ВЕСТНИК НОВОСИБИРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
Серия «Математика, механика, информатика»
Том V, 2005 г. Выпуск 1. С. 15–21

УДК 621.39

Сжатие неподвижных изображений и опорных кадров для видео по алгоритму сферической классификации с использованием предварительной обработки[1]

Н. А. Ваганова

§ 1. Трехфакторный анализ для выделения компонент цветного изображения

Пусть в цветовом пространстве RGB задан набор используемых цветных характеристик или так называемая палитра:

$$\pi_N = \{P_i = (R_i,\ G_i,\ B_i),\ i = 1, 2, \ldots, N\}, \tag{1}$$

где $R_i,\ G_i,\ B_i$ – компоненты вектора интенсивности красного, зеленого и синего цветов, соответствующие точке из палитры π_N. Определим так называемый «центр тяжести» — точку, минимально удаленную в среднеквадратичном смысле от точек палитры. Это будет вектор P^0 с координатами

$$R^0 = \frac{\sum_{i=1}^N R_i}{N},\ G^0 = \frac{\sum_{i=1}^N G_i}{N},\ B^0 = \frac{\sum_{i=1}^N B_i}{N},\ . \tag{2}$$

Теперь определим «главное направление» ν в палитре π_N, то есть найдем прямую L, проходящую через «центр тяжести» и наименее удаленную в среднеквадратичном смысле от этих точек. Прямую L будем искать в виде

$$L = \{P \in R^3 :\ P = P_0 + t\nu,\ -\infty < t < +\infty\}, \tag{3}$$

[1]Работа выполнена при финансовой поддержке: Фонда «ЦСР-Сибирь», Фонда содействия отечественной науке, интеграционного проекта № 6 «Развитие теории приближения функций с помощью сплайнов, тригонометрических полиномов и фракталов с приложением к созданию моделей гидротурбин и сжатию телевизионных изображений», выполняемом в СО РАН в содружестве с УрО РАН., проекта «Университеты России» № ур.03.01.195

где ν — это вектор единичной евклидовой длины. Тогда нам необходимо минимизировать целевой функционал

$$\Phi(\nu) = \sum_{i=1}^{N} \rho^2(L,\ P_i) \to \min_{\nu}. \tag{4}$$

Здесь через $\rho(L,\ P_i)$ обозначено евклидово расстояние между прямой L и точкой палитры P_i. После простой замены переменных $P_i' = P - P_0$ получаем

$$\rho^2(L,\ P_i) = \rho^2(L',\ P_i') = \|P_i'\|^2 - (P_i',\ \nu)^2.$$

В этом случае, задача минимизации переписывается в следующем виде:

$$\Phi(\nu) = \sum_{i=1}^{N} [\|P_i'\|^2 - (P_i',\ \nu)^2] \to \min_{\nu}. \tag{5}$$

с естественным дополнительным условием

$$\|\nu\|^2 = 1.$$

Возникает задача максимизации

$$\psi(\nu) = \sum_{i=1}^{N} (P_i', \nu)^2 \to \max_{\nu}. \tag{6}$$

Используя функцию Лагранжа

$$L(\nu,\ \lambda) = \psi(\nu) + \lambda(1 - \|\nu\|^2), \tag{7}$$

после дифференцирования получаем

$$\frac{1}{2}\frac{\partial L}{\partial \nu_k} = (\sum_{i=1}^{N} x_i^{(k)} P_i',\ \nu) - \lambda \nu_k = 0,\ k = 1,\ 2,\ 3, \tag{8}$$

где $x_i^{(1)} = R_i - R^0,\ x_i^{(2)} = G_i - G^0,\ x_i^{(3)} = B_i - B^0,\ i = 1,\ 2,\ ,\ldots,\ N$. Обозначим через B прямоугольную матрицу размера $N \times 3$:

$$B = \{x_i^{(k)}\}_{i=1,\ k=1}^{N,\ 3}, \tag{9}$$

тогда соотношение (8) перепишется в матричной форме

$$B^T B \nu = \lambda \nu. \tag{10}$$

Здесь $B^T B$ - это матрица 3×3 и имеет три собственных значения $\lambda_1,\ \lambda_2,\ \lambda_3$, которым соответствуют собственные вектора $\nu^{(1)},\ \nu^{(2)},\ \nu^{(3)}$. В этом случае, условие

$$\|\nu\|^2 = 1$$

выполняется автоматически. Поскольку матрица B^TB является симметрической, вещественнозначной, то все собственные значения этой матрицы различны и вещественнозначны. Пусть $\nu^{(1)},\ \nu^{(2)},\ \nu^{(3)}$ – это три собственных вектора для задачи (10) и пусть $B^TB\nu_l = \lambda_l\nu^{(l)},\ \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0, l = 1,\ 2,\ 3$. Найдем $\psi(\nu^{(l)})$,

$$\psi(\nu^{(l)}) = \sum_{i=1}^{N}(\sum_{k=1}^{3} x_i^{(k)}\nu_k^{(l)})^2 = (B^TB\nu^{(l)},\ \nu^{(l)}) = \|B\nu^{(l)}\|^2 = \lambda_l^2. \tag{11}$$

Таким образом, максимальное значение $\psi(\nu)$ или минимальное значение $\Phi(\nu)$будет соответствовать максимальному собственному значению λ_1 матрицы B^TB.

Если построить плоскость, проходящую через «центр тяжести» P^0 палитры π_N и ортогональную собственному вектору $\nu^{(1)}$, а затем найти в этой плоскости лучшее направление, то это будет соответствовать второму собственному вектору $\nu^{(2)}$ для матрицы B^TB. Последний собственный вектор $\nu^{(3)}$ определяет наименее важное направление в палитре π_N. Таким образом, собственные вектора $\nu^{(1)},\ \nu^{(2)},\ \nu^{(3)}$ определяют три фактора в порядке «важности» в соответствии с априорно заданной палитрой π_N. Уровень «важности», как мы предполагаем, регулируется соотношением $\lambda_1 \div \lambda_2 \div \lambda_3$.

В соответствии с рассмотренным 3-факторным анализом палитры цветное RGB изображение разлагается в три компоненты (3FA-разложение) $F_1,\ F2,\ F3$ по следующему правилу:

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \nu_1^{(1)} & \nu_2^{(1)} & \nu_3^{(1)} \\ \nu_1^{(2)} & \nu_2^{(2)} & \nu_3^{(2)} \\ \nu_1^{(3)} & \nu_2^{(3)} & \nu_3^{(3)} \end{pmatrix} \cdot \begin{pmatrix} R - R^0 \\ G - G^0 \\ B - B^0 \end{pmatrix}. \tag{12}$$

или в матричной форме

$$F = V(P - P^0).$$

Поскольку V – ортогональная матрица, обратное преобразование является также очень простым $P = P^0 + V^TF$. Этот тип разложения в дальнейшем будем называть 3FA-стандартом.

§ 2. Алгоритмы сжатия с потерями качества

Основной принцип сжатия цветного изображения может быть описан следующей схемой: исходное изображение, как RGB сигнал сначала преобразуется посредством умножения на матрицу, с целью выделения главных компонент $(\mathrm{F1}, \mathrm{F2}, \mathrm{F3})$, каждая из которых подвергается сжатию, то есть представлению в виде линейной комбинации заранее известных функций, либо представлению другой функцией, хранение которой в памяти компьютера занимает меньшее количество памяти. Далее полученные, в результате этого приближения вещественные коэффиценты разложения квантуются (округляются) (именно здесь происходит дополнительная потеря качества) и квантованные коэффициенты, а также некоторая дополнительная информация об изображении, передается на вход алгоритмов кодирования без потери информации или так называемой упаковки данных. В результате этой операции мы получаем код изображения.

Восстановление изображения заключается в следующем: код изображения, распаковывается, при этом полученные коэффициенты поступают на вход алгоритма декомпрессии (восстановления) компонент, далее проиходит постобработка изображения с целью устранения артефактов (например, блочности), обработанные компоненты подвергаются умножению на обратную матрицу, в результате которого мы получаем декодированное изображение.

Все общепринятые на сегодняшний день стандарты сжатия изображений с потерей информации условно можно разбить на две категории: алгоритмы основанные на Фурье-преобразовании и квантовании коэффициентов Фурье (JPEG-стандарт), алгоритмы основанные на разложении по всплескам (вейвлетам) (JPEG2000-стандарт). Фрактальное сжатие изображений из-за большой трудоемкости процесса кодирования, основанные на представлении изображения в качестве неподвижной точки некоторого сжимающего оператора (фрактальное сжатие), в настоящее время не является стандартом. Хотя по сравнению с JPEG2000 стандартом фрактальные методы позволяют восстанавливать изображение в различных масштабах. И имеют гораздо больший коэффициент сжатия. Эта особенность дает право признать, что некоторые модификации фрактальных алгоритмов позволят в будущем воспроизводить передавать большие изображения по каналам связи не увеличивая времени передачи данных. Основная математическая идея этого метода состоит в том, что для получения искомого сжимающего оператора в полутоновом изображении ищут подобные области различных размеров. Предположим, что мы кодируем не одно, а несколько изображений одного типа, тогда разумно составить базу в которую поместить по одному из повторяющихся (похожих)элементов изображения. База может быть большой и сравнивать площадки перебором будет длительным по времени. В связи с этим уместно использовать алгоритмы классификации площадок [1]. В этом случае алгоритм фрактального сжатия модифицируется при этом сохраняется масштабируемость изображения.

§ 3. Сравнение PAL и 3FA стандартов в экспериментах по сжатию изображений при помощи сферической классификации

В наших экспериментах для сравнения с 3FA стандартом мы будем использовать стандарт PAL [3]. Перед применением любых алгоритмов сжатия согласно этому стандарту производят умножение каждого сигнала цветности изображения на матрицу (матрицирование сигнала цветности):

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ -0.15 & -0.29 & 0.44 \\ 0.62 & -0.51 & -0.10 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \tag{13}$$

В качестве алгоритма сжатия в технических экспериментах использовался так называемый алгоритм сферической классификации площадок [1]. Этот алгоритм позволяет каждой полутоновой площадке изображения по математическим ожиданиям квадрантов сопоставить точку на сфере и соответствующую функцию из искусственной базы, так, что в процессе декодирования приходится лишь подсвечивать заданную площадку из фрактальной базы, контрастировать, ортогонально преобразовывать и ставить на место декодируемой площадки изображения. Предложенный алгоритм здесь рассматривать мы не будем,так как он подробно описан в [1]. Здесь мы а лишь остановимся на некоторых результатах его применения в комплексе с предобработкой изображения. Мы будем проводить сравнение стандартов PAL и 3FA (сэмплинг 4:4:4) на следующих изображениях: «гепард», «Лена», «кадр телепередачи»

Пример 1 (фотография гепарда). Исходное изображение имеет разрешение 640x480, представлено 256 цветами.

Рис. 1. Фотография гепарда. Коэффициент сжатия 88 раз, сигнал-шум 30 дБ

Трехфакторный анализ палитры дает следующие результаты: $P^0 = (149.23,\ 114.56,\ 100.96)$, $\lambda_1 \div \lambda_2 \div \lambda_3 = 2312 \div 266 \div 10$,

$$V = \begin{pmatrix} 0.56 & 0.62 & 0.55 \\ -0.72 & -0.05 & 0.69 \\ 0.40 & -0.79 & 0.47 \end{pmatrix}.$$

Пример 2 («Лена», фотография). Кодируемое изображение имеет разрешение 512x512. Количетво цветов или глубина изображения равна 16777216.

Рис. 2. Лена. Коэффициент сжатия 54 раза, сигнал-шум 30 дБ

Трехфакторный анализ палитры дает следующие результаты: $P^0 = (181.89,\ 104.32,\ 110.79)$, $\lambda_1 \div \lambda_2 \div \lambda_3 = 8324 \div 908 \div 106$,

$$V = \begin{pmatrix} 0.59 & 0.70 & 0.40 \\ -0.71 & 0.21 & 0.67 \\ 0.38 & -0.68 & 0.62 \end{pmatrix}.$$

Пример 3 (Опорный кадр видеопоследовательности). Исходное цветное изображение взято из трансляции телепередачи, задано как 24-битовое изображение с разрешением 384×288.

Рис. 3. Кадр видеопоследовательности. Коэффициент сжатия 76 раз, сигнал-шум 30 дБ

Трехфакторный анализ палитры дает следующие результаты $P^0 = (150.64,\ 115.38,\ 111.07)$, $\lambda_1 \div \lambda_2 \div \lambda_3 = 8.29 \div 7.47 \div 4.19$,

$$V = \begin{pmatrix} 0.58 & 0.58 & 0.57 \\ -0.63 & -0.11 & 0.77 \\ 0.50 & -0.81 & 0.30 \end{pmatrix}.$$

Эксперименты показали, что применение трехфакторного анализа вместе с алгоритмом сферической классификации увеличивает качество декодированного изображения в среднем на 1-2 децибелла.

§ 4. Заключение

В наших исследованиях по разбиению и сжатию изображений мы не стремились точно ответить на вопрос, что лучше общепринятое PAL разбиение цветового пространства или 3FA? На самом деле наши эксперименты показывают, что в данном случае 3FA разбиение лучше и это не удивительно, так как предложенный алгоритм настраивается на палитру кодируемого изображения. С другой стороны стандарт PAL является универсальным и дает примерно одинаковые коэффициенты сжатия для различных типов изображений. Опасно подвергать первую компоненту как PAL, так и 3FA стандартов большему коэффициенту сжатия, чем вторую и третью компоненты, так как первая полутоновая компонента носит более информативный характер, в свою очередь остальные полутоновые компоненты отвечают за степень окрашивания декодируемого изображения.

Литература

[1] *Н. А. Ваганова*, Построение дисперсионной и сферической фрактальных баз для задачи кодирования изображений, Вестник НГУ, Серия: матем., механика, информ., **1**, № 1 (2001), 23–39.

[2] *И. С. Грузман, В. С. Киричук, и др.*, Цифровая обработка изображений в информационных системах, Новосибирск, НГУ, 2000.

[3] *Ю. Г. Зубарев, Г. Л. Глориозова*, Телевизионная техника. Справочник, М., Радио и связь, 1994.