

ЭТАЛОННАЯ ОЦЕНКА КАЧЕСТВА ОЦИФРОВАННЫХ СТАТИЧЕСКИХ ИЗОБРАЖЕНИЙ И ВИДЕОСИГНАЛОВ В СИСТЕМАХ БЕЗОПАСНОСТИ

Филиппов А.К.¹, Шахтарин Б.И.²

Рассмотрены существующие проблемы оценки качества оцифрованных статических изображений и видеосигналов в системах безопасности с целью обеспечения таких важных элементов информационной безопасности, как достоверность и целостность. Проведен анализ ряда современных метрик (PSNR, SSIM и MS-SSIM) и методик эталонной оценки качества, наиболее широко распространенных в таких сферах, как цифровое кино и телевидение, видеоконференции. Рассмотрен ряд свойств и особенностей указанных метрик и методик, которые ограничивают их применение в системах охранных телевизионных, также известных как системы видеонаблюдения. Приведено краткое описание обобщенной математической модели, в которой видеоизображение рассматривается как комбинация четырех компонентов: гладких областей, контуров, текстур и шума. Базируясь на этой математической модели, разработан новый метод эталонной оценки качества статических изображений и видеосигналов, заключающийся в использовании метрик AEF и PWAE, предназначенных соответственно для оценки искаженности границ и величины максимального искажения яркости или цветности. Представлены детальное теоретическое обоснование предлагаемых метрик, описание методики их применения, а также экспериментальные результаты, демонстрирующие пригодность данного метода для оценки достоверности в системах, где применяется сжатие с потерями информации (в первую очередь, в системах видеонаблюдения).

Ключевые слова: сжатие с потерями информации, статические изображения, видеосигналы, системы охранные телевизионные, видеонаблюдение, эталонная оценка качества.

DOI: 10.21681/2311-3456-2017-2-61-74

Введение

Современные угрозы (такие, как терроризм, нелегальная миграция, организованная преступность и многочисленные приграничные конфликты) обуславливают широкое распространение систем безопасности, функционирующих, используя различные физические принципы. К ним относятся технические средства охраны и защиты периметра, системы контроля и управления доступом к объектам, охранно-пожарные системы, а также системы видеонаблюдения, для которых в соответствии с ГОСТ Р 51558-2008 используется термин «системы охранные телевизионные» (СОТ). В последние 5-10 лет роль СОТ непрерывно возрастала: от вспомогательной подсистемы, роль которой сводилась, как правило, к проверке информации, поступающей от других подсистем, СОТ эволюционировали до статуса ключевого элемента систем безопасности. Во многом такой прогресс в развитии СОТ стал возможен благодаря переходу от аналоговых технологий к цифровым, которые и используются в настоящее время совместно с сетевыми технологиями, построенными, как правило, на основе IP(Internet Protocol)-протокола.

Именно такое сочетание позволяет разрабатывать и проектировать недорогие, надежные и масштабируемые системы СОТ, использующие не уникальное (как было для аналоговых СОТ), а широко распространенное оборудование (в первую очередь, сетевые устройства) и элементную базу, предназначенную для цифровой обработки видеосигналов. Не удивительно, что одними из критически важных методов цифровой обработки видеосигналов, применяемых в СОТ, являются алгоритмы сжатия с потерями информации. Они позволяют существенно сокращать изначально большие объемы видеоданных до такого размера, который может быть передан по широко распространенным сетям связи (в частности, соответствующим стандартам IEEE 802.3 и IEEE 802.11) за счет ухудшения качества видеосигнала по сравнению с исходным, делая задачу оценки качества практически актуальной и теоретически значимой.

1. Постановка задач исследования

Как известно, одна из основных проблем при использовании сжатия с потерями состоит в выборе такого соотношения между объемом данных (англ. rate) и искажением (англ. distortion) видео-

1 Филиппов Алексей Константинович, кандидат технических наук, доцент, ООО «Техкомпания Хуавей», Москва, email: aleks_filippov@mail.ru

2 Шахтарин Борис Ильич, доктор технических наук, профессор, МГТУ им. Н.Э. Баумана, Москва, email: shakhtarin@mail.ru

изображения или видеопоследовательности, которое бы наилучшим образом соответствовало задачам, решаемым системой (в рассматриваемом случае, СОР). Эта проблема известна как RD-оптимизация (Rate-Distortion optimization, RD optimization, RDO). С математической точки зрения, она может быть сформулирована как условной оптимизации (constrained optimization) [1]:

$$\begin{cases} D(R) \rightarrow \min \\ R < R_c \end{cases}, \quad (1)$$

где $D(R)$ – величина искажения видеоизображения или видеопоследовательности (это значение зависит от объема данных R , получаемых на выходе кодера, что отражено в виде обозначения $D(R)$); R_c – целевое значение объема данных, которое не должно быть превышено при кодировании, то есть объем данных R на выходе кодера должен быть меньше величины R_c . Принимая во внимание тот факт, что задачи условной более сложны, чем задачи безусловной оптимизации, целесообразно сводить задачи первого типа к эквивалентным (то есть с совпадающим решением) задачам второго типа.

Используя метод множителей Лагранжа, эквивалентная задача безусловной оптимизации может быть сформулирована как [1]

$$J_{\text{opt}} = \min_i \{J_i\}, \quad (2)$$

где J_i – это функционал, подлежащий оптимизации для всех доступных значений i :

$$J_i = D_i + \lambda R_i, \quad (3)$$

λ – скалярный параметр функционала оптимизации, известный как множитель Лагранжа. Следовательно, для поиска оптимального соотношения между потерей качества исходного изображения D и объемом данных R необходимо знать значения двух множеств (наборов) величин: $\{D_i\}_{i=0}^{N-1}$ и $\{R_i\}_{i=0}^{N-1}$, где N – это число доступных значений i (по сути, это число таких комбинаций D и R , при которых значения обоих величин находятся в приемлемых пределах). Если методика определения объема данных R очевидна, то оценка качества видеоизображений и видеопоследовательностей, а также степени их искаженности D по сравнению с оригиналом является актуальной научной проблемой, не имеющей удовлетворительного решения до настоящего времени как в общем виде, так и для многих практически важных приложений. Так, для СОР критически важным требованием, предъявляемым к качеству видеосигнала, является обеспечение его *достоверности* (англ. signal

fidelity) [2]. В настоящей работе предлагается новый метод оценки качества статических изображений и видеосигналов, ориентированный именно на оценку их достоверности.

2. Аналитический обзор известных методов оценки качества оцифрованных изображений и видеосигналов

Как известно, методы оценки качества (англ. quality assessment) статических изображений и видеоданных разделяются на

- *субъективные*, то есть когда оценка осуществляется человеком посредством его органов зрения;

- *объективные*, то есть когда используются различные математически сформулированные критерии качества видеоизображений, называют также метриками (англ. metric), мерами (англ. measure) или индексами оценки качества (англ. quality assessment index, QAI).

Безусловно, субъективные методы имеют большее значение во многих применениях, где требуется оценка качества видеосигнала. Более того, для подавляющего большинства случаев именно субъективная оценка используется для верификации результатов объективной, и это совершенно обоснованно там, где пригодность качества видеосигнала для человеческого восприятия является конечной целью (например, киноиндустрия, телевидение, видеоконференции, сервисы передачи видеоданных, снятых пользователями). Однако, как было сказано выше, СОР представляют собой особый случай, где, в первую очередь, требуется *достоверность сигнала*, определяемая как репрезентативность оцениваемого сигнала R по отношению эталонному S , который считается практически равным реальному сигналу S_r (искажениями в оптической подсистеме, шумами сенсора и ошибками преобразования сигнала из аналоговой формы в цифровую, как правило, или стараются компенсировать или пренебрегают в виду их малого влияния на сигнал S). Очевидно, что репрезентативность численно может оцениваться через степень близости R и S , что весьма похоже на математическое определение метрики, являющейся как раз мерой близости. Таким образом, особую роль в СОР приобретают объективные методы оценки качества видеосигнала, которые включают в себя три большие группы [3]:

- *эталонная оценка* (англ. full-reference quality assessment): полностью известны как эталонный $S(x, y)$, так и оцениваемый $R(x, y)$ сигналы, то есть возможно провести их поэлементное сравнение;

– *псевдоэталонная оценка* (англ. reduced-reference quality assessment): оцениваемый сигнал $R(x, y)$ известен полностью, а об эталонном сигнале $S(x, y)$ имеются лишь некоторые сведения, а именно значения его характеристик (англ. features): например, диапазон частот, положение наиболее важных участков изображений (англ. Region of interest, ROI), максимальное и минимальное значение яркости пикселей, а также их распределения;

– *неэталонная оценка* (англ. no-reference quality assessment, известная также как «blind» quality assessment): известен только оцениваемый сигнал, а информация об эталонном сигнале отсутствует полностью, но имеются априорные сведения о типах присутствующих искажений. Например, оценка может производиться путем анализа перепадов между блоками, известных как блочные артефакты (англ. blocking artifacts) и образуемые после сжатия рядом кодеков (в частности, совместимыми со стандартом JPEG).

Очевидно, что эталонные методы оценки качества изображений обладают принципиально большей точностью, чем псевдоэталонные и неэталонные, а точность оценки – один из ключевых факторов при определении достоверности видеосигнала. Далее в настоящей работе рассматриваются только эталонные методы оценки качества, поскольку именно они, как правило, используются в кодеках для RD-оптимизации. Как показано на рис. 1, на стороне кодера имеется RDO-механизм, отсутствующий на стороне декодера. Как было показано выше, именно в нем производится эталонная оценка качества с тем, чтобы установить степень искажения D сжатого сигнала относительно эталонного, поскольку на стороне кодера доступны как эталонный, так и оцениваемый сигналы. Кодер осуществляет перебор пар значений $\{D_i(p_i)\}_{i=0}^{N-1}$ и $\{R_i(p_i)\}_{i=0}^{N-1}$ для набора значений параметра p , принимающего N различных значений, чтобы в результате найти такое его значение p_{opt} , при котором функционал оптимизации $J(p)$ оказывается в точке минимума:

$$p_{opt} = \arg \min_i \{J(p_i)\}. \quad (4)$$

Эти переборы могут осуществляться независимо друг от друга (например, при реализации алгоритма полного перебора значений параметра p) или решение о необходимости следующей итерации может приниматься на основе информации о предыдущей итерации (например, при использовании RDO на основе метода градиентного спуска). Именно при осуществлении этих переборов, количество которых, как правило, определяет вычислительную сложность кодера, можно контролировать качество той части видеосигнала, которая обрабатывается в настоящий момент: если объем искажений D_i для p_i превышает некоторое заранее установленное значение D_{THR} , то такие значения параметра p_i просто исключаются из дальнейшего рассмотрения:

$$J_i(p_i) = \infty \mid D_i(p_i) > D_{THR}. \quad (5)$$

Как видно из рис. 1, на стороне декодера оценка качества для дополнительного подтверждения достоверности восстановленного видеосигнала может проводиться с применением псевдоэталонных или неэталонных методов.

Наиболее распространенным и широко применяемым критерием для эталонной оценки качества является пиковое значение сигнал/шум (англ. Peak Signal-to-Noise Ratio, PSNR), рассчитываемое как

$$PSNR = 10 \log_{10} \left(\frac{N \times M \times \left(\max_{x=0, N-1, y=0, M-1} \{S(x, y)\} \right)^2}{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (S(x, y) - R(x, y))^2} \right), \quad (6)$$

где $S(x, y)$ и $R(x, y)$ эталонный и оцениваемый (в случае оценки качества после сжатия с потерями, соответственно исходный и восстановленный) двумерные сигналы размером N по оси абсцисс и M по оси ординат. Операция $\max_{x=0, N-1, y=0, M-1} \{S(x, y)\}$ означает выбор максимально возможного значения, принимаемого исходным сигналом. $\max_{x=0, N-1, y=0, M-1} \{S(x, y)\} = 2^n - 1$ для видеоизображений, где цветовая плоскость p представляется через n бит, то есть значение пикселя $S_p(x, y)$ для

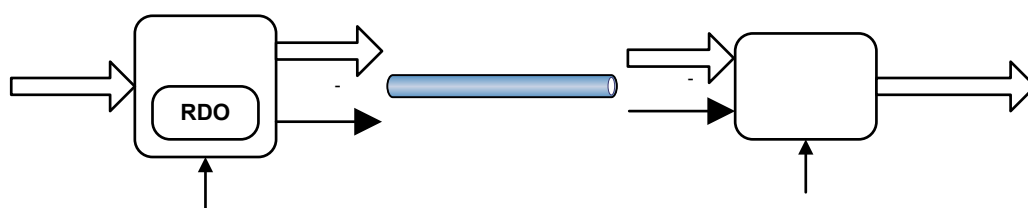


Рис. 1. Сжатие видеосигналов с потерями в COT

цветовой плоскости p находится в диапазоне от 0 до $2^n - 1$: $S_p(x, y) \in [0, 2^n - 1]$ (например, для 8-битного представления $2^8 - 1 = 255$). В ряде источников (например, в [4, 5]) считается, что именно PSNR отражает достоверность сигнала. Это утверждение базируется на том простом факте, что в расчете PSNR участвуют попиксельные разницы между исходным и восстановленным изображениями: $S(x, y) - R(x, y)$. Однако вряд ли такую точку зрения можно считать верной. В том, что в PSNR никак не учитывается корреляция между ошибками, которая может существенно влиять как субъективное восприятие видеосигнала, так и на его достоверность. Рассмотрим результаты простого эксперимента, представленные на рис. 2:

- на рис. 2а показано исходное изображение Lenna, имеющее 8-битное представление для каждого пикселя и разрешение 512x512 пикселей;
- на рис. 2б показано изображение Lenna, к которому добавлен аддитивный гауссов белый шум (PSNR=34,65 дБ);
- на рис. 2в показано изображение Lenna, где на лицо нанесено искажение в виде линии черного цвета (PSNR=34,65 дБ).

При сравнении всех изображений между собой можно заметить, что

- изображения на рис. 2а и рис. 2б практически идентичны, то есть достоверность данного изображения, к которому добавлен некоррелированный шум, не вызывает сомнений, поскольку значение PSNR достаточно велико;
- изображения на рис. а и рис. 2в отличаются весьма существенно, поскольку подобное искажение при идентификации объекта может быть использовано как отличительный признак, то есть изображение на рис. в не может быть признано

достоверным, что свидетельствует о том, что значение PSNR при добавлении коррелированных искажений не может использоваться для определения достоверности видеосигнала.

Весьма очевидна причина, порождающая указанную проблему: оценка производится для всего изображения, при этом величина искажений в отдельных его частях фактически усредняется – это, безусловно, не подходит для решения задачи определения достоверности видеосигнала. Следует подчеркнуть, что данная проблема типична не только для PSNR, но и для других широко распространенных индексов оценки качества, в частности, для SSIM (англ. Structural Similarity, структурное подобие), которая представляет собой комбинацию трех величин, сравнивающих отличия соответственно яркости l , контраста c и структуры s у видеосигналов x и y (рис. 3) [3]:

$$SSIM(x, y) = f(l(x, y), c(x, y), s(x, y)). \quad (7)$$

Выражение (7) может быть преобразовано к следующему виду [3]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) \cdot (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) \cdot (\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (8)$$

$$\text{где } \mu_x = \frac{\sum_{i=1}^N x_i}{N}, \quad \sigma_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu_x)^2}{N-1}} \quad \text{и}$$

$$\sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \mu_x) \cdot (y_i - \mu_y)}{N-1}, \quad \text{а } C_1 = (K_1 \cdot L)^2,$$

$C_2 = (K_2 \cdot L)^2$ и C_3 – малые константы, которые принимают следующие значения: $K_1 = 0,01$, $K_2 = 0,03$ и $C_3 = C_2/2$ [3, 7].



а) Исходное изображение Lenna



б) Изображение, к которому добавлен аддитивный гауссов белый шум (PSNR=34,65 дБ)



в) Изображение, к которому добавлена линия черного цвета (PSNR=34,65 дБ)

Рис. 2. Сравнение влияния некоррелированных и коррелированных искажений на достоверность видеосигнала, а также на величину PSNR [6]

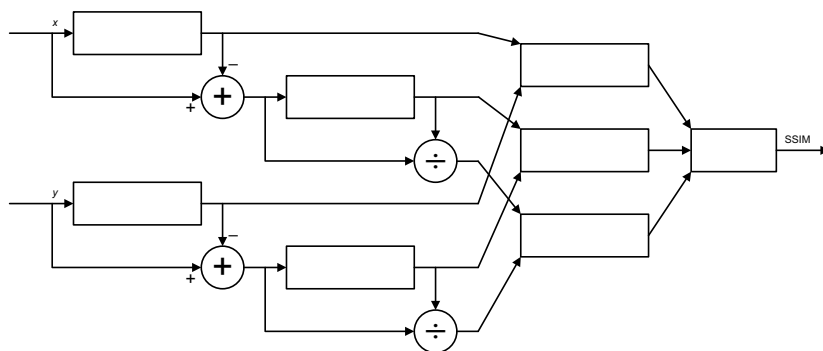


Рис. 3. Индекс оценки качества SSIM [3]

Выражение (8) широко используется для эталонной оценки качества видеосигналов после сжатия с потерями. Как правило, указанная формула применяется не ко всему изображению, а только к его локальной области Ω . Для практической реализации парадигмы локальной обработки при оценке качества изображений применяется метод скользящего окна (рис. 4), когда расчет некоторой метрики производится в окне, имеющем прямоугольную форму и размеры $H \times W$, где H и W – соответственно её высота и ширина. После завершения очередного вычисления окно смещается на величину Δ_H по вертикали и на Δ_W по горизонтали, образуя перекрытие между соседними окнами, равное $(H - \Delta_H) \cdot (W - \Delta_W)$. Как правило, последовательные смещения производятся только по одной из координатных осей. В [3] используется окно размером 8×8 и смещение в один пиксель. Изменения положения окна в обоих направлениях производится только в специфических случаях (например, при оценке качества вдоль границ). В завершении рассчитывается усредненное значение SSIM (англ. mean SSIM, MSSIM) для всех

M значений, которые были получены при обработке изображения скользящим окном:

$$MSSIM(x, y) = \frac{1}{M} \sum_{j=0}^{M-1} SSIM(x_j, y_j), \quad (9)$$

где x_j и y_j – эталонный и оцениваемый (искаженный) сигналы в окне j .

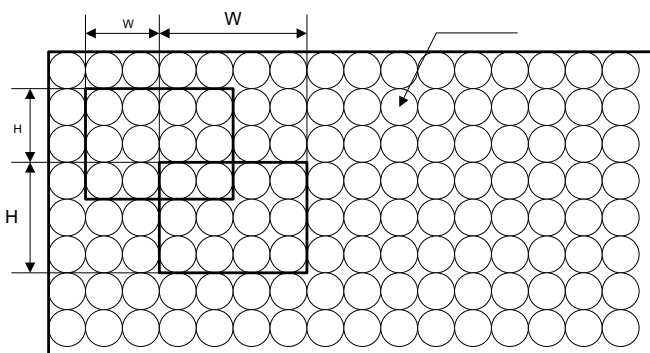
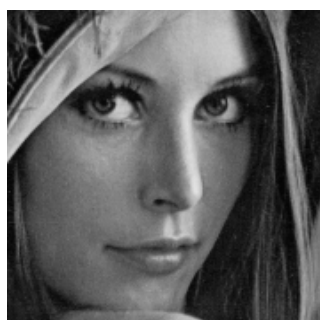
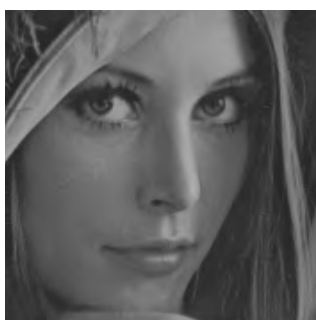


Рис. 4. Применение метода скользящего окна при решении задачи оценки качества видеосигналов

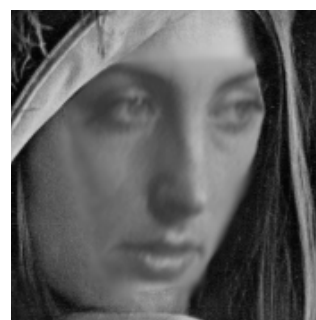
Рассмотрим то, насколько SSIM может быть использован для оценки достоверности сигнала, снова проведя простой эксперимент. На рис. 5 представлены три изображения:



а) Фрагмент исходного изображения Lenna



б) Фрагмент изображения, сжатого Kakadu кодером, совместимым со стандартом кодирования JPEG 2000:
PSNR=21,64 дБ;
SSIM=0.8095



в) Фрагмент изображения, искаженного путем добавления части лица другого человека:
PSNR=21,23 дБ;
SSIM=0.8028

Рис. 5. Сравнение влияния сжатия с потерями и коррелированных искажений на достоверность видеосигнала, а также на величины PSNR и SSIM

- фрагмент исходного изображения Lenna,
- упомянутый выше фрагмент, подвергнутый компрессии с потерями;
- упомянутый выше фрагмент, искаженный путем подмены лица одного человека на лицо другого.

Для эталонной оценки качества изображений, представленных на рис. 5б и 5в, использовались одновременно PSNR и SSIM. Полученные значения метрик отличаются на 0,41 дБ для PSNR и на 0,0067 для SSIM, что не является значительной разницей, хотя, по сути, с оригиналом сравниваются лица двух разных людей. Следует подчеркнуть, что подмена лиц могла бы быть произведена более качественно, а это позволило бы еще больше сблизить значения PSNR и SSIM. Таким образом, можно сделать вывод, что ни SSIM, ни PSNR в сочетании с PSNR не могут с достаточно высокой надежностью гарантировать достоверность видеосигнала.

Основная причина, которая не позволяет указанным метрикам быть эффективными при оценке достоверности видеосигнала, обусловлена, в первую очередь, не в их математическими свойствами, а тем, что локальные оценки искажений D_j усредняются для всего видеосигнала. Очевидно, что в примерах, показанных на рис. 5в и 5в, искажения D_j , локализованные на отдельных фрагментах j изображений, значительно превышают средний уровень \bar{D} ошибки для изображений в целом, т.е. $D_j > \bar{D}$. Таким образом, для получения достоверных результатов необходимо производить сравнение величины искажения D с пороговым значением D_{THR} , отражающим максимально допустимую величину искажения при сохранении достоверности, не только для всего изображения, но и локально, чтобы исключить недостоверные фрагменты. Очевидно, что если для некоторой области Ω изображения величина искажения D_Ω превышает D_{THR} ($D_\Omega > D_{THR}$), то эта область Ω должна быть помечена как недостоверная, или все изображение целиком должно быть признано недостоверным. Не стоит пояснять, что в случае первого варианта возникает весьма много сложностей. В частности, не понятно, каких размеров и формы должна быть

область Ω . Если она будет достаточно большой, то это не позволит решить обозначенную выше проблему. В противном случае, объем дополнительно передаваемых и/или сохраняемых метаданных будет весьма значительным. В предельном случае, когда значение бинарного флага «Достоверен / Не достоверен» указывается для каждого пикселя отдельно, необходимо к основному изображению добавить еще и бинарное, что может существенно снизить эффективность процедуры компрессии. Таким образом, решение лучше принимать или для видеоизображения целиком или отдельно для каждого важного участка изображения (ROI), что не приведет к существенному увеличению объема метаданных. С другой стороны, при таком подходе даже относительно небольшое по размерам искажение может привести к тому, что все изображение или его достаточно большой фрагмент будут признаны недостоверными, а это может также негативно сказаться на эффективности сжатия, поскольку потребует увеличения объема передаваемых данных R без реальной на то необходимости. Каким же образом следует производить локальную оценку качества с целью определения достоверности?

Давайте рассмотрим одну из модификаций метрики SSIM, которая называется MS-SSIM (Multi-Scale SSIM) и является его многомасштабным расширением [7], то есть оценка производится не только для исходного изображения, но и для её уменьшенных вариантов (рис. 6). Значение метрики рассчитывается по формуле

$$MS-SSIM(x, y) = [L_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} \cdot [s_j(x, y)]^{\gamma_j}, \quad (10)$$

где $j \in [1, M]$ – порядковый индекс текущего масштаба, что означает уменьшение размера изображения по вертикали и горизонтали по сравнению с исходным в 2^{j-1} раз (что обозначено на рис. 9 символом « $2 \downarrow$ »); α_M , β_j и γ_j – величины, используемые в качестве весовых показателей для каждой из компонент (чтобы можно было настроить их относительную важность) [7]. С целью упрощения выбора параметров для MS-SSIM в [7] принято, что $\alpha_M = \beta_j = \gamma_j$ и $\sum_{j=1}^M \gamma_j = 1$.

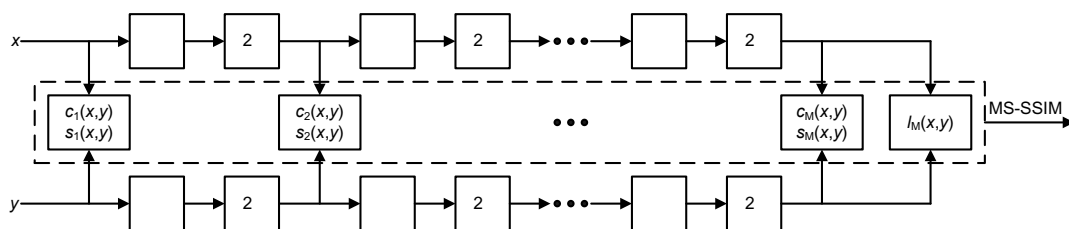


Рис. 6. MS-SSIM: многомасштабное расширение индекса оценки качества SSIM [7]

Как видно из рис. 6, оценка производится итерационно по мере масштабирования размеров изображений (как оригинального, так и искаженного). Во избежание эффекта наложения частот перед уменьшением размера изображения производится фильтрация с помощью фильтра низких частот (ФНЧ). Результаты, полученные после выполнения всех итераций, используются для расчета окончательного значения MS-SSIM по формуле (10). Подход, заключающийся в применении многомасштабной оценки качества видеосигналов и предложенный в [7], может быть применен к рассматриваемой задаче с целью установления размера минимально достоверного элемента. Необходимость такой процедуры обусловлена тем фактом, что в СОТ решаются задачи, предъявляющие различные требования к качеству достоверного видеоизображения [8]. Задачи можно разделить на три большие группы в порядке возрастания требований к качеству видеосигнала:

- детектирование (например, присутствия человека или транспортного средства);
- классификация (например, квалификация преступления и/или подсчет числа его участников);
- идентификация (например, человека или транспортного средства).

Таким образом, определение размера минимально достоверного элемента видеоизображения позволяет оценить пригодность того или иного видеосигнала для решения перечисленных задач. Предлагаемый алгоритм оценки, совмещающий в себе метод скользящего окна с многомасштабным анализом и отвечающий требованиям, описанным выше, представлен на рис. 7, из которого видно, что для оценки искажений в принципе может быть использована любая метрика D . На входе алгоритма задается начальный масштаб $s_0 = 1$. В процессе анализа рассчитывается значение метрики D_k внутри скользящего окна с номером k и сравнивается с пороговым значением D_{THR} . Размер скользящего окна k и количество M_j таких окон зависит от номера итерации j , то есть определяется

масштабом s_j . На выходе алгоритма получается значение достоверного масштаба s_j , соответствующее номеру итерации j , или констатируется тот факт, что достоверных масштабов, превышающих величину минимально допустимого масштаба S_{min} , не обнаружено по причине искажений, присутствующих в оцениваемом видеоизображении. Безусловно, сама по себе методология проведения измерений и принятия решения о достоверности изображения крайне важна, но это не исключает необходимости разработки такой метрики, которая бы наилучшим образом соответствовала требованиям, предъявляемым СОТ.

Следует подчеркнуть, что не всегда достаточно одной метрики, чтобы корректно оценить качество искаженного изображения. Однако при использовании двух и более метрик они должны быть логически непротиворечивыми (что выра-

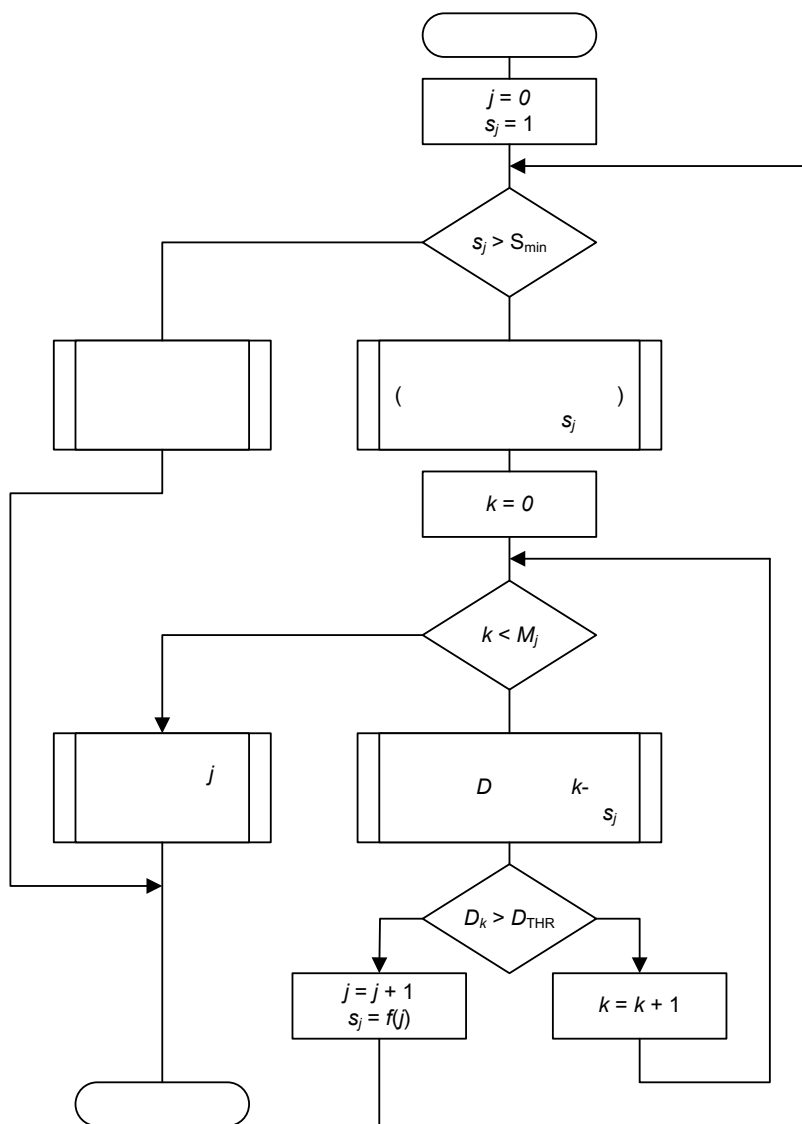


Рис. 7. Оценка достоверности изображения с использованием многомасштабного анализа

жается, например, в одинаковых трендах), что не всегда выполняется. В частности, для сочетания PSNR и MS-SSIM могут наблюдаться различные тренды относительно среднего значения для некоторой базы видеопоследовательностей, как было показано в [9], например, для последовательности Sparkler. С тем чтобы избежать подобных ситуаций, можно использовать несколько метрик, ориентированных на оценку различного типа артефактов.

3. Обобщенная математическая модель оцифрованных изображений и видеосигналов

Математически оцифрованное статическое изображение P в градациях серого цвета (англ. grayscale still image) представляется как двумерная матрица $M \times N$, элементы которой находятся в диапазоне $D = [0, 2^n - 1]$: $P_{M \times N} = (p_{ij}) : p_{ij} \in [0, 2^n - 1] \subset \mathbb{N}_0$. Аналогичным образом видеосигнал V в градациях серого цвета можно определить как трехмерную матрицу $M \times N \times T$: $V_{M \times N \times T} = (v_{ijt}) : v_{ijt} \in [0, 2^n - 1] \subset \mathbb{N}_0$, где $t \in [0, T - 1] \subset \mathbb{N}_0$ используется для обозначения дискретного времени. Математическая модель, используемая в настоящей работе, исходит, главным образом, из тех же самых предположений, которые изложены в [10], а именно: изображение представляется собой композицию трёх геометрически разделенных компонентов: гладкие области $f_{smooth}(x, y)$, текстуры $f_{tex}(x, y)$ и контуры $f_{edge}(x, y)$. Представленная модель отличается от трехкомпонентного варианта [10], как минимум, в двух аспектах. Во-первых, данная модель максимально формализована, что позволяет описать изображение в целом, а не просто ограничиться констатацией свойств каждой из компонент. Во-вторых, вводится четвертый компонент – шумовой сигнал $f_{noise}(x, y)$. Следует подчеркнуть, что шум $f_{noise}(x, y)$, генерируемый сенсором (шум сенсора) или носителем видеоданных (например, зернистость пленки, используемой для записи видеоданных в аналоговых системах), по своим свойствам отличается тем, что его воздействие на изображение заключается в искажении других компонент, с которыми шум пространственно не разделён. Таким образом, в самом общем виде изображение $P(x, y)$ может быть представлено как

$$P(x, y) = \Phi(f_{smooth}(x, y), f_{tex}(x, y), f_{edge}(x, y), f_{noise}(x, y)), \tag{11}$$

где $\Phi(\cdot)$ – оператор, генерирующий изображение путем композиции всех 4 компонент. Указан-

ный оператор может быть формально раскрыт следующим уравнением:

$$P(x, y) = f_{smooth}(x, y) \cdot \chi_{\Omega_{smooth}}(x, y) + f_{tex}(x, y) \cdot \chi_{\Omega_{tex}}(x, y) + f_{edge}(x, y) \cdot \chi_{\Omega_{edge}}(x, y) + f_{noise}(x, y), \tag{12}$$

где χ_{Ω} – характеристическая (индикаторная) функция множества $\Omega \subset \mathbb{N}_0^2$, определяемая как

$$\chi_{\Omega}(x, y) = \begin{cases} 1, & (x, y) \in \Omega \\ 0, & (x, y) \notin \Omega \end{cases}. \text{ Принимая обозначения}$$

$$\begin{aligned} \chi_{\Omega_0}(x, y) &= \chi_{\Omega_{smooth}}(x, y), & f_0(x, y) &= f_{smooth}(x, y), \\ \chi_{\Omega_1}(x, y) &= \chi_{\Omega_{tex}}(x, y), & f_1(x, y) &= f_{tex}(x, y), \\ \chi_{\Omega_2}(x, y) &= \chi_{\Omega_{edge}}(x, y), & f_2(x, y) &= f_{edge}(x, y), \end{aligned}$$

выражение (11) может быть переписано следующим образом:

$$P(x, y) = \sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \cdot f_i(x, y) + f_{noise}(x, y), \tag{13}$$

где $R = 3$ – количество пространственно разделенных компонент. Возьмем частные производные $P(x, y)$:

$$\begin{aligned} \frac{\partial P(x, y)}{\partial x} &= \sum_{i=0}^{R-1} \left(\frac{\partial \chi_{\Omega_i}(x, y)}{\partial x} f_i(x, y) + \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial x} \right) + \frac{\partial f_{noise}(x, y)}{\partial x} = \\ &= \sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial x} + \frac{\partial f_{noise}(x, y)}{\partial x} + \sum_{i=0}^{R-1} \frac{\partial \chi_{\Omega_i}(x, y)}{\partial x} f_i(x, y) = \\ &= \sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial x} + \frac{\partial f_{noise}(x, y)}{\partial x} + \sum_{i=0}^{R-1} \delta_{x_{0i}}(x - x_{0i}) f_i(x, y) \end{aligned} \tag{14}$$

$$\begin{aligned} \frac{\partial P(x, y)}{\partial y} &= \sum_{i=0}^{R-1} \left(\frac{\partial \chi_{\Omega_i}(x, y)}{\partial y} f_i(x, y) + \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial y} \right) + \frac{\partial f_{noise}(x, y)}{\partial y} = \\ &= \sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial y} + \frac{\partial f_{noise}(x, y)}{\partial y} + \sum_{i=0}^{R-1} \frac{\partial \chi_{\Omega_i}(x, y)}{\partial y} f_i(x, y) = \\ &= \sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial y} + \frac{\partial f_{noise}(x, y)}{\partial y} + \sum_{i=0}^{R-1} \delta_{y_{0i}}(y - y_{0i}) f_i(x, y) \end{aligned} \tag{15}$$

где $\frac{\partial \chi_{\Omega_i}(x, y)}{\partial x} = \delta_{x_{0i}}(x - x_{0i})$ и $\frac{\partial \chi_{\Omega_i}(x, y)}{\partial y} = \delta_{y_{0i}}(y - y_{0i})$

– частные производные характеристической функции множества, которые после дифференцирования фактически обращаются в дельта-функции Дирака, как показано на рис. 8. Прежде, чем перейти к дальнейшему рассмотрению проблемы, следует пояснить, почему к дискретизированным по времени и квантованным по уровню сигналам, которыми являются изображение $P(x, y)$ и видеосигнал $V(x, y, t)$, применяется операция взятия частных производных, типичная для использования с непрерывными сигналами. Это связано с тем, что как изображения, так и видеосигналы могут подвергаться субпиксельной интерполяции с целью увеличения их разрешения. Необходи-

мость такой операции может быть обусловлена различными причинами, например, требованием повышенной точности при предсказании видеосигналов, используя метод компенсации движения (англ. motion compensation) на основе расчета векторов смещения (движения). В целом, можно утверждать, что сигналы, подвергаемые субпиксельной интерполяции, будут близки по свойствам к непрерывным сигналам, что позволяет применять к ним методы дифференцирования и интегрирования.

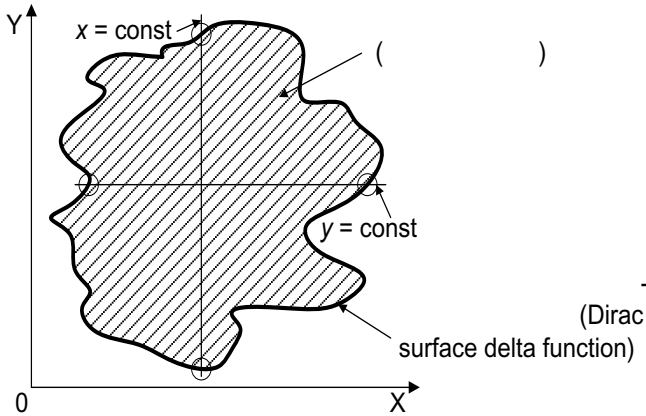


Рис. 8. Взаимосвязь характеристической функции множества и поверхностной дельта-функции Дирака δ_{Ω_i} (Dirac surface delta function), используя условное геометрическое представление

Проведем анализ слагаемых, входящих в выражения (14) и (15). $\sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial x}$ и $\sum_{i=0}^{R-1} \chi_{\Omega_i}(x, y) \frac{\partial f_i(x, y)}{\partial y}$ представляют собой значения производных четырех компонент в областях, ограниченных характеристическими функция-

ми множеств, а $\frac{\partial f_{\text{noise}}(x, y)}{\partial x}$ и $\frac{\partial f_{\text{noise}}(x, y)}{\partial y}$ – частные производные шумового сигнала. Наконец, сингулярные слагаемые $\sum_{i=0}^{R-1} \delta_{x_{0i}}(y - y_{0i})f_i(x, y)$ и $\sum_{i=0}^{R-1} \delta_{y_{0i}}(x - x_{0i})f_i(x, y)$ описывают границы между различными сегментами изображения. Таким образом, изменения указанных границ может оказывать существенное влияние на выражения (14) и (15), что и будет использовано нами в метриках, ориентированных на определение достоверности видеосигналов.

4. Предлагаемый метод эталонной оценки качества оцифрованных изображений и видеосигналов

Предлагаемый метод оценки качества оцифрованных изображений и видеосигналов заключается в использовании двух метрик [6, 11], одна из которых (PWAE, англ. Peak Weighted Absolute Error, пиковое значение взвешенной абсолютной ошибки) оценивает величину максимального искажения яркости или цветности, а вторая (AEF, англ. Activity Error Function, функция ошибки активности) используется для оценки искажения границ, а также подобных объектов (например, шума) в изображении. PWAE определяется как

$$PWAE = \max_{c=0, C-1} \left(w_c \max_{x, y \in \Omega} |S(x, y, c) - R(x, y, c)| \right), \quad (16)$$

где c – это номер цветовой плоскости, C – общее число цветных плоскостей, w_c – весовой коэффициент, оценивающий значимость цветовой плоскости. PWAE позволяет контролировать максимальное значение яркостных и цветных

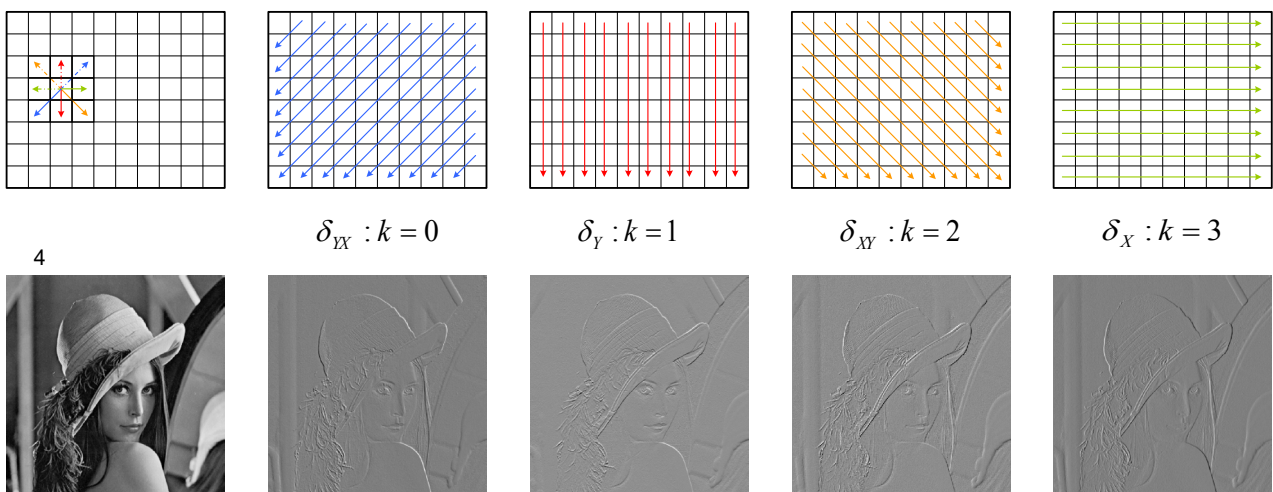


Рис. 9. Изображение и его производные по 4 направлениям

искажений, появляющихся внутри области Ω (в рассматриваемом случае – внутри скользящего окна), путем его сравнения с пороговым значением. Расчет значения данной метрики не требует значительных вычислительных ресурсов и может быть частично совмещен с расчетом АЕФ, которая использует значения численных производных как эталонного, так и оцениваемого изображений (рис. 9). Это связано с тем, что АЕФ ориентирована на оценку искажений геометрических контуров, которые по определению представляют собой границы объектов и сравнительно легко локализируются в пространстве, порождаемом путем взятия численной производной по определенному направлению, как это показано на рис. 9.

Очевидно, что для получения полной информации о контурах производные должны быть взяты, как минимум, в двух взаимно ортогональных направлениях (например, δ_x и δ_y или δ_{xy} и δ_{yx}). В рассматриваемом примере производные вычисляются по формулам

$$\delta_s(x, y, k) = \begin{cases} S(x-1, y+1) - S(x, y), k = 0 \\ S(x, y+1) - S(x, y), k = 1 \\ S(x+1, y+1) - S(x, y), k = 2' \\ S(x+1, y) - S(x, y), k = 3 \end{cases} \quad (17)$$

$$\delta_r(x, y, k) = \begin{cases} R(x-1, y+1) - R(x, y), k = 0 \\ R(x, y+1) - R(x, y), k = 1 \\ R(x+1, y+1) - R(x, y), k = 2' \\ R(x+1, y) - R(x, y), k = 3 \end{cases} \quad (18)$$

где переменная k используется для выбора направления, по которому берётся производная. Выражения (17) и (18) используются для расчета производных по направлению соответственно для эталонного и оцениваемого изображений. Затем полученные результаты интерпретируются путем сопоставления значений производных (19-21):

$$\Delta_b(x, y, k) = \begin{cases} |\delta_r(x, y, k) - \delta_s(x, y, k)|, \text{ если } |\delta_s(x, y, k)| < |\delta_r(x, y, k)| \ \& \ \text{sgn}(\delta_s(x, y, k)) = \text{sgn}(\delta_r(x, y, k)), \\ 0, \text{ в остальных случаях} \end{cases} \quad (19)$$

$$\Delta_d(x, y, k) = \begin{cases} |\delta_s(x, y, k) - \delta_r(x, y, k)|, \text{ если } |\delta_s(x, y, k)| > |\delta_r(x, y, k)| \ \& \ \text{sgn}(\delta_s(x, y, k)) = \text{sgn}(\delta_r(x, y, k)) \\ |\delta_s(x, y, k)|, \text{ если } \text{sgn}(\delta_s(x, y, k)) \neq \text{sgn}(\delta_r(x, y, k)) \\ 0, \text{ в остальных случаях} \end{cases} \quad (20)$$

$$\Delta_f(x, y, k) = \begin{cases} |\delta_r(x, y, k)|, \text{ если } \text{sgn}(\delta_s(x, y, k)) \neq \text{sgn}(\delta_r(x, y, k)) \\ 0, \text{ в остальных случаях} \end{cases} \quad (21)$$

$$\text{где } \text{sgn}(x) = \begin{cases} -1, \text{ если } x < 0 \\ 0, \text{ если } x = 0 \\ 1, \text{ если } x > 0 \end{cases} \text{ . Формулы (19)–(21)}$$

фактически представляют собой классификатор, позволяющий качественно и количественно оценить те изменения, которые произошли, прежде всего, с пикселями, имеющими достаточно большие значения производных, по крайней мере, по одному направлению. Если исходное изображение, подвергаемое компрессии с потерями, изначально не было сильно зашумлено, то такие пиксели, как правильно, находятся на границах объектов. Каким искажениям может быть подвергнута граница? Теоретически возможны четыре варианта (рис. 10):

- значения яркостной и/или цветовой компонент, соответствующих пикселям на границе, может быть или увеличены (вариант 1 – ложно усиленные границы), или уменьшены (вариант 2 – ложно ослабленные границы);

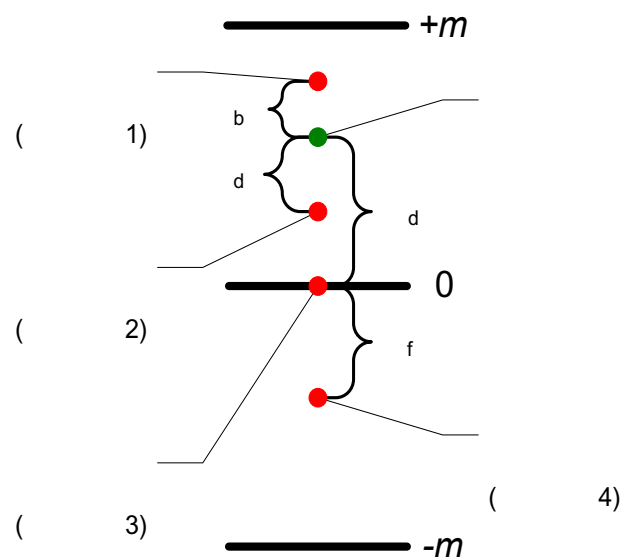


Рис. 10. Возможные варианты искажения пикселей, находящихся на границе объекта: физический смысл формул (19)–(21) [6, 11]

- граница может исчезнуть (вариант 3 – исчезающая граница), что, по сути, является частным случаем варианта 2, когда значение производной для данного пикселя по всем направлениям становится равным нулю ($\forall k \in [0,3]: \delta_R(x, y, k) = 0$);

- может появиться ложная граница (вариант 4).

Для определения степени влияния выше перечисленных эффектов и были введены 3 метрики, описываемые в виде формул (19) – (21):

$\Delta_b(x, y, k)$ и $\Delta_d(x, y, k)$ используются для оценки изменения амплитуды пикселя, если граница соответственно была ложно усилена или ложно ослаблена;

$\Delta_f(x, y, k)$ применяется для оценки изменения амплитуды пикселя при появлении ложной границы.

Очевидно, что метода расчета искажений для единичного пикселя не достаточно для того, чтобы оценить артефакты внутри области Ω (в частном случае, внутри скользящего окна) – требуется формула для вычисления интегральной метрики, которой, собственно говоря, и является AEF, определяемое как

$$AEF = \sum_{k=0}^K w_1(k) \cdot (D_f(k) + D_d(k) + D_b(k)), \quad (22)$$

где $D_b(k)$, $D_d(k)$ и $D_f(k)$ – значения интегральных величин (компонент AEF), оценивающих искажения, вызванные соответственно ложным усилением границ, ложных ослаблением границ и ложным появлением границ; $w_1(k)$ – весовой коэффициент, учитывающий направление, по которому была рассчитана производная. Как было сказано выше, для полноты информационной картины производные должны быть взяты, как минимум, в двух взаимно ортогональных направлениях, то есть $K \geq 1$. В случае, если используются производные по 4 направлениям, $K = 3$, то есть

$$AEF = \sum_{k=0}^3 w_1(k) \cdot (D_f(k) + D_d(k) + D_b(k)), \quad (23)$$

где $w_1(k) = \begin{cases} \frac{1}{\sqrt{2}}, k = \{0, 2\} \\ 1, k = \{1, 3\} \end{cases}$. Величины $D_b(k)$, $D_d(k)$ и $D_f(k)$ рассчитываются следующим образом:

$$D_b(k) = \frac{w_b(k)}{H \cdot W} \sum_{i=0}^N \sum_{j=0}^M \Delta_b^2(i, j, k), \quad (24)$$

$$D_d(k) = \frac{w_d(k)}{H \cdot W} \sum_{i=0}^N \sum_{j=0}^M \Delta_d^2(i, j, k), \quad (25)$$

$$D_f(k) = \frac{w_f(k)}{H \cdot W} \sum_{i=0}^N \sum_{j=0}^M \Delta_f^2(i, j, k), \quad (26)$$

где H и W – соответственно высота и ширина окна в пикселях, для которого и производится расчет значений; $w_b(k)$, $w_d(k)$ и $w_f(k)$ – коэффициенты, используемые для настройки веса каждого типа искажения. Подставляя (24)-(26) в (23), после выполнения элементарных математических преобразований получаем следующую формулу:

$$AEF = \sum_{k=0}^3 \frac{w_1(k)}{N \cdot M} \cdot \left(w_f(k) \sum_{i=0}^N \sum_{j=0}^M \Delta_f^2(i, j, k) + w_d(k) \sum_{i=0}^N \sum_{j=0}^M \Delta_d^2(i, j, k) + w_b(k) \sum_{i=0}^N \sum_{j=0}^M \Delta_b^2(i, j, k) \right). \quad (27)$$

На рис. 11 представлена геометрическая интерпретация формулы (27). Следует подчеркнуть, что масштаб каждой из координатных осей настраивается с помощью весовых коэффициентов $w_b(k)$, $w_d(k)$ и $w_f(k)$. Таким образом, PWAЕ и AEF представляют собой набор метрик, позволяющий оценить искажения как яркостных и цветовых компонент, так и границ объектов, что в совокупности с рассмотренной выше методикой многомасштабной обработки и позволяет гарантировать достоверность изображения.

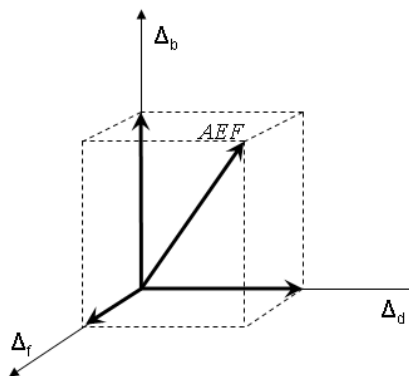


Рис. 11. Геометрическая интерпретация метрики AEF [6,11]

5. Экспериментальные результаты

На рис. 12б и 12в представлены увеличенные фрагменты изображений, декодированные в соответствии со стандартами JPEG и JPEG2000(JP2k) после сжатия с потерями. Как известно, стандарты кодирования JPEG и JPEG2000(JP2k), являющиеся самими популярными в области сжатия с потерями для статических изображений, радикально отличаются технологиями обработки видеоизображений и характеризуются существенно различающимися визуальными искажениями (артефактами).

Если для стандарта JPEG характерны блочные артефакты (англ. blocking artifacts, blockiness) по границам блоков 8x8, а также в несколько меньшей степени шумовые ореолы (англ. ringing artifacts) вокруг резких границ, являющиеся проявлениями эффекта Гиббса, то для JPEG2000 типично, в первую очередь, размытие мелких деталей (англ. blurring artifacts) и только во вторую – артефакты, обусловленные эффектом Гиббса. Следует заметить, что артефакты, вызванные блочной архитектурой JPEG кодека и эффектом Гиббса, приводят к ложному усилению существующих или к появлению ложных границ в изображении, в то время как размытие деталей проявляется в виде ложного ослабления или полного исчезновения границ.

Данные объективной оценки качества изображений, показанных на рис. 12б и 12в, представлены в табл. 1. Как видно, значения SSIM и MS-SSIM для каждого из сжатых изображений очень близки, а PSNR отличаются более, чем на 2 дБ. Как в таком случае можно оценить достоверность каждого из изображений? Чтобы ответить на этот вопрос, проведем анализ результатов, полученных при вычислении AEF.

На рис. 12б, сжатым JPEG-совместимым кодеком, доминируют искажения, обусловленные ложными ($D_f = 5.0204$) и ложно усиленными ($D_b = 6.7319$, $D_f + D_b = 11.7523$), а не ложно ослабленными ($D_d = 10.7087$) границами. Обратная ситуация наблюдается для рис. 12в, сжатого JPEG2000-совместимым кодеком: артефакты, вызванные ложно ослабленными ($D_d = 8.8564$) границами, превалируют над искажениями, обусловленными ложными ($D_f = 3.0519$) и ложно усиленными ($D_b = 4.1457$, $D_f + D_b = 7.1976$) границами. Всё это полностью согласуется с тех-

нологиями кодирования, используемыми в JPEG и JPEG2000, а также с приведенным выше описанием типовых визуальных артефактов, характерных для данных кодеков.

Принимая во внимание численные данные о компонентах AEF, можно сделать вывод о том, что достоверность изображения, показанного на рис. 12в выше, чем у изображения, представленного на рис. 12б. Это связано, прежде всего, с тем, что

- во-первых, абсолютные значения D_f , D_d и D_b в случае, показанном на рис. 12б, больше, чем для рис. 12в;

- во-вторых, доля ложных (D_f/D_d) и ложно усиленных (D_b/D_d) границ на рис. 12б больше, чем на рис. 12в, а это означает появление ложных объектов, которые в случае криминалистической экспертизы могут быть неверно интерпретированы. В свою очередь, ложно ослабленные границы означают снижение амплитуды границы (вплоть до её исчезновения), что, безусловно, негативно сказывается на достоверности видеосигнала, но всё-таки в меньшей степени, чем ложные границы.

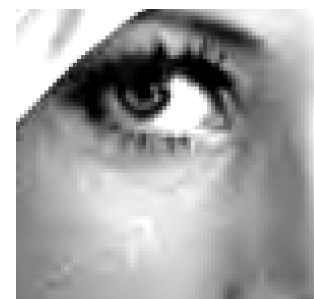
Другим примером, демонстрирующим эффективность предложенного подхода, является успешное определение того, что достоверность изображения на рис. 5в является значительно более низкой, чем у рис. 5б. Значения PWEA и компонент AEF представлены в табл. 2. Из этих данных следует, что максимальное искажение амплитуды яркости сигнала в случае рис. 5в превышает аналогичный показатель у рис. 5б более, чем в 2.6 раза. Отношения для величин D_f и D_b составляют соответственно почти 2.3 и 3.2 раза, что свидетельствует о значительном росте числа ложных и ложно усиленных границ у рис. 5в по сравнению с рис. 5б. Однако число ложно ослабленных пиксе-



а) Фрагмент исходного изображения Леппа



б) Фрагмент изображения, сжатого JTG кодером, совместимым со стандартом JPEG



в) Фрагмент изображения, сжатого Kakadu кодером, совместимым со стандартом JPEG2000 (JP2K)

Рис. 12. Увеличенные фрагменты изображения Леппа, два из которых подвергнуты сжатию с потерями, используя стандартизованные кодеки (JPEG и JPEG2000)

Таблица 1.

Значения различных метрик, используемых для оценок качества сжатых изображений, представленных на рис. 12

Номер рисунка	PSNR	SSIM	MS-SSIM	AEF				
				D_f	D_d	D_b	D_f/D_d	D_b/D_d
Рис. 14б	33.9230	0.9733	0.9853	5.0204	10.7087	6.7319	0.4688	0.6286
Рис. 14в	36.0920	0.9733	0.9860	3.0519	8.8564	4.1457	0.3446	0.4681

Таблица 2.

Значения различных метрик, используемых для оценок качества искаженных изображений, представленных на рис. 5

Номер рисунка	PWAE	AEF		
		D_f	D_d	D_b
Рис. 8б	54	2.0929	14.7946	1.5134
Рис. 8в	142	4.7552	13.0222	4.7883
Отношения значений метрик для рис. 8в к значениям у рис. 8б	2.6296	2.2721	0.8802	3.1639

лей на границах объектов меньше для рис. 5в, чем для рис. 5б. Подобные соотношения явным образом приводят к выводу о том, что изображение, представленное на рис. 5в, является очевидно недостоверным. Более подробная информация об экспериментальной верификации предложенного метода оценки качества может быть найдена в [6] и [11].

Заключение

В настоящей работе предложен метод оценки достоверности видеосигналов, позволяющий не только производить анализ изображений и видеоданных, но и адаптировать стандартные кодеки

(такие, как JPEG, JPEG 2000, H.264/AVC, H.265/HEVC) к требованиям систем безопасности, обеспечивая целостность и достоверность видеoinформации, что является одним из важнейших элементов кибербезопасности. Предложенный метод является вычислительно простым, что позволяет реализовывать его на стороне кодера в режиме реального времени. При этом изменения, вносимые в кодер, не являются нормативными, то есть сжатый поток остается совместимым со стандартом, что крайне важно в контексте тех усилий, которые прикладывает индустрия безопасности для унификации решений в области видеонаблюдения.

Рецензент: Цирлов Валентин Леонидович, кандидат технических наук, доцент МГТУ им. Н.Э. Баумана. E-mail: v.tsirlov@spno.ru

REFERENCE QUALITY ASSESSMENT OF THE DIGITIZED STATIC IMAGES AND VIDEO SIGNALS IN THE SECURITY SYSTEMS

Filippov A.K.³, Shakhtarin B.I.⁴

The paper considers the existing problems in assessing quality of the digitized static images and video-signals in the security systems to ensure such important elements of information security as accuracy and integrity. A set of modern metrics (PSNR, SSIM and MS-SSIM) and methods of reference quality assess-

3 Filippov Alexey, Ph.D., Associate Professor, Huawei Technologies Ltd., Moscow, aleks_filippov@mail.ru
 4 Shakhtarin Boris, Dr.Sc., Professor, Bauman Moscow State Technical University, Moscow, shakhtarin@mail.ru

ment, which are most widely spread in such industries as digital cinema and TV, videoconferences have been analyzed. We have reviewed a set of properties and specifics of the above metrics and methods which limit their use in the closed circuit television systems, also known as video surveillance systems. It provides a brief description of a generic mathematical model, in which video image is considered to be a combination of four components: smooth areas, contours, textures and noise. Based on this mathematical model, a new method of reference quality assessment of static images and video signals, which consists in using AEF and PVAE metrics, which are designed accordingly for assessment of distortion of borders and values of maximum distortion of brightness or chromacity. The paper provides a detailed theoretical substantiation of the suggested metrics, description of the methods of their use, and experimental findings, which show suitability of this method for accuracy assessment in the systems that use compression with the loss of information (firstly, in the video surveillance systems).

Keywords: lossy compression, still and moving pictures, video surveillance systems, forensics, full-reference quality assessment

Литература (References):

1. Sullivan G.J.; Wiegand T. Rate-distortion optimization for video compression // IEEE Signal Processing Magazine. 1998. №6 (15). P. 74-90.
2. Filippov A., Norkin A., Alvarez J.R. Video Codec Requirements and Evaluation Methodology // The IETF NETVC (Internet Video Codec) Working Group draft, draft-ietf-netvc-requirements-06, 2017, 25 p. URL: <https://datatracker.ietf.org/doc/draft-ietf-netvc-requirements/>
3. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing, 2004. №4 (13). P. 600-612.
4. Gonzalez R. C., Woods R.E. Digital Image Processing. Prentice Hall, 3rd edition, 2008, 976 p.
5. Iqbal M.I., Zepernick H.-J., Engelke U. Error sensitivity analysis for wireless JPEG2000 using perceptual quality metrics // The 2nd International Conference on Signal Processing and Communication Systems (ICSPCS 2008). 2008. P. 1-9.
6. Filippov A.K., Ruffitskiy V.A. Image Quality Assessment for Forensic Applications // Contribution WG1N5672 to the 54th JPEG Meeting (Tokyo, Japan), 2011.
7. Wang Z., Simoncelli E.P., Bovik A.C. Multi-scale Structural Similarity for Image Quality Assessment // The 37th Asilomar IEEE Conference on Signals, Systems and Computers. 2003. №2. P. 1398-1402.
8. Nicholson D. Image Quality metrics requirements for video-surveillance forensics // Contribution WG1N5482 to the 52nd JPEG meeting (Brussels, Belgium), 2010.
9. Sychev M., Filippov A., Sixin L. «Future Video Coding 4K test sequences evaluation report» // Contribution m37074 to the 113th MPEG meeting (Geneva, Switzerland), 2015.
10. Ran X., Farvardin N. A perceptually motivated three-component image model – Part I: Description of the Model // IEEE Transactions on Image Processing, 1995, №4 (4). P. 401-415.
11. Filippov A.K., Ruffitskiy V.A. A New Method of Image Quality Assessment for Video Surveillance and Forensic Applications // Proceedings of the 9th International Conference «Perspective technology in the mass media – PTMM'2011» (Vladimir, Russia, 2011), № 2. P. 21-24.

