

Алгоритмы сегментации кадров и сжатия видеоинформации для систем видеонаблюдения

Л.Д. Вилесов, С.А. Кузьмин

Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП)

Аннотация: Статья посвящена решению задачи сжатия видеоинформации для системы панорамного видеонаблюдения. Актуальность задачи связана как с ограничениями существующих решений, так и с особенностями видеопотоков, которые во многих кодеках не учитываются. Новизна работы связана с тем, что роль ключевого кадра выполняет не один из кадров входного видео, а изображение оценки фона. В данной работе используются идеи всех четырех основных направлений в сжатии: сегментация (разделение на динамическую составляющую и статический фон), JPEG-сжатие, уменьшенная частота передачи опорного кадра (оценки фона), возможность изменения разрешения кадра, синтез выходного изображения. Для решения задачи сжатия производится разделение сцены на отдельные семантически значимые элементы (объекты, фон), передаваемые по каналу связи с разной степенью подробности в зависимости от семантической значимости и сопровождаемые метаданными. В ходе эксперимента установлено, что значение коэффициента сжатия видеопотока $K=135$ раз при пиковом отношении сигнал-шум $PSNR=32$ дБ, что позволяет передавать панораму разрешением 3072×720 , сшитую из кадров от трех камер по каналу с пропускной способностью в 10-11 Мбит/с.

Ключевые слова: сжатие видеоинформации, сегментация, обработка изображений, видеонаблюдение, метаданные.

Введение

Большой поток данных от камер наблюдения создает сильную нагрузку на каналы связи. Для несжатых видеоданных характерен большой объем данных, что также существенно усложняет и увеличивает стоимость их хранения. Например, для хранения несжатых видеоданных, накопленных за сутки в панорамной системе видеонаблюдения с формируемым изображением разрешением 3072×720 требуются больше десятка носителей данных с суммарным объемом 13348,39 ГБ и ценой 734 доллара [15]. При этом существует тенденция увеличения разрешения видеокамер в системах видеонаблюдения.

Поток данных надо согласовывать с пропускной способностью каналов связи. Таким образом, возникает задача сжатия видеоинформации, особенно актуальная в случае систем видеонаблюдения, в которых осуществляется сшивка панорам. Данная статья посвящена описанию алгоритма сжатия видеоинформации для панорамных систем видеонаблюдения.

Актуальность разработки алгоритма сжатия

Актуальность задачи связана как с ограничениями существующих решений, так и с особенностями видеопоследовательностей, которые во многих кодеках не учитываются.

Основные предпосылки для разработки нового способа сжатия:

- 1) нестандартное разрешение панорам, которое выше, чем поддерживает стандарт H.264;
 - 2) движущиеся объекты занимают малую часть кадра, поэтому частая передача опорных кадров будет лишним действием для 80% пикселей;
 - 3) значительная часть кодеков (серия стандартов MPEG) работает с группами кадров, что приводит к задержкам при кодировании в реальном времени (чем больше длина группы кадров, тем больше задержка между моментом получения первого и последнего кадров в группе) и при декодировании (для доступа к произвольному кадру надо декодировать несколько кадров).
 - 4) опорные кадры для панорамных изображений требуют в несколько раз больше бит, чем предсказанные. Это приводит к скачкам объема передаваемых данных, что усложняет реализацию буфера кодера и буфера декодера. Для борьбы с этой проблемой был предложен способ нарезки опорного кадра на вертикальные полоски (слайсы) и передачи полосок вместе с предсказанными кадрами [8].
 - 5) многие алгоритмы сжатия не обеспечивают метаданные для объектов (Motion JPEG, Dirac, VP3-Theora, VP8-WebM, RealVideo);
-

- 6) применяемые в сжатии (серия стандартов MPEG) алгоритмы сегментации кадров видеопотоков имеют низкую точность в условиях приближения/удаления объектов;
- 7) обычно алгоритмы сжатия нацелены на сохранение или ухудшение параметров, в то время как можно сжимать с улучшением визуального качества (подавлением теней, коррекцией геометрических искажений, компенсацией дрожания камеры) и повышением характеристик видеоряда (частоты кадров, отношения сигнал-шум, размера кадра).

Подходы к сжатию

Сжатие видеoinформации возможно из-за наличия избыточности – статистической, психофизиологической, структурной. Направления в сжатии видеорядов [9]:

1. использование сегментации с последующей обработкой. Пример: методы на основе векторов движения (семейство MPEG). Последующая обработка включает в себя дифференциальную импульсно-кодую модуляцию (ДИКМ);
2. применение сжатия с потерями, основанного на более грубом представлении яркости пикселей кадра. Пример: алгоритм JPEG - преобразование в частотную область, обработка в частотной области и обратное преобразование в пространственную область;
3. уменьшение параметров видеопотока (размера изображения, частоты кадров, длительности). Примеры: чересстрочная развертка, малокадровое ТВ, запись по тревоге.

Отдельного внимания заслуживает четвертое направление, связанное с исключением передачи части кадра и синтезом изображения. Примеры: цветовая субдискретизация, «нарезка»[10], «разделение на контексты, удаление мелких деталей, у которых контраст ниже порогового контраста, генерация фактуры» [11, 12].

Разработанный алгоритм сжатия видеoinформации

В данной работе используются идеи всех четырех основных направлений: сегментация (разделение на динамическую составляющую и статический фон), JPEG-сжатие, уменьшенная частота передачи опорного кадра (оценки фона), возможность изменения разрешения кадра, синтез выходного изображения.

Для решения задачи сжатия разработан алгоритм видеосжатия [2] на основе сегментации кадра и передачи областей с объектами на приемную сторону, где они вставляются в изображение оценки фона. Задача реализации высокоточной сегментации является очень сложной, поскольку в условиях яркостной априорной неопределенности при пороговой обработке возникают ошибки ложной тревоги и пропуска. Это связано с многомодальным характером гистограмм кадров видеопотока, состоящих из смеси распределений, соответствующих теням, темному фону, темным объектам, светлому фону, светлым объектам.

Алгоритм сегментации

В условиях, когда яркость фона неоднородна в пространстве и шум имеет негауссово распределение, для выделения точечных объектов оптимальным является применение выбеливающего фильтра с последующим применением согласованного фильтра. При обработке изображений такой фильтр фактически реализуется вычислением контурного препарата и применением решающего правила (операции бинаризации) [1 с. 50-80, 13].

В условиях, когда объекты являются протяженными, такой подход малоэффективен, т.к. большие части однородно окрашенных объектов теряются при вычислении контурного препарата. Поэтому необходимо дополнить информацию о контурах информацией обо всех движущихся участках кадра. Основные этапы предложенного подхода: вычитание оценки фона (детектор «движущиеся области»), устранение высоких частот оценки

фона (детектор «движущиеся контуры»), объединение этих алгоритмов сегментации, коррекция выходного изображения, в том числе путем устранения детектированных вместе с объектами теней из бинарного изображения. Алгоритм сегментации представлен на рис. 1. Блок «выделение объектов» (рис. 2): семейство алгоритмов сегментации с точностью, регулируемой от пониженной до субпиксельной [3].

На его вход подается панорамное изображение. Для сшивки панорамы используется метод, основанный на векторах смещения характерных точек [1 с.203-218, 16].

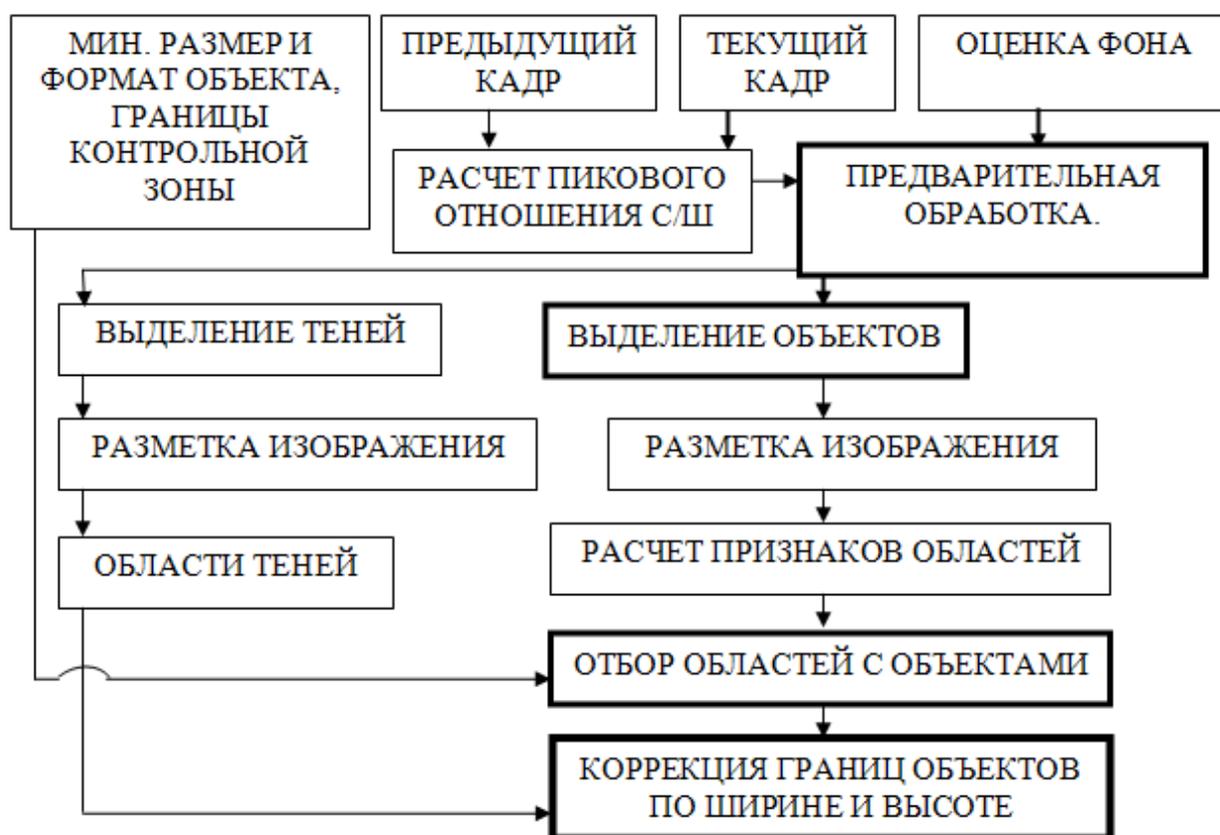


Рис. 1. Последовательность операций при сегментации текущего кадра.

Для попиксельной оценки фона используется переключение между двумя алгоритмами: фильтр Калмана используется при низком отношении сигнал/шум, медиана – при высоком отношении сигнал/шум. В [14] описан

алгоритм априорной оценки пригодности алгоритмов оценивания фоновой составляющей. Различие во внешнем виде и значениях СКО разностных изображений рассматривается как разница в условиях наблюдения (отношении сигнал-шум) некоего идеального разностного изображения. Для двух сравниваемых алгоритмов вычисляются СКО разностных изображений (без взятия модуля) и вычисляется показатель, показывающий насколько условия наблюдения для одного алгоритма лучше, чем для второго, в дБ:

$$\Delta\psi = 20 \log_{10} \frac{MSE1}{MSE(n(t))} - 20 \log_{10} \frac{MSE2}{MSE(n(t))} = 20 \log_{10} \frac{MSE1}{MSE2}.$$

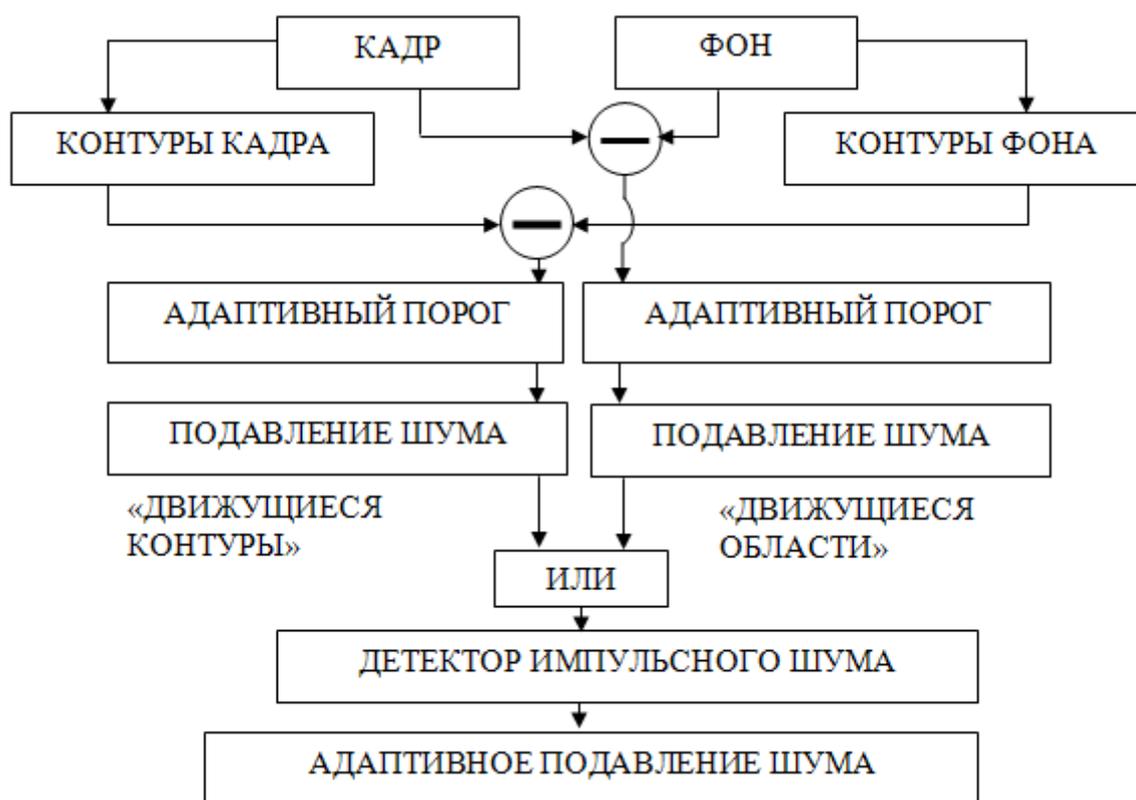


Рис. 2. Блок-схема блока «выделение объектов» для номинальной точности

Для коррекции изображений после объединения результатов работы детекторов интересных пикселей используется детектор импульсного биполярного шума и адаптивные фильтры на основе порядковой статистики [5].

При добавлении априорной информации о площади объектов происходит переход из класса задач с полной априорной неопределенностью в класс задач с неполной априорной информацией. В результате отбора уменьшается количество областей, которые могут быть приняты за объекты.

Семантическое сжатие видеoinформации

При реализации алгоритма видеосжатия учитывались четыре актуальных направления в видеосжатии. Первое направление связано с созданием систем синтеза изображения лучшего качества, чем исходное. Это осуществляется устранением дефектов, например геометрических искажений [1 с.232-245] или размытости [7]. В [7] для таких изображений сформулирован термин *superimage* – сверхизображение. В данной статье под синтезом сверхизображения также подразумеваются: 1) возможность удаления теней, движущихся рядом с объектами [4]; 2) возможность увеличения изображения для дальнейшего субпиксельного анализа [3]. Второе направление заключается в том, что из-за перехода к системам панорамного видеонаблюдения приходится сжимать не исходные кадры, а преобразованные в результате сшивки [6]. Это усложняет сжатие для известных кодеков, т.к. получающееся разрешение изображения является нестандартным. Третье направление заключается в том, что стремятся комбинировать идеи из разных алгоритмов сжатия. В работе [17] это направление названо *supercompression* – сверхсжатие. Четвертое направление связано с возможностью передачи видеопотока с разрешением, которое допустимо для данного канала связи или требуется потребителю.

Блок-схема алгоритма видеосжатия представлена на рис. 3. На основе фильтрации набора кадров в направлении оси времени получается изображение оценки фона.

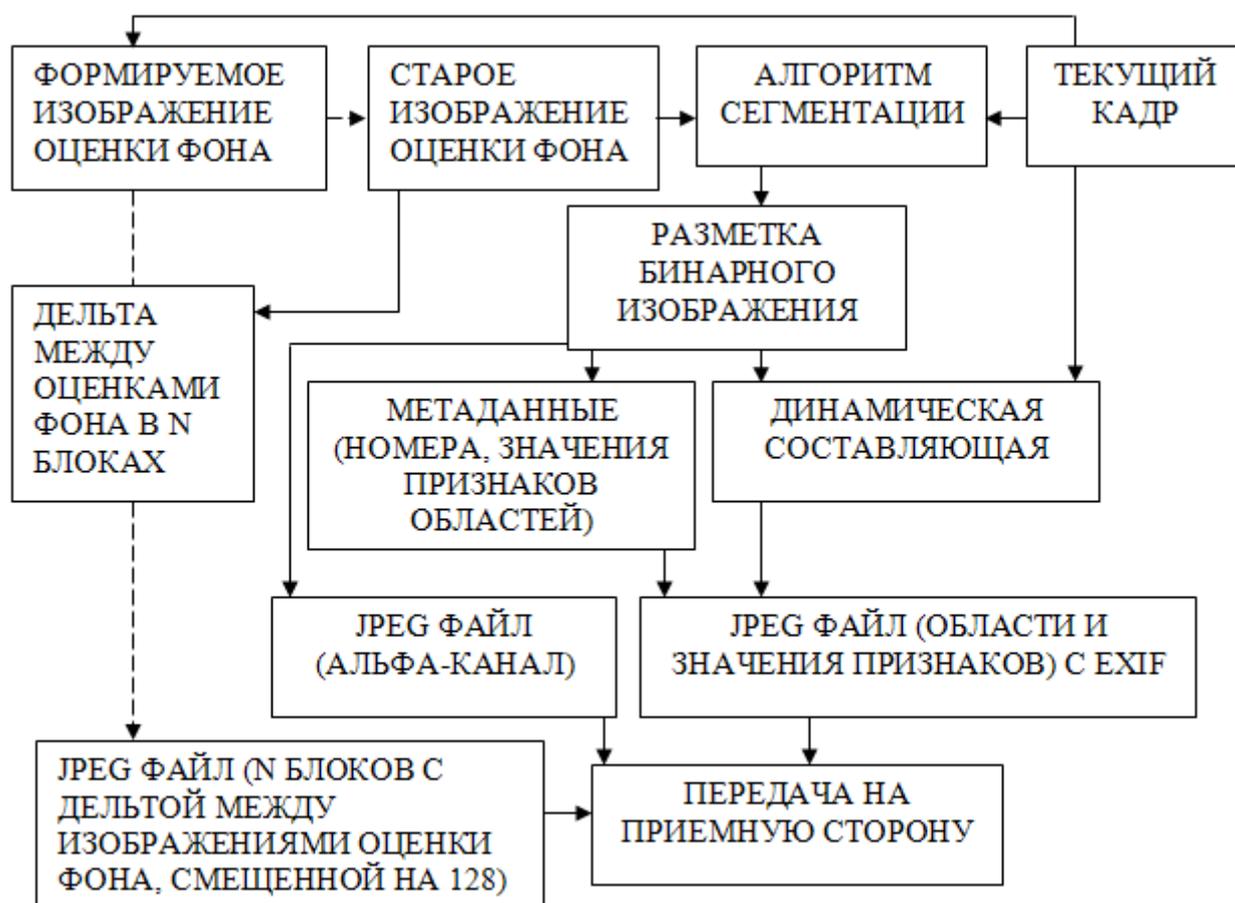


Рис. 3. Блок-схема алгоритма кодирования одного кадра

Если проводить аналогию с классическими алгоритмами сжатия MPEG версий 1 и 2, то изображение оценки фона является ключевым кадром. Но если в MPEG версий 1 и 2 ключевые кадры надо полностью обновлять раз в 12 кадров, то в данном случае оценку фона можно обновлять гораздо реже. В MPEG версий 1 и 2 передача I-кадров видеопанорамы - проблема, потому что они требуют намного большего количества битов, чем P (предсказанные) и B (двунаправлено предсказанные) кадры. Фактически это приводит к огромному увеличению количества переданных битов для I-кадров, которое может быть измерено с точки зрения пик-фактора (отношения пикового значения к среднему). Способ уменьшить эту проблему был описан в [8] разработчиком кодека x264. Ключевой кадр разделяется на несколько вертикальных столбцов, которые передаются вместе с предсказанными

кадрами. Результат этого приёма - намного более низкий пик-фактор. В алгоритме, описываемом в данной статье, передаваемое изображение делится не на столбцы, а на блоки размером 8x8 пикселей.

Передаются не значения интенсивности, а разницы (положительные или отрицательные небольшие значения) между интенсивностью внутри блока текущей оценки фона и уже переданной оценки фона, сдвинутые на 128 уровней. Такие сдвинутые разностные сигналы больше похожи на базисные функции ДКП, чем исходные сигналы. Скорость обновления оценки фона определяется количеством блоков N оценки фона, передаваемых с каждой порцией данных о движущихся объектах и их параметрах. Каждая порция данных включает в себя альфа-канал и динамическую составляющую (движущиеся объекты), формирование которых описано далее.

В изображение фоновой составляющей, переданное на приемную сторону, можно вставлять движущиеся объекты (рис.4).

Информацию о них необходимо передавать с частотой видеосъемки. На основании изображения и результата работы блока сегментации создается изображение динамической составляющей (ДС), в котором есть участки с яркостью исходного изображения в пикселях, где в ходе сегментации определен класс «объект», и черный фон в остальных пикселях. ДС и изображение результата сегментации (альфа-канала) сжимаются в файлы формата JPEG. В тэгах JPEG файла передаются метаданные (позиции объектов). На приемной стороне объекты вставляются в изображение оценки фона в соответствии с координатами, записанными в метаданных.

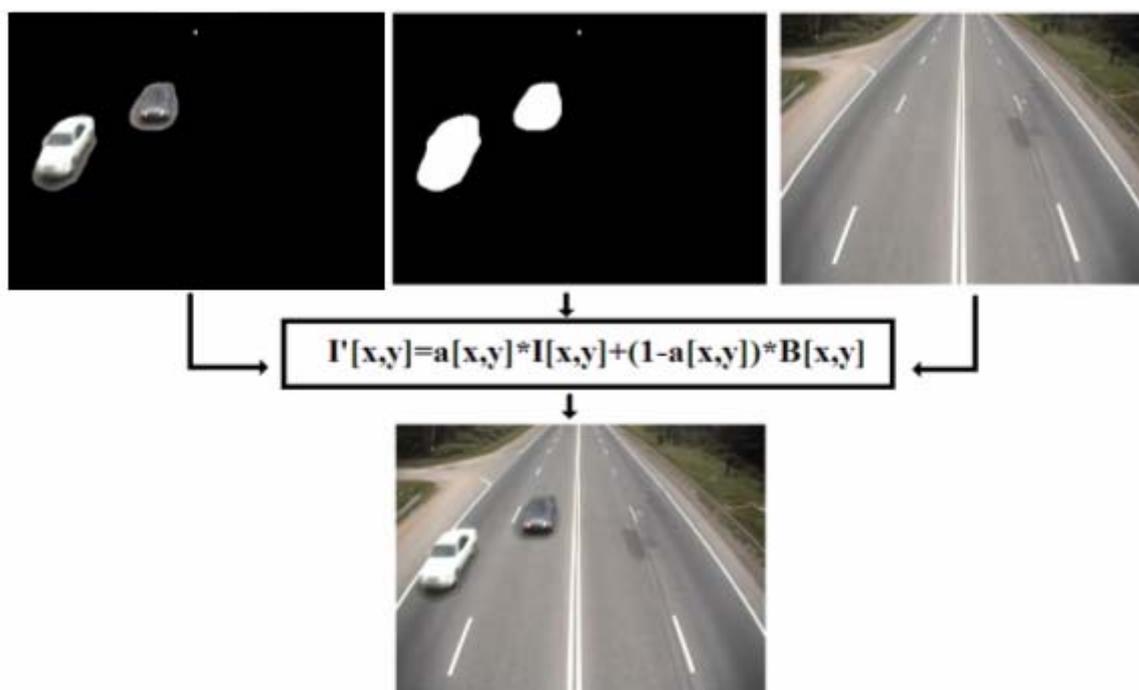


Рис. 4. Принцип синтеза изображения при декодировании – взвешенное суммирование динамической составляющей и оценки фона, при котором веса задаются бинарным альфа-каналом.

Экспериментальные характеристики алгоритма сжатия

Основными показателями алгоритмов сжатия являются коэффициент сжатия K и качество сигнала. В данной работе для оценки качества использовался показатель PSNR.

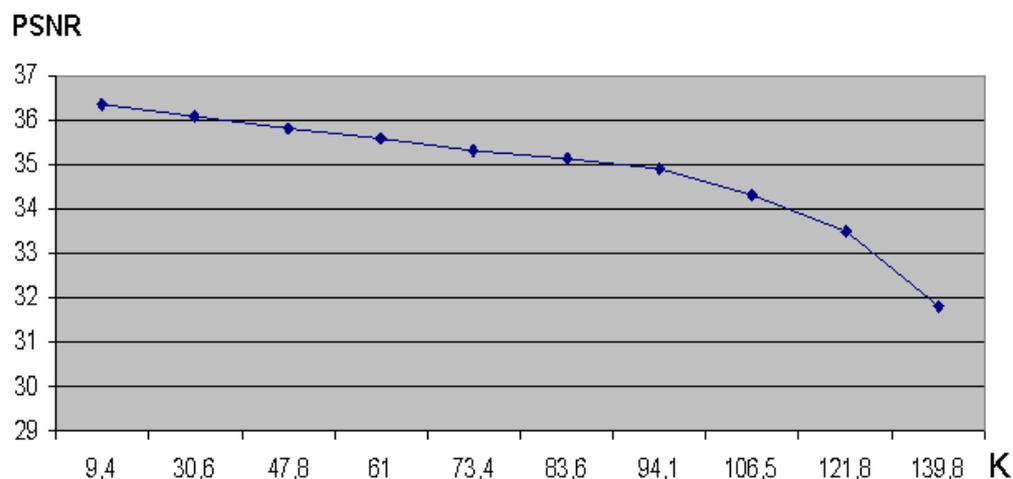


Рис. 5. Зависимость PSNR от коэффициента сжатия.

Значение коэффициента сжатия видеопотока достигает 135 раз (скорость ввода информации в канал 9,83 Мбит/с) при PSNR 32 дБ, что позволяет передавать панораму разрешением 3072x720, сшитую из кадров от трех камер, по каналу со стандартной пропускной способностью в 10-11 Мбит/с.

Заключение

В данной работе отражена работа по созданию алгоритма сжатия видеоинформации с синтезом выходного изображения. Данным научным направлением начали заниматься в СПбГУАП более 10 лет назад [10-11], а данным конкретным алгоритмом – с конца 2008 года. Новизна работы связана с тем, что роль ключевого кадра выполняет не один из кадров входного видео, а изображение оценки фона. Комбинирование идей основных направлений в сжатии позволило достичь требуемых показателей, решив актуальную практическую и теоретическую задачу. В результате уменьшается стоимость эксплуатации (обслуживания и хранения данных) панорамной системы видеонаблюдения, разрабатываемой в СПбГУАП.

Изменения по сравнению со статьей [2]:

1) сегментация с регулируемой пространственной точностью [3], т.е. можно менять разрешение передаваемой видеопоследовательности в зависимости от разрешения и количества видеокамер, пропускной способности канала связи или требуемого разрешения на приёмной стороне;

2) на приёмную сторону передается не полный кадр оценки фона, а N блоков значений отличий новой оценки фона относительно предыдущей оценки фона, смещенных на 128 уровней яркости. Этот приём позволяет увеличить сжатие этого типа кадров в 3,56 раза;

3) построена зависимость $PSNR=f(K)$. В процессе измерения параметры PSNR и K менялись из-за изменения параметра качества Q кодека JPEG.

Литература

1. Обработка изображений в прикладных телевизионных системах/ Астратов О.С., Вилесов Л.Д. и др. СПб.: ГУАП, 2012. 272 с.: ил.
2. Кузьмин С.А. Семантическое сжатие видеоинформации в системах видеонаблюдения// Журнал Радиоэлектроники (электронный журнал), №4, 2012 URL: jre.cplire.ru/jre/apr12/5/text.html.
3. Кузьмин С.А. Сегментация последовательностей изображений с регулируемой точностью и визуализация эффективности// Технические науки — от теории к практике. № 8 (21): сборник статей по материалам XXV международной научно-практической конференции. Новосибирск: Изд. «СибАК», 2013. С. 44-54.
4. Кузьмин С.А. Устранение влияния теней на точность выделения объектов в видеопоследовательностях// Журнал Радиоэлектроники (электронный журнал), №5, 2012 URL: jre.cplire.ru/jre/may12/2/text.html.
5. Кузьмин С.А. Исследование комбинации детектора импульсного шума в бинарных изображениях и процентильных фильтров// «Естественные и математические науки в современном мире»: материалы IX международной заочной научно-практической конференции. Новосибирск: Изд. «СибАК», 2013. С. 34-44.
6. Тимофеев Б.С. Видеопанорамы – новое направление развития систем телевизионного наблюдения// Материалы 7-ой МК «Телевидение: передача и обработка изображений». СПб., 2009. С. 165-172.
7. Shcherbakov M.A., Schegolev W.Y. A Wavelet-based Technique for Image Refinement, EUSIPCO-2000, Tampere, pp.1737-1739.
8. x264: the best low-latency video streaming platform in the world / J. Garrett-Glaser. [Б. м.], 2010 URL: x264dev.multimedia.cx/archives/249.

9. Кулешов С. В., Зайцева А. А., Аксенов А. Ю. Информационная избыточность в мультимедийных данных// Труды СПИИРАН, № 4 (23), 2012, с. 472–481.
 10. Обухова Н.А., Тимофеев Б.С. Методы повышения эффективности систем видеонаблюдения// Материалы 4-ой МК «Телевидение: передача и обработка изображений». СПб., 2005. С. 85-86.
 11. Красильников Н.Н. Методы увеличения степени сжатия изображений энтропийными кодерами// Информационно-управляющие системы, N1, 2004. С.10-13.
 12. Красильников Н.Н. Принципы обработки изображений, основанные на учете их семантической структуры// Информационно-управляющие системы, N1, 2008. С.2-6.
 13. Вилесов Л.Д. Обнаружение-измерение параметров объектов на изображении// Информационно-управляющие системы. 2004. № 1. С. 22-29.
 14. Кузьмин С.А. Сравнительное исследование способов оценки фона в видеопоследовательностях // Журнал Радиоэлектроники (электронный журнал), №3, 2012. - URL: jre.cplire.ru/jre/mar12/12/text.html.
 15. Шарапов Р.В. Аппаратные средства хранения больших объёмов данных // Инженерный вестник Дона, 2012, №4 (часть 2) URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1371.
 16. Толкачев Д.С. Формирование панорамного изображения с учетом параллакса при известной модели окружающего мира // Инженерный вестник Дона, 2013, №3 URL: ivdon.ru/ru/magazine/archive/n3y2013/1872.
 17. Mastriani M. Supercompression for Full-HD and 4k-3D (8k) Digital TV Systems//World Academy of Science, Engineering and Technology, International Science Index 48, 4(12), 2010. Pp. 485 - 497.
-

References

1. Astratov O.S., Vilesov L.D. et al. Obrabotka izobrazheniy v prikladnykh televizionnykh sistemakh [Image processing in applied TV systems] (Rus). Saint Petersburg: SUAI, 2012. 272 p.
 2. Kuzmin S.A. Journal of Radio Electronics (Rus), 2012, №4 URL: jre.cplire.ru/jre/apr12/5/text.html
 3. Kuzmin S.A. Novosibirsk, 2013, № 8 (21). Pp. 44-54.
 4. Kuzmin S.A. Journal of Radio Electronics (Rus), 2012, №5 URL: jre.cplire.ru/jre/may12/2/text.html
 5. Kuzmin S.A. Novosibirsk, 2013, № 9. Pp. 34-44.
 6. Timofeev B.S. Videopanoramy – novoe napravlenie razvitiya sistem televizionnogo nabljudeniya [Videopanorams - a new direction of development of video surveillance systems] Materialy 7-oj MK «Televidenie: peredacha i obrabotka izobrazhenij», SPb. 2009, pp. 165-172.
 7. Shcherbakov M.A., Schegolev W.Y. A Wavelet-based Technique for Image Refinement, EUSIPCO-2000, Tampere, pp.1737-1739.
 8. J. Garrett-Glaser, x264: the best low-latency video streaming platform in the world, 2010. URL: x264dev.multimedia.cx/archives/249
 9. Kuleshov S. V., Zaytseva A. A., Aksenov A. Y. Trudy SPIIRAN, 2012, № 4 (23), pp. 472–481.
 10. Obukhova N.A., Timofeev B.S., Materialy 4-oj MK «Televidenie: peredacha i obrabotka izobrazhenij», SPb., 2005, pp. 85-86.
 11. Krasilnikov N.N. Informatsionno-upravliaiushchie sistemy, 2004, №1. Pp. 10-13.
 12. Krasilnikov N.N. Informatsionno-upravliaiushchie sistemy, 2008, №1. Pp. 2-6.
 13. Vilesov L.D. Informatsionno-upravliaiushchie sistemy, 2004, №1. Pp. 22-29.
-



14. Kuzmin S.A. Zhurnal Radioelektroniki (jelektronnyj zhurnal), 2012, №3 URL: jre.cplire.ru/jre/mar12/12/text.html
15. Sharapov R.V. Inženernyj vestnik Dona (Rus), 2012, №4 (part 2) URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1371
16. Tolkachev D.S. Inženernyj vestnik Dona (Rus), 2013, №3 URL: ivdon.ru/ru/magazine/archive/n3y2013/1872
17. Mastriani M., Supercompression for Full-HD and 4k-3D (8k) Digital TV Systems, World Academy of Science, Engineering and Technology, International Science Index 48, 4(12), 2010. Pp. 485 - 497.