

Кодирование видеоданных в мобильных системах на основе объектной компенсации движения видеокамеры



Цветков В.Ю., доцент, к.т.н., доцент кафедры «Сети и устройства телекоммуникаций» БГУИР

Одной из важнейших проблем создания систем видеоаналитики является сжатие видеоинформации для минимизации ее объема при передаче и хранении. Особенно остро проблема сжатия видеоданных стоит в мобильных системах, где для передачи используется радиоканал. В этом случае сжатие рассматривается как средство достижения компромисса между скоростью и дальностью передачи, мощностью и массогабаритными параметрами передатчика. В настоящее время в мобильных системах видеоаналитики большое распространение получили видеокодеки MPEG-4 и H.264, основанные на блочной компенсации движения. Эффективность применения данных видеокодеков в мобильных системах ограничена отсутствием учета параллакса, возникающего при движении видеокамеры и проявляющегося в различной скорости перемещения изображений объектов в зависимости от их удаленности от видеокамеры.

Для сжатия видеоданных, формируемых движущейся видеокамерой в условиях параллакса, предлагается метод эффективного кодирования видеоданных на основе объектной компенсации движения видеокамеры. Суть метода состоит в сегментации и одноракурсной объектной декомпозиции опорного кадра видеоданных, соответствующего исходному положению видеокамеры в момент времени t_i и объектном предсказании следующего прогнозируемого кадра видеоданных, соответствующего смещенному положению видеокамеры в момент времени t_{i+1} , с использованием информации о характеристиках, параметрах установки и перемещения видеокамеры на интервале $(t_{i+1} - t_i)$. Пиксельная разность областей перекрытия полученного в результате прогнозного и прогнозируемого кадров состоит в основном из нулевых или близких к нулю значений и может быть эффективно сжата. Для реализации метода выполняются следующие шаги.

1. Сегментация опорного кадра

$M_F(t_i) = \|m_F(y, x, t_i)\|_{(y=0, Y-1, x=0, X-1)}$ видеоданных размером $Y \times X$ пикселей. Формируется матрица $M_S(t_i) = \|m_S(y, x, t_i)\|_{(y=0, Y-1, x=0, X-1)}$ сегментации, содержащая информацию о количестве S и расположении сегментов опорного кадра $M_F(t_i)$, значение каждого элемента $m_S(y, x, t_i) \in [0, S-1]$ которой указывает на номер сегмента, к которому он принадлежит. Для обеспечения устойчивости результатов сегментации к изменению яркости и действию шума предлагается использовать иерархическую сегментацию на основе разделения и слияния областей [1].

2. Одноракурсная объектная декомпозиция опорного кадра $M_F(t_i)$ видеоданных.

На основе матрицы $M_S(t_i)$ сегментации формируется матрица $M_D(t_i) = \|m_D(y, x, t_i)\|_{(y=0, Y-1, x=0, X-1)}$ объектной декомпозиции, содержащая информацию о количестве D и расположении объектов опорного кадра $M_F(t_i)$, которые представляются совокупностями одинаковых по значениям элементов $m_D(y, x, t_i) \in [0, D]$. Объектная декомпозиция основана на оценке положения сегментов $m_S(y, x, t_i)$ относительно предполагаемой проекции линии горизонта, предсказываемой с помощью модели $M_C(t_i)$ монокулярного дальномера при известных высоте установки, угле тангажа и параметрах внутренней калибровки видеокамеры [2]. К изображениям объектов относятся сегменты, пересекающие проекцию предполагаемой линии горизонта или расположенные рядом с ней. Элементам матрицы $M_D(t_i)$, соответствующим таким сегментам, присваивается значение номера объекта ($m_D(y, x, t_i) > 0$). Остальным элементам матрицы $M_D(t_i)$ присваивается нулевое значение ($m_D(y, x, t_i) = 0$) они соответствуют поверхности, на которой размещаются объекты, и заднему плану.

3. Определение расстояний от видеокамеры до объектов опорного кадра $M_F(t_i)$.

На основе модели $M_C(t_i)$ монокулярного дальномера и предположения о плоском характере поверхности, на которой находятся объекты опорного кадра $M_F(t_i)$, формируется матрица $D_Z(t_i) = \|d_Z(y, x, t_i)\|_{(y=0, Y-1, x=0, X-1)}$ расстояний от видеокамеры до объектов (рис. 1).

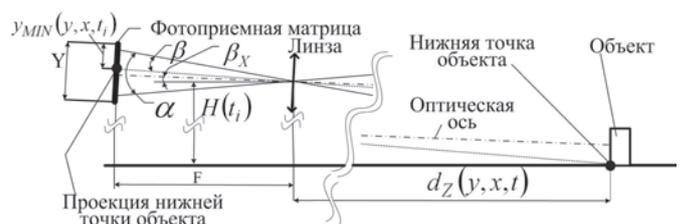


Рис. 1. Графическая интерпретация модели монокулярного дальномера

Расстояния от видеокамеры до элементов фона и объектов оцениваются по смещениям проекций элементов фона и точек местоположения объектов, определяемых значениями $y_{MIN}(y, x, t_i)$ минимальных y -координат соответствующих им сегментов на опорном кадре $M_F(t_i)$, от нижней границы опорного кадра $M_F(t_i)$ после компенсации крена видеокамеры [3]. Значения элементов $d_Z(y, x, t_i)$ матрицы $D_Z(t_i)$ определяют расстояния до соответствующих им элементов фона или объектов и вычисляются с помощью выражения $d_Z(y, x, t_i) = H(t_i) / \text{tg}(\alpha/2 + \beta_X(t_i) - \beta(t_i))$, (1) где $H(t_i)$ — высота установки видеокамеры в момент t_i ; α — угол обзора видеокамеры, соответствующий разрешению видеокамеры по вертикали (Y пикселей); $\beta_X(t_i)$ — угол тангажа видеокамеры; $\beta(t_i)$ — угол, соответствующий смещению проекции элемента фона или точки местоположения объекта относительно нижней границы опорного кадра.

4. Объектное предсказание прогнозируемого кадра $M_F(t_{i+1})$ видеоданных.

Формируется прогнозный кадр

$\hat{M}_F(t_{i+1}) = \|\hat{m}_F(y, x, t_{i+1})\|_{(y=0, Y-1, x=0, X-1)}$ для положения видеокамеры в момент t_{i+1} . При перемещении видеокамеры изображения объектов опорного кадра $M_F(t_i)$ сдвигаются на прогнозируемом кадре $M_F(t_{i+1})$ в зависимости от их удаленности от видеокамеры. Смещения изображений объектов на прогнозном кадре $\hat{M}_F(t_{i+1})$ вычисляются с помощью модели $M_C(t_i)$ монокулярного дальномера [3]. Значения пикселей прогнозного кадра, заслоненных объектами на опорном кадре, устанавливаются равными среднему значению фона.

5. Вычисление ошибки предсказания. На основе прогнозируемого $M_F(t_{i+1})$ и прогнозного $\hat{M}_F(t_{i+1})$ кадров формируется разностный кадр

$$\tilde{M}_F(t_{i+1}) = \|\tilde{m}_F(y, x, t_{i+1})\|_{(y=0, Y-1, x=0, X-1)}, \text{ каждый пиксель } \tilde{m}_F(y, x, t_{i+1}) \text{ которого вычисляется с помощью выражения } \tilde{m}_F(y, x, t_{i+1}) = m_F(y, x, t_{i+1}) - \hat{m}_F(y, x, t_{i+1}). \quad (2)$$

6. Сжатие ошибки предсказания. Разностный кадр $\tilde{M}_F(t_{i+1})$ сжимается как неподвижное изображение и передается в декодер.

7. Оценка ошибки объектного предсказания и выбор способа кодирования следующего кадра видеоданных. В кодере и декодере формируется восстановленный прогнозируемый кадр

$$\bar{M}_F(t_{i+1}) = \|\bar{m}_F(y, x, t_{i+1})\|_{(y=0, Y-1, x=0, X-1)}, \text{ учитывающий потери информации при сжатии. На основе прогнозируемого } M_F(t_{i+1}) \text{ и восстановленного прогнозируемого } \bar{M}_F(t_{i+1}) \text{ кадров в кодере вычисляется ошибка } E(t_{i+1}) \text{ предсказания с использованием выражения}$$

$$E_P(t_{i+1}) = \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} |m_F(y, x, t_{i+1}) - \bar{m}_F(y, x, t_{i+1})| / (YX). \quad (3)$$

Если ошибка предсказания не превышает заданной пороговой величины, восстановленный прогнозируемый кадр $\bar{M}_F(t_{i+1})$ считается опорным. Иначе, в качестве опорного выбирается следующий кадр $M_F(t_{i+2})$. В обоих случаях осуществляется переход на шаг 1.

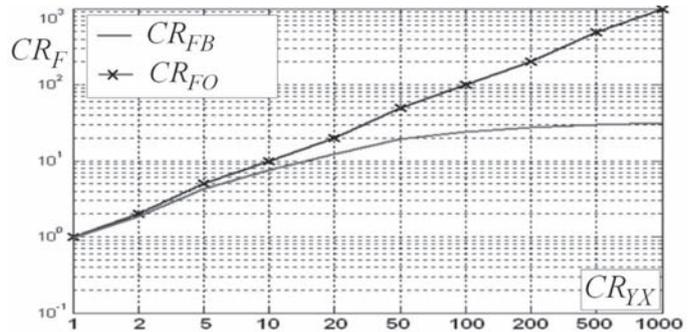


Рис. 2. Зависимость коэффициента CR_F от коэффициента CR_{YX}

Для оценки эффективности предложенного метода кодирования на основе объектной компенсации движения относительно метода кодирования на основе блочной компенсации движения получены зависимости коэффициента CR_F сжатия (CR_{FO} — для объектной компенсации движения и CR_{FB} — для блочной компенсации движения) одного прогнозируемого кадра $M_F(t_i)$ учетом информации о компенсации движения от коэффициента CR_{YX} ошибки предсказания $\tilde{M}_F(t_{i+1})$ (рис. 2).x

Из рис. 2 следует, что при сжатии ошибки предсказания $\tilde{M}_F(t_{i+1})$ в 100 раз ($CR_{YX} = 100$) блочная компенсация движения обеспечивает сжатие прогнозируемого кадра всего в 24,2 раза. Это связано с включением информации о компенсации движения, объем которой для кадра 480×640 пикселей, размера блока 8×8 пикселей и окрестности поиска 64×64 пикселей составляет $(480/8) \times (640/8) \times 16 = 76800$ бит. Использование для эффективного кодирования в тех же условиях объектной компенсации движения обеспечивает сжатие прогнозируемого кадра в 99,8 раза за счет минимизации информации о компенсации движения — всего 56 бит на кадр. В результате выигрыш в коэффициенте сжатия прогнозируемого кадра от использования объектной компенсации движения составляет $99,8/24,2 = 4,1$ раза. Выигрыш тем больше, чем больше размер кадра и коэффициент CR_{YX} .

Литература

1. Аль-Джубури Т.М. Адаптивная вейвлет-сегментация изображений на основе иерархического выращивания областей / Аль-Джубури Т.М., Конопелько В.К., Цветков В.Ю. // Доклады БГУИР. — 2011. — №2 (56). — С. 102—108.
2. Аль-Джубури Т.М. Монокулярная объектная декомпозиция изображения на основе предсказания положения линии горизонта / Аль-Джубури Т.М., Цветков В.Ю. Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных: материалы доклада международного научно-технического семинара. — Минск, январь-декабрь 2011 г. // БГУИР. — Минск, 2011. — С. 62—66.
3. Аль-Джубури Т.М. Метод объектно-ориентированной компенсации движения камеры на основе двухмерной декомпозиции видеоданных / Аль-Джубури Т.М., Конопелько В.К., Цветков В.Ю. // Инженерный вестник. — 2010. — Т.2(30). — №30. — С. 20—26. ■