании (последние существенно отличаются от первых)
- Методу вычисления доверия к векторам оптического потока, оптимизированного для задачи матирования видео

Входными данными предлагаемого в данной работе алгоритма являются: исходная видеопоследовательность, последовательность тернарных масок. Алгоритм выполняет следующие основные шаги:

1. Применить к исходной видеопоследовательности алгоритм восстановления фрагментов фона, скрытых объектами переднего плана, описанный в работе [2]
2. Применить к каждому кадру исходной последовательности модифицированную версию алгоритма матирования изображений на основе обучения [1], использующую восстановленный фон в качестве дополнительных входных данных
3. Применить итеративный алгоритм пространственно-временного сглаживания последовательности карт прозрачности

Следует отдельно отметить использование итеративного подхода к сглаживанию карт прозрачности. Формально это может быть записано следующим образом: пусть α_t^0 — кадр t последовательности карт прозрачности, вычисленной изначально; $S(\cdot)$ — оператор сглаживания, α_t^i — кадр t результата фильтрации на итерации i , тогда

$$\alpha_t^i = S(\alpha_{t-1}^{i-1}, \alpha_t^{i-1}, \alpha_{t+1}^{i-1}). \quad (3)$$

Доминирующим подходом в литературе на сегодняшний день является метод последовательной фильтрации, результат которой (α'_t) может быть записан как

$$\alpha'_1 = \alpha_1^0 \quad (4)$$

$$\alpha'_t = S(\alpha'_t - 1, \alpha_t^0, \alpha_t^0 + 1). \quad (5)$$

Как было замечено в ходе проведенных экспериментов, последовательная фильтрация склонна накапливать ошибку, то есть расстояние между вычисленной картой прозрачности и эталоном увеличивается с номером кадра. Итеративный подход лишен данного недостатка, а также допускает параллельное вычисление карт прозрачности нескольких кадров.

Далее мы рассмотрим основные шаги алгоритма более детально.

3.1 Восстановление фрагментов фона

Для восстановления фрагментов фона, скрытых передним планом, был использован алгоритм, описанный в [2]. Входными данными алгоритма являются видеопоследовательность и набор масок для каждого ее кадра, отмечающих области, подлежащие восстановлению. Результатом работы алгоритма является видеопоследовательность, совпадающая с исходной в областях, не отмеченных на масках, и содержащая фрагменты восстановленного фона в остальных областях. Дополнительными выходными данными алгоритма является набор масок, на которых отмечены области, восстановление которых не было выполнено. Принцип работы алгоритма состоит в экстраполяции поля векторов движения в область восстановления и последующем копировании фрагментов фона других кадров согласно экстраполированным векторам.

В данной работе в качестве маски восстановления используется объединение областей тернарной маски, отвечающих переднему плану и неизвестной области.

3.2 Базовый метод матирования изображений

Для лучшего понимания дальнейшего описания предлагаемого алгоритма матирования видео рассмотрим краткое описание алгоритма матирования изображений, лежащего в его основе (подробное описание содержится в работе [1]).

Входными данными алгоритма являются изображение в цветовом пространстве RGB и тернарная маска. Алгоритм выполняет обход изображения с использованием скользящего окна 3×3 пиксела. Далее для каждого окна ω , содержащего хотя бы один неизвестный пиксель, выразим прозрачность каждого его пиксела α' как линейную комбинацию w' прозрачностей его соседей α_ω :

$$\alpha' = w' \alpha_\omega. \quad (6)$$

Для получения w' воспользуемся приемом, схожим с линейной регрессией. Пусть, зная цвет пиксела $x' = [R; G; B; 1]$, требуется предсказать его прозрачность α' . Нам известны цвета всех его соседей X_ω (каждая строка матрицы состоит из цвета пиксела в системе RGB и единицы). Дополнительно предположим, что нам известны прозрачности всех его соседей в текущем окне α_ω , тогда:

$$\alpha' = x'^T \beta, \quad (7)$$

где β — столбец из четырех элементов, таких, что:

$$\beta = \arg \min_{\beta} \|X_\omega \beta - \alpha_\omega\|^2 + \epsilon \|\beta\|^2. \quad (8)$$

$\epsilon = 10^{-6}$ — параметр регуляризации. Последнее выражение содержит задачу минимизации квадратичного многочлена с положительно определенным гессианом, а значит может быть сведено к:

$$\beta = (X_\omega^T X_\omega + \epsilon I)^{-1} X_\omega^T \alpha_\omega. \quad (9)$$

Подставим правую часть (9) в уравнение (7) и сопоставим с уравнением (6), получим:

$$w' = x'^T (X_\omega^T X_\omega + \epsilon I)^{-1} X_\omega^T. \quad (10)$$

Вычислим вектор-строку w для каждого пиксела в каждом окне, содержащем хотя бы один пиксель неизвестной области. Для удобства записи дополним полученные вектора нулями, чтобы было верно:

$$\alpha' = w' \alpha, \quad (11)$$

где α — коэффициенты прозрачности всех пикселей изображения. Образую матрицу W из дополненных нулями векторов. Теперь задача вычисления карты прозрачности сводится к следующей задаче оптимизации:

$$\alpha = \arg \min_{\alpha} \|\alpha - W\alpha\| + (\alpha - T)^T D(\alpha - T), \quad (12)$$

где T — тернарная маска, D — диагональная матрица, у которой элементы главной диагонали, отвечающие пикселям неизвестной области, равны нулю, а остальные — 10^6 . Решение последней задачи может быть сведено к решению разреженной системы линейных уравнений. Ее решение и есть результат работы базового метода матирования изображений.

3.3 Использование восстановленного фона

Рассмотрим, как базовый метод может быть модифицирован для использования восстановленного фона в качестве дополнительных данных. Важно, чтобы данная модификация была выполнена с учетом следующих фактов:

- Результаты работы метода восстановления фона могут содержать неверно восстановленные фрагменты
- Восстановленный фон зачастую смещен на несколько пикселей из-за накопления ошибки восстановления

Для тех окон, которые не содержат восстановленного фона, базовый метод применяется без модификаций. Для каждого окна ω , содержащего хотя бы один пиксель восстановленного фона, дополним матрицу X_ω (см. уравнение (8)) цветами пикселей восстановленного фона. Использование соответствий окон фона окнам изображений вместо прямых соответствий между пикселями обеспечивает устойчивость метода к сдвигам фона до 3 пикселей. Для обеспечения устойчивости метода к ошибкам восстановления фона модифицируем уравнение (8), взвесив образцы пикселей в матрице X_ω на доверие к ним:

$$\beta = \arg \min_{\beta} \|C (X_\omega \beta - \alpha_\omega)\|^2 + \epsilon \|\beta\|^2, \quad (13)$$

где C — диагональная матрица, элементы главной диагонали которой определяют доверие к соответствующим строкам матрицы X_ω . В предлагаемом методе доверие ко всем образцам фона устанавливается равным 0,08 и 1 для остальных пикселей. Задача (13), аналогично предыдущему разделу, может быть переписана в виде:

$$\beta = (X_\omega^T C X_\omega + \epsilon I)^{-1} X_\omega^T C \alpha_\omega, \quad (14)$$

тогда (аналогично уравнению (10)):

$$w' = x'^T (X_\omega^T C X_\omega + \epsilon I)^{-1} X_\omega^T C. \quad (15)$$

Так как прозрачность пикселей восстановленного фона равна нулю, в уравнении (6) подставим нули вместо соответствующих им элементов вектора α_ω , эта операция эквивалентна удалению из вектора w' элементов, отвечающих пикселям восстановленного фона. Аналогично базовому методу получаем задачу вида (12), которая сводится к решению системы линейных уравнений. Таким образом, полученный метод расширяет каждое окно базового метода образцами пикселей восстановленного фона, прозрачность которых заведомо равна нулю, и учитывает их с низким доверием для обеспечения устойчивости к неточностям восстановленного фона.

3.4 Обеспечение стабильности результата во времени

Как уже упоминалось в начале раздела 3, предложенный метод выполняет итеративную фильтрацию последовательности карт прозрачности по схеме, описанной в уравнении (3). Здесь и далее записью v_k^n мы будем обозначать поле векторов движения от кадра n к кадру k . Рассмотрим, какие операции выполняются при фильтрации кадра t :

1. Вычисление полей векторов движения v_{t-1}^t , v_{t+1}^t , v_t^{t-1} и v_t^{t+1} с помощью алгоритма, описанного в [24]
2. Вычисление карт доверия к полям векторов движения K_{t-1}^t и K_{t+1}^t
3. Применение базового метода матирования, учитывающего восстановленный фон, с дополнительными требованиями стабильности результата во времени

Для вычисления доверия к векторам движения, вычисленным на первом шаге, используются две вспомогательные величины: **FBC** и E . **FBC** — несоответствие между полями векторов движения, вычисленных в прямом и обратном направлениях. Значение **FBC** в точке с координатой x может быть вычислено по формуле:

$$\mathbf{FBC}_k^n(x) = \|v_k^n(x) - v_n^k(x + v_k^n(x))\|_2^2. \quad (16)$$

$E_k^n(x)$ — L_1 расстояние между блоком 8×8 пикселей с центром в точке x кадра n исходной видеопоследовательности и аналогичным блоком кадра k с центром в точке $x + v_k^n(x)$.

С использованием введенных величин доверие к векторам движения вычисляется по формуле:

$$K_{t-1}^t = \exp\left(-\frac{\mathbf{FBC}_{t-1}^t}{\sigma_1} - \frac{E_{t-1}^t}{\sigma_2}\right), \quad (17)$$

где σ_1, σ_2 — константы, выбранные с использованием двух видеопоследовательностей с известной эталонной картой прозрачности из набора [3]. Для этого были предвычислены вектора движения для этих последовательностей. Затем исключены из рассмотрения вектора, принадлежащие областям с малой дисперсией значений эталонной прозрачности. Для оценки качества вектора $v_k^n(x)$ было вычислено пиковое отношение сигнала к шуму $P_k^n(x)$ между блоком 8×8 пикселей с центром в точке x эталонной карты прозрачности кадра n и аналогичным блоком кадра k с центром в точке $x + v_k^n(x)$. Путем визуального контроля было установлено, что значения менее 12 децибел соответствуют ошибочно найденным векторам. σ_1, σ_2 были выбраны как:

$$\langle \sigma_1, \sigma_2 \rangle = \arg \min_{\sigma_1, \sigma_2} \sum_{P_k^n(x)} \exp\left(-\frac{\mathbf{FBC}_{t-1}^t}{\sigma_1} - \frac{E_{t-1}^t}{\sigma_2}\right) (12 - P_k^n(x)). \quad (18)$$

Для решения последней задачи был использован метод внутренней точки. Найденные значения: $\sigma_1 = 3, 9$; $\sigma_2 = 29$.

Рассмотрим шаг матирования. Под операцией компенсации кадра k к кадру n мы будем понимать отображение R :

$$I' = R(I, v_k^n), \quad (19)$$

такое, что

$$I'(x + v_k^n(x)) = I(x). \quad (20)$$

Компенсируем предыдущий кадр, следующий кадр и их карты прозрачности (вычисленные на предыдущей итерации) к текущему кадру, будем обозначать их I_B, I_F, α_B и α_F соответственно. Аналогично описанному в разделе 3.3 способу расширим для каждого окна ω матрицу X_ω пикселями из I_B и I_F . Вычислим коэффициенты доверия к этим образцам по формуле (17) и домножим их на 0,2 (экспериментально выбранное значение). Используем значения прозрачности в окне ω карт α_B и α_F в качестве значений прозрачности добавленных пикселей. Добавим в уравнение (12) новые слагаемые:

$$\psi [(\alpha - \alpha_B)^T \text{Diag}(K_{t-1}^t)(\alpha - \alpha_B) + (\alpha - \alpha_F)^T \text{Diag}(K_{t+1}^t)(\alpha - \alpha_F)], \quad (21)$$

где $\text{Diag}(K_{t-1}^t)$ — диагональная матрица, образованная значениями карты доверия K_{t-1}^t ; $\psi = 0, 1$.

Предложенный метод выполняет две итерации сглаживания для вычисления итоговой карты прозрачности.

| Метод | Ср. место | Тестовая последовательность | | | | | | | | | |
|-----------------------------|-----------|-----------------------------|----------------------|---------------------|----------------------|---------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| | | №1 | №2 | №3 | №4 | №5 | №6 | №7 | №8 | №9 | №10 |
| Предложенный | 1,9 | 29,39 ¹ | 27,79 ² | 8,83 ² | 122,22 ⁵ | 13,96 ¹ | 21,01 ¹ | 31,83 ³ | 27,66 ¹ | 49,26 ¹ | 59,67 ² |
| Learning Based [1] | 3,1 | 29,98 ² | 33,08 ⁵ | 10,84 ⁶ | 134,82 ⁷ | 15,23 ² | 23,35 ² | 31,46 ² | 30,20 ² | 58,38 ² | 56,60 ¹ |
| Closed Form [10] | 4,3 | 31,34 ³ | 31,80 ⁴ | 10,96 ⁷ | 149,48 ⁹ | 15,86 ³ | 23,67 ³ | 30,15 ¹ | 30,60 ³ | 69,94 ⁴ | 82,13 ⁶ |
| Comprehensive Sampling [26] | 4,3 | 34,48 ⁴ | 30,18 ³ | 12,10 ⁸ | 79,60 ¹ | 16,17 ⁴ | 25,02 ⁴ | 35,25 ⁵ | 41,05 ⁵ | 77,72 ⁵ | 70,18 ⁴ |
| Robust Matting [12] | 4,6 | 37,54 ⁶ | 26,15 ¹ | 9,61 ³ | 88,15 ³ | 18,52 ⁵ | 33,31 ⁶ | 34,15 ⁴ | 46,91 ⁷ | 78,56 ⁶ | 80,41 ⁵ |
| Shared Matting [21] | 4,9 | 36,79 ⁵ | 46,72 ⁷ | 10,30 ⁴ | 87,64 ² | 24,41 ⁶ | 35,18 ⁷ | 35,26 ⁶ | 45,23 ⁶ | 68,52 ³ | 67,17 ³ |
| Refine Edge [13] | 5,8 | 37,78 ⁷ | 38,48 ⁶ | 8,64 ¹ | 129,56 ⁶ | 27,47 ⁷ | 31,36 ⁵ | 36,65 ⁷ | 38,75 ⁴ | 102,02 ⁸ | 94,50 ⁷ |
| Bayesian Matting [5] | 7,9 | 58,76 ⁹ | 51,66 ⁸ | 10,79 ⁵ | 144,53 ⁸ | 30,36 ⁸ | 53,28 ⁸ | 43,70 ⁸ | 70,50 ⁹ | 96,69 ⁷ | 140,12 ⁹ |
| KNN Matting [23] | 8,8 | 67,83 ¹⁰ | 92,22 ⁹ | 64,73 ¹¹ | 112,46 ⁴ | 32,96 ⁹ | 61,50 ⁹ | 50,87 ⁹ | 79,20 ¹⁰ | 113,41 ⁹ | 123,63 ⁸ |
| Nonlocal matting [27] | 9,7 | 54,21 ⁸ | 92,37 ¹⁰ | 33,86 ¹⁰ | 243,47 ¹¹ | 45,13 ¹⁰ | 64,05 ¹⁰ | 61,75 ¹⁰ | 67,30 ⁸ | 135,08 ¹⁰ | 171,79 ¹⁰ |
| Spectral Matting [28] | 10,7 | 86,32 ¹¹ | 105,24 ¹¹ | 13,15 ⁹ | 210,04 ¹⁰ | 48,76 ¹¹ | 65,72 ¹¹ | 68,83 ¹¹ | 109,44 ¹¹ | 157,18 ¹¹ | 182,82 ¹¹ |

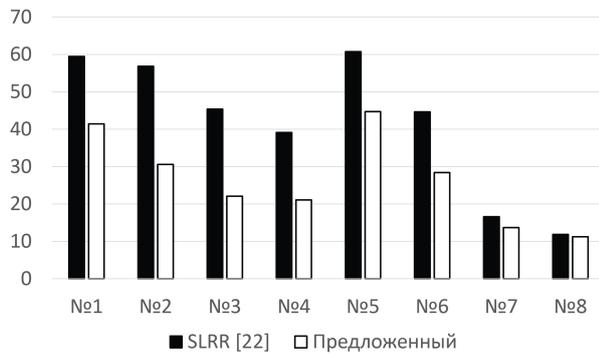
Таблица 1: Результаты сравнения предложенного алгоритма с 10 аналогами на 10 тестовых видеопоследовательностях из набора, описанного в [3]. Для каждой видеопоследовательности были вычислены L_2 расстояния между результатами работы алгоритмов и эталоном (приведены в столбцах 3-12). Затем методы были упорядочены по близости результата к эталону (надстрочные числа в столбцах 3-12). Для каждого метода было вычислено среднее место во всех тестах (второй столбец). Меньшие значения соответствуют более высоким местам в рейтинге.

| Метод | Ср. место | Тестовая последовательность | | | | | | | | | |
|-----------------------------|-----------|-----------------------------|---------------------|--------------------|---------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|
| | | №1 | №2 | №3 | №4 | №5 | №6 | №7 | №8 | №9 | №10 |
| Предложенный | 1,3 | 0,44 ¹ | 0,941 ² | 0,07 ² | 2,65 ¹ | 0,15 ¹ | 0,25 ¹ | 0,66 ² | 0,51 ¹ | 1,60 ¹ | 2,20 ¹ |
| Closed Form [10] | 3,1 | 0,49 ² | 1,35 ⁴ | 0,09 ⁶ | 2,88 ² | 0,19 ³ | 0,35 ³ | 0,61 ¹ | 0,61 ² | 2,75 ³ | 3,38 ⁵ |
| Learning Based [1] | 3,2 | 0,50 ³ | 1,42 ⁵ | 0,09 ⁷ | 2,88 ³ | 0,18 ² | 0,34 ² | 0,68 ³ | 0,64 ³ | 2,37 ² | 2,43 ² |
| Refine Edge [29] | 4,6 | 0,55 ⁴ | 1,48 ⁶ | 0,06 ¹ | 3,99 ⁶ | 0,34 ⁶ | 0,54 ⁵ | 0,70 ⁴ | 0,81 ⁴ | 3,80 ⁶ | 3,16 ⁴ |
| Comprehensive Sampling [26] | 5,1 | 0,74 ⁵ | 1,06 ³ | 0,19 ⁹ | 3,15 ⁴ | 0,22 ⁴ | 0,39 ⁴ | 0,86 ⁶ | 1,24 ⁵ | 3,65 ⁵ | 3,40 ⁶ |
| Robust Matting [12] | 5,6 | 0,89 ⁷ | 0,936 ¹ | 0,07 ³ | 4,82 ⁸ | 0,25 ⁵ | 0,88 ⁶ | 0,77 ⁵ | 1,58 ⁷ | 3,89 ⁷ | 3,92 ⁷ |
| Shared Matting [21] | 5,9 | 0,79 ⁶ | 2,73 ⁷ | 0,09 ⁵ | 4,50 ⁷ | 0,35 ⁷ | 1,00 ⁷ | 0,91 ⁷ | 1,43 ⁶ | 2,94 ⁴ | 2,94 ³ |
| KNN Matting [23] | 8,4 | 2,28 ¹⁰ | 6,40 ⁹ | 1,22 ¹¹ | 3,47 ⁵ | 0,61 ⁸ | 1,02 ⁸ | 1,73 ⁹ | 1,94 ⁸ | 6,42 ⁸ | 7,73 ⁸ |
| Bayesian Matting [5] | 9,1 | 1,85 ⁹ | 3,58 ⁸ | 0,12 ⁸ | 13,33 ⁹ | 0,74 ⁹ | 2,31 ¹¹ | 1,47 ⁸ | 4,51 ¹⁰ | 8,01 ⁹ | 16,23 ¹⁰ |
| Spectral Matting [28] | 9,8 | 7,03 ¹¹ | 13,25 ¹¹ | 0,08 ⁴ | 15,36 ¹⁰ | 2,09 ¹¹ | 1,84 ⁹ | 2,45 ¹⁰ | 6,59 ¹¹ | 8,71 ¹⁰ | 19,97 ¹¹ |
| Nonlocal matting [27] | 9,9 | 1,56 ⁸ | 8,17 ¹⁰ | 0,70 ¹⁰ | 20,97 ¹¹ | 1,11 ¹⁰ | 1,88 ¹⁰ | 2,51 ¹¹ | 2,35 ⁹ | 11,17 ¹¹ | 13,37 ⁹ |

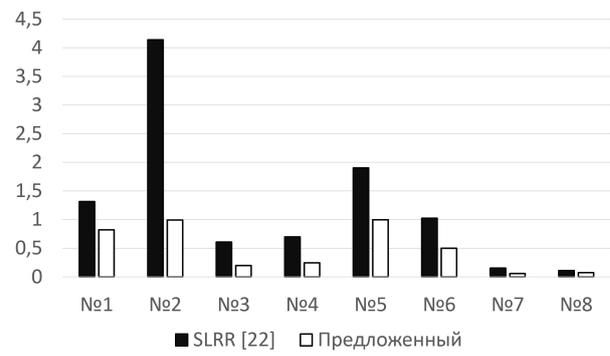
Таблица 2: Результаты сравнения предложенного алгоритма с 10 аналогами на 10 тестовых видеопоследовательностях из набора, описанного в [3]. Для каждой видеопоследовательности были вычислены значения квазиметрики **meSSDdt** (в работе [25] показана высокая корреляция данного способа оценки с результатами экспертной оценки) между результатами работы алгоритмов и эталоном (приведены в столбцах 3-12). Затем методы были упорядочены по близости результата к эталону (надстрочные числа в столбцах 3-12). Для каждого метода было вычислено среднее место во всех тестах (второй столбец). Меньшие значения соответствуют более высоким местам в рейтинге.

4 Экспериментальная оценка

Описанный метод был сравнен с другими методами матирования изображений и видео на 10 тестовых видео с известными эталонными картами прозрачности из набора, предложенного в [3]. Критериями оценки выступали: L_2 расстояние до эталона и **meSSDdt** (метод сравнения, предложенный в работе [25], имеющий высокую корреляцию с результатами экспертной оценки). Результаты сравнения предложенного метода с 10 ранее существовавшими приведены в таблицах 1 и 2. Для каждой тестовой последовательности методы были упорядочены согласно критерию сравнения, а затем было вычислено среднее место метода во всех тестах. Как видно из таблиц, предложенный метод превзошел аналоги по данному показателю. Рассмотрим также сравнение предложенного метода с алгоритмом матирования видео, предложенным в 2015 году [22]. Сравнение было выполнено на устаревшей версии набора тестовых видео (к сожалению, авторы метода не предоставили результатов тестирования своего метода на обновленном наборе). Результаты представлены на рисунке 1, согласно которому предложенный метод превзошел аналог.



(а) L_2 расстояние до эталона



(б) **meSSDdt** (предложена в [25])

Рис. 1: Сравнение предложенного алгоритма с алгоритмом, описанным в [22], на 8 тестовых видео последовательностях с использованием L_2 расстояния до эталона (а) и **meSSDdt** (б). Меньшие значения соответствуют лучшим результатам согласно выбранному способу сравнения.

5 Заключение

В работе был предложен алгоритм матирования видеопоследовательностей, использующий восстановленный фон в качестве дополнительных данных. К основным особенностям можно отнести:

- Модификацию базового алгоритма матирования, позволившую использовать фон в качестве дополнительных данных
- Итеративный подход к пространственно-временному сглаживанию карт прозрачности, позволивший избежать накопления ошибки и вычислять несколько кадров параллельно
- Способ вычисления доверия к векторам движения, выбранный путем анализа двух видео с эталонной картой прозрачности

Преимущества предложенного метода были продемонстрированы путем объективного сравнения результатов его работы с 11 аналогами.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 15-01-08632 а.

Список литературы

- [1] Zheng Yuanjie, Kambhamettu C. Learning based digital matting // International Conference on Computer Vision (ICCV). — 2009. — P. 889–896.
- [2] Зачесов А. А., Ерофеев М. В., Ватолин Д. С. Использование карт глубины при восстановлении фона в видеопоследовательностях // Новые информационные технологии в автоматизированных системах: материалы научно-практического семинара. — М.: МИЭМ НИУ ВШЭ, 2015.
- [3] Методика объективного сравнения алгоритмов матирования видео / М. В. Ерофеев, Ю. А. Гитман, Д. С. Ватолин, А. А. Федоров // Цифровая обработка сигналов. — 2015. — № 3. — С. 53–59.

- [4] Video matting of complex scenes / Yung-Yu Chuang, Aseem Agarwala, Brian Curless et al. // ACM Transactions on Graphics (TOG). — 2002. — Vol. 21, no. 3. — P. 243–248.
- [5] A bayesian approach to digital matting / Yung-Yu Chuang, Brian Curless, David H. Salesin, Richard Szeliski // Computer Vision Pattern Recognition (CVPR). — Vol. 2. — 2001. — P. II-264–II-271.
- [6] Apostoloff Nicholas, Fitzgibbon Andrew. Bayesian video matting using learnt image priors // Computer Vision Pattern Recognition (CVPR). — Vol. 1. — 2004. — P. I-407–I-414.
- [7] Corrigan David, Robinson S, Kokaram A. Video matting using motion extended grabcut // European Conference on Visual Media Production (CVMP). — 2008. — P. 3–3(1).
- [8] Rother Carsten, Kolmogorov Vladimir, Blake Andrew. Grabcut: Interactive foreground extraction using iterated graph cuts // ACM transactions on graphics (TOG) / ACM. — Vol. 23. — 2004. — P. 309–314.
- [9] Video snapcut: Robust video object cutout using localized classifiers / Xue Bai, Jue Wang, David Simons, Guillermo Sapiro // ACM Transactions on Graphics (TOG). — 2009. — Vol. 28, no. 3. — P. 70:1–70:11.
- [10] Levin Anat, Lischinski Dani, Weiss Yair. A closed-form solution to natural image matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). — 2008. — Vol. 30, no. 2. — P. 228–242.
- [11] Lee Sun-Young, Yoon Jong-Chul, Lee In-Kwon. Temporally coherent video matting // Graphical Models. — 2010. — Vol. 72, no. 3. — P. 25–33.
- [12] Wang Jue, Cohen Michael F. Optimized color sampling for robust matting // Computer Vision Pattern Recognition (CVPR). — 2007. — P. 1–8.
- [13] Bai Xue, Wang Jue, Simons David. Towards temporally-coherent video matting // International Conference on Computer Vision (ICCV). — 2011. — P. 63–74.
- [14] Linear time euclidean distance transform algorithms / Heinz Breu, Joseph Gil, David Kirkpatrick, Michael Werman // Pattern Analysis and Machine Intelligence, IEEE Transactions on. — 1995. — Vol. 17, no. 5. — P. 529–533.
- [15] Video matting via opacity propagation / Zhen Tang, Zhenjiang Miao, Yanli Wan, Dianyong Zhang // The Visual Computer. — 2012. — Vol. 28, no. 1. — P. 47–61.
- [16] Spatio-temporally coherent interactive video object segmentation via efficient filtering / Nicole Brosch, Asmaa Hosni, Christoph Rhemann, Margrit Gelautz // Pattern Recognition. — Vol. 7476. — 2012. — P. 418–427.
- [17] He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering // European Conference on Computer Vision (ECCV). — 2010. — P. 1–14.
- [18] Sindeev Mikhail, Konushin Anton, Rother Carsten. Alpha-flow for video matting // Computer Vision–ACCV 2012. — Springer, 2012. — P. 438–452.
- [19] Choi Inchang, Lee Minhaeng, Tai Yu-Wing. Video matting using multi-frame nonlocal matting laplacian // European Conference on Computer Vision (ECCV). — 2012. — P. 540–553.

- [20] Temporally coherent and spatially accurate video matting / Ehsan Shahrian, Brian Price, Scott Cohen, Deepu Rajan // Computer Graphics Forum. — 2014. — Vol. 33, no. 2. — P. 381–390.
- [21] Gastal Eduardo S.L., Oliveira Manuel M. Shared sampling for real-time alpha matting // Computer Graphics Forum. — 2010. — Vol. 29, no. 2. — P. 575–584.
- [22] Video matting via sparse and low-rank representation / Dongqing Zou, Xiaowu Chen, Guangying Cao, Xiaogang Wang // Proceedings of the IEEE International Conference on Computer Vision. — 2015. — P. 1564–1572.
- [23] Chen Qifeng, Li Dingzeyu, Tang Chi-Keung. KNN matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). — 2013. — Vol. 35, no. 9. — P. 2175–2188.
- [24] Fast video super-resolution via classification / Karen Simonyan, Sergey Grishin, Dmitriy Vatolin, Dmitriy Popov // International Conference on Image Processing (ICIP). — 2008. — P. 349–352.
- [25] Perceptually motivated benchmark for video matting / Mikhail Erofeev, Yury Gitman, Dmitriy Vatolin et al. // Proceedings of the British Machine Vision Conference (BMVC). — BMVA Press, 2015. — September. — P. 99.1–99.12. — URL: <https://dx.doi.org/10.5244/C.29.99>.
- [26] Improving image matting using comprehensive sampling sets / E. Shahrian, D. Rajan, B. Price, S. Cohen // Computer Vision Pattern Recognition (CVPR). — 2013. — P. 636–643.
- [27] Lee Philip, Wu Ying. Nonlocal matting // Computer Vision Pattern Recognition (CVPR). — 2011. — P. 2193–2200.
- [28] Levin A., Rav Acha A., Lischinski D. Spectral matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). — 2008. — Vol. 30, no. 10. — P. 1699–1712.
- [29] <http://www.adobe.com/en/products/aftereffects.html>.