УДК 621.327

ОЦЕНКА КОЛИЧЕСТВА ИНФОРМАЦИИ ИЗОБРАЖЕНИЯ ПО ЧИСЛУ СЕРИЙ ОДИНАКОВЫХ ЭЛЕМЕНТОВ

проф. А.В. Королёв, к.т.н. В.В. Баранник

Выводится выражение для определения количества информации в изображении с учетом ограничений на число серий одинаковых элементов. Приведены расчетные значения количества устраняемой избыточности в зависимости от вероятности появления элементов изображения.

Введение. Определение количества информации в изображении до и после кодирования по числу серий позволит сделать вывод о степени устраняемой избыточности, а, следовательно, о коэффициенте сжатия. При этом представляет интерес оценка информационной емкости изображения, элементы которого $\mathbf{a_{ij}}$ принадлежат двоичному множеству $\mathbf{a_{ij} \in \{0,1\}}$. Это могут быть черно - белые изображения или изображения, приведенные к двоичному виду (двухцветные).

Нахождение количество информации $\mathbf{H_0}$ в двухцветном изображении с учетом закона распределения элементов $\mathbf{a_{ij}}$. В общем случае количество избыточности $\mathbf{R_{01}}$ в изображении по информационному признаку находится по формуле [1]:

$$\mathbf{R_{01} = 1 - H_{01} / H_0}, \tag{1}$$

где $\mathbf{H_{01}}$ - количество информации в изображении по информационному признаку; $\mathbf{H_0}$ - количество информации с учетом вероятностей появления значений элементов изображений.

На основе формулы (1) для определения количества избыточности необходимо оценить количество информации $\mathbf{H_0}$ и $\mathbf{H_{01}}$.

Поскольку элементы $\mathbf{a_{ij}}$ входной последовательности принадлежат двоичному множеству, то в качестве статистической модели такого представления изображений используется Бернуллевский процесс [1]. Для таких процессов появление элемента изображения $\mathbf{a_{ij} = 0}$ и появление элемента изображения $\mathbf{a_{ij} = 1}$ являются событиями взаимонезависимыми с вероятностями, соответственно равными $\mathbf{q}$ и $\mathbf{p=(1-q)}$. В этом случае количество информации $\mathbf{H_0}$ в последовательности из $\mathbf{m}$ элементов двухцветного изображения без учета дополнительных ограничений находится по формуле

$$\mathbf{H_0 = -m\left(p\,\ell og_2\, p + q\,\ell og_2\, q\right)}. \tag{2}$$

Нахождение количества информации $\mathbf{H_{01}}$ в двухцветном изображении с учетом ограниченного числа серий одинаковых элементов. Введем случайную величину $\mathbf{N_j}$ - значение кода по числу серий одинаковых элементов. Тогда случайное событие $\mathbf{N_j = N}$ состоит в том, что случайная величина $\mathbf{N_j}$ будет равна значению $\mathbf{N}$ для заданного количества элементов изображения **m** (длина столбца $\mathbf{A^{(j)}}$ в фрагменте изображения) и конкретного значения числа серий $\mathbf{S}$:

$$\mathbf{N_j} = \varphi_{01}\left(\mathbf{m}, \mathbf{A^{(j)}}, \mathbf{S}\right) = \mathbf{N}\ , \tag{3}$$

где $\varphi_{01}(\bullet)$ - оператор вычисления значения $\mathbf{N}$ по заданному количеству элементов изображения **n** и известному числу серий $\mathbf{S}$ [2,3,4].

Поскольку для заданного значения **S** между величинами $\mathbf{N_j}$ и $\mathbf{A^{(j)}}$ существует взаимоднозначное соответствие, то вероятность появления события $\mathbf{N_j = N}$ будет равна вероятности появления конкретной последовательности $\mathbf{A^{(j)}}$ двоичных элементов изображения, равной $\mathbf{P}\left(\mathbf{A^{(j)}} = \mathbf{p^{m_1} q^{m_2}}\right)$:

$$\mathbf{P}\left(\mathbf{N_j = N}\right) = \mathbf{P}\left(\mathbf{A^{(j)}} = \mathbf{p^{m_1} q^{m_0}}\right), \tag{4}$$

где $\mathbf{P}\left(\mathbf{A^{(j)}} = \mathbf{p^{m_1} q^{m_0}}\right)$ - вероятность появления последовательности $\mathbf{A^{(j)}}$ элементов изображений, состоящей из $\mathbf{m_1}$ единиц и $\mathbf{m_0}$ нулей. Причем выполняется равенство $\mathbf{m = m_0 + m_1}$.

Для известных вероятностей случайной величины $\mathbf{N_j}$ количество информации $\mathbf{H_{01}}$ определяется как [1]:

$$\mathbf{H_{01}} = -\sum_{\mathbf{N_j = N_{min}}}^{\mathbf{N_{max}}} \mathbf{P}\left(\mathbf{N_j = N}\right) \ell\mathbf{og_2}\, \mathbf{P}\left(\mathbf{N_j = N}\right), \tag{5}$$

где $\mathbf{N_{min}}$, $\mathbf{N_{max}}$ - соответственно минимальное и максимальное значения величины $\mathbf{N_j}$, равные:

$$\mathbf{N_{min}} = \mathbf{0}\,; \quad \mathbf{N_{max}} = \mathbf{V}\left((\mathbf{n}+\mathbf{1})/\mathbf{4}, \mathbf{n}\right),$$

в которых $\mathbf{V}\left((\mathbf{n}+\mathbf{1})/\mathbf{4}, \mathbf{n}\right)$ - число комбинаций из $\mathbf{n}$ по $(\mathbf{n}+\mathbf{1})/\mathbf{4}$.

Чтобы в выражении (5) перейти от вероятностей $\mathbf{P}\left(\mathbf{N_j = N}\right)$ к вероятностям $\mathbf{P}\left(\mathbf{A^{(j)}} = \mathbf{p^{n_1} q^{n_0}}\right)$, необходимо учитывать случаи (если конкретное значение числа серий $\mathbf{S}$ не задано), когда для различных по составу последовательностей $\mathbf{A}^{(j)}$ могут формироваться одинаковые значения чисел $\mathbf{N}_j$. В соответствии с этим для $\mathbf{A}^{(1)} \neq \mathbf{A}^{(2)}$ выполняется условие

$$\mathbf{N}_1 = \mathbf{N}_2, \tag{6}$$

где $\mathbf{N}_1, \mathbf{N}_2$ - значения величины $\mathbf{N}_j$ для разных последовательностей $\mathbf{A}^{(1)} \neq \mathbf{A}^{(2)}$ (разных значений числа серий $\mathbf{S}=\vartheta_1$ и $\mathbf{S}=\vartheta_2$) элементов изображений.

В связи с замечанием (6) формулу (4) для вероятности $\mathbf{P}\left(\mathbf{N}_j = \mathbf{N}\right)$ нельзя использовать при вычислении количества информации $\mathbf{H}_{01}$. Поэтому в начале требуется просуммировать вероятности событий $\mathbf{A}^{(j)} = \mathbf{p}^{\mathbf{n}_1} \mathbf{q}^{\mathbf{n}_0}$ (согласно теоремы о сумме независимых событий [4]), для которых величина $\mathbf{N}_j$ будет иметь равные значения.

В соответствии с теоремой о непересекаемости классов комбинаций количество $\boldsymbol{\omega}$ одинаковых значений величины $\mathbf{N}_j$ определяется из неравенства

$$\mathbf{N}_j = \mathbf{N} \leq \left|\boldsymbol{\Omega}_s\right|, \tag{7}$$

где $\left|\boldsymbol{\Omega}_s\right|$ - количество элементов в множестве $\boldsymbol{\Omega}_s$, равное количеству комбинаций $\mathbf{V}(\mathbf{S}, \mathbf{n})$.

Отсюда для заданного $\mathbf{N}_j = \mathbf{N}$ число $\boldsymbol{\omega}$ равно количеству выполнения неравенства (7) по всем возможным значениям признака $\mathbf{S}$, $\left(0 \leq \mathbf{S} \leq \dfrac{\mathbf{n}+1}{2}\right)$. Тогда вероятность события $\mathbf{N}_j = \mathbf{N}$ по всем значениям признака $\mathbf{S}$ равна

$$\mathbf{P}\left(\mathbf{N}_j = \mathbf{N}\right) = \sum_{\mathbf{S}=0}^{\frac{\mathbf{n}+1}{2}} \mathbf{P}\left(\mathbf{A}^{(j)} \,\middle|\, \mathbf{N} \leq \mathbf{V}(\mathbf{S},\mathbf{n})\right), \tag{8}$$

где $\mathbf{P}\left(\mathbf{A}^{(j)} \,\middle|\, \mathbf{N} \leq \mathbf{V}(\mathbf{S},\mathbf{n})\right)$ - вероятность события $\mathbf{A}^{(j)} = \mathbf{p}^{\mathbf{n}_1} \mathbf{q}^{\mathbf{n}_0}$, при котором выполняется неравенство (7).

Учет неравенства (7) позволит при суммировании по всем значениям признака $\mathbf{S}$ выбрать вероятности только тех событий $\mathbf{A}^{(j)} = \mathbf{p}^{\mathbf{n}_1} \mathbf{q}^{\mathbf{n}_0}$, для которых величина $\mathbf{N}_j$ будет принимать одинаковые значения.

Подставив формулу (8) в (5), получим выражение для определения количества информации $\mathbf{H}_{01}$:

$$
\mathbf{H_{01}} = -\sum_{N_j=0}^{V\left(\frac{n+1}{4},n\right)} \left(\sum_{S=0}^{\frac{n+1}{2}} P\left(A^{(j)} \middle| N \le V(S,n) \right) \times \right.
$$

$$
\left. \times \ell og_2 \sum_{S=0}^{\frac{n+1}{2}} P\left(A^{(j)} \middle| N \le V(S,n) \right) \right). \tag{9}
$$

Таким образом, выражение (9) позволяет оценить количество информации $\mathbf{H_{01}}$ в двухцветном изображении (двоичном изображении) с учетом ограничений на число серий одинаковых элементов в зависимости от вероятности появления элементов изображения **p** и **q**. На основе выражений (1), (2) и (9) можно определить количество (выраженное в **%**) избыточности $\mathbf{R_{01}}$, исключаемое в результате формирования кодов по числу серий. Расчеты показывают, что при изменении вероятности **q** от **0.1** до **0.9** количество сокращаемой избыточности соответственно равно $\mathbf{50\,\% \le R_{01} \le 90\,\%}$.

Выводы. Получено выражение для определения количества информации с учетом ограниченного числа серий одинаковых элементов изображения. Результаты расчетов, проведенных по этому выражению, показали наличие в изображениях структурной избыточности. При этом избыточность, обусловленная чередованием областей, закрашенных одним цветом, достигает значения **90 %**.

ЛИТЕРАТУРА

1. *Шеннон К. Работы по теории информации и кибернетике. – М.: Изд - во иностр. лит - ры, 1963. – 793 с.*
2. *Бабкин В.Ф., Крюков А.Б., Штарьков Ю.М. Сжатие данных. – В кн.: Аппаратура для космических исследований. – М.: Наука,1972. – С. 172 - 209.*
3. *Прэтт У. Цифровая обработка изображений. – М.: Мир, 1982. – 790 с.*
4. *Корн Г. Корн Т. Справочник по математике. – М.: Наука, 1973. – 831 с.*

Поступила 28.03.2002

Королёв Анатолий Викторович, *Заслуженный изобретатель Украины, профессор, кандидат технических наук, профессор кафедры ХВУ. В 1969 году окончил Харьковское ВКИУ. Область научных интересов – обработка и передача информации.*

Баранник Владимир Викторович, *кандидат технических наук, научный сотрудник ХВУ. В 1994 году окончил Харьковский военный университет. Область научных интересов – обработка и передача информации.*