

УДК 681.325

Н.Н. Пономаренко

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

## КОРРЕЛЯЦИОННЫЙ КРИТЕРИЙ СОГЛАСИЯ ДЛЯ ОЦЕНКИ ГАУССОВОСТИ ВЫБОРОК ДАННЫХ В ЗАДАЧАХ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

*Предложен новый критерий согласия, основанный на вычислении коэффициента корреляции Пирсона между упорядоченной выборкой данных и упорядоченной среднестатистической выборкой данных для нормального закона распределения. Показано, что для задач обработки изображений, в которых анализ данных осуществляется в области спектральных коэффициентов дискретного косинусного преобразования, предложенный критерий существенно превосходит по эффективности наиболее часто используемый на практике критерий согласия  $\chi^2$ . Достоинством предложенного критерия также является инвариантность к дисперсии шума в выборке.*

**Ключевые слова:** критерий согласия, обработка изображений, дискретное косинусное преобразование.

### Введение

**Постановка проблемы.** Проверку заданной выборки данных на соответствие нормальному (Гауссовскому) закону распределения достаточно часто требуется осуществлять во многих задачах обработки сигналов, например, в задаче слепой оценки параметров шума на изображениях, когда из множества выборок данных, часть из которых содержит смесь шума и информационной составляющей, требуется выделить те выборки, которые содержат только нормальный шум.

Основной особенностью проверки выборки на гауссовость в задачах обработки изображений является достаточно малый размер выборок данных (от 25 до 256 отсчетов). При этом существующие критерии согласия, например, широко используемый критерий согласия  $\chi^2$  [1], плохо приспособлены для работы с выборками таких размеров. Кроме того, не существует доказательства оптимальности того или иного критерия согласия в общем виде. В зависимости от вида возможных негауссовых примесей в оцениваемых выборках лучшую эффективность может показывать тот или иной критерий. Поэтому актуальной является разработка новых критериев согласия, разработанных с учетом специфики задач обработки изображений.

**Анализ литературы.** В последнее время в задачах обработки изображений все большее применение находит использование дискретного косинусного преобразования (ДКП) для предварительной декорреляции значений пикселей изображения. Так, лучшую эффективность в сравнении с другими методами демонстрируют основанные на ДКП методы сжатия изображений с потерями [2], методы подавления шума на изображениях [3, 4], методы слепой оценки параметров шума [5, 6], методы реконструкции изображений [7], методы оценки визуального качества изображений [8] и др. При этом изображе-

ние обычно разбивается на блоки (чаще всего  $8 \times 8$  пикселей), для которых и выполняется ДКП. Чаще всего оценку гауссовости в этих задачах приходится выполнять для выборок именно такого размера (64 отсчета). В ряде случаев [2] могут использоваться и блоки другого размера.

Значения спектральных коэффициентов естественных (зарегистрированных с помощью тех или иных оптических приборов) изображений распределяются в соответствии с обобщенным гауссовским законом распределения [7, 9].

Поэтому критерий согласия должен позволять эффективно отличать выборки данных, содержащие гауссовский шум, от смеси гауссовского шума с информационной составляющей, распределенной по обобщенному гауссовскому закону.

Такие выборки характеризуются небольшим отличием формы распределения от гауссовского и при небольших размерах выборок плохо идентифицируются с помощью известных критериев согласия, таких как критерий согласия  $\chi^2$  [1] или критерий согласия Колмогорова [10].

**Цель исследования.** В данной работе предлагается новый критерий согласия, который может эффективно оценивать гауссовость небольших по размеру выборок коэффициентов ДКП изображений с возможной примесью информационной составляющей.

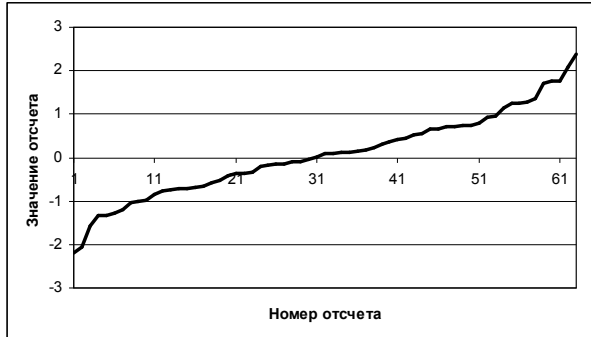
В подразделе 1 описывается алгоритм вычисления предлагаемого критерия.

В подразделе 2 рассматривается математическая модель смеси информационной и шумовой составляющей коэффициентов ДКП изображений и осуществляется сравнение эффективности предложенного критерия и критерия  $\chi^2$  для такой смеси.

В подразделе 3 показан пример использования предложенного критерия в задаче слепой оценки параметров шума.

### 1. Описание предлагаемого корреляционного критерия согласия

Пусть  $D$  – анализируемая выборка данных, состоящая из  $n$  элементов. Упорядочим эту выборку, например, по возрастанию значений и получим выборку  $D^o$  (рис. 1, а).



а



б

Рис. 1. Пример упорядоченной выборки данных (63 отсчета) с гауссовским шумом: а – реальная выборка  $D^o$ , б – среднестатистическая выборка  $I^o$

На рис. 1, б приведена “идеальная” упорядоченная выборка  $I^o$ , полученная усреднением большого числа (несколько миллионов) упорядоченных выборок нормального шума.

Идея состоит в использовании корреляции Пирсона [11] для оценки степени соответствия между  $D^o$  и  $I^o$ :

$$C(D^o, I^o) = \frac{\sum_{i=1}^n D_i^o I_i^o - n \bar{D}^o \bar{I}^o}{n \sigma_D \sigma_I}, \quad (1)$$

где  $\sigma_D$  и  $\sigma_I$  – соответственно СКО выборок  $D^o$  и  $I^o$ .

Выражение (1) далее будем называть корреляционным критерием согласия (ККС). Если значение ККС больше определенного порога, то гипотезу о гауссовости выборки  $D^o$  можно принять с соответствующим данному порогу уровнем значимости. Максимальным возможным значением ККС является единица, минимальным – минус единица.

Одним из преимуществ ККС является то, что значение выражения (1) инвариантно к дисперсиям выборок  $D^o$  и  $I^o$ . Поэтому для вычисления предло-

женного критерия согласия совсем не требуется знания дисперсии анализируемой выборки, как это требуется, например, при вычислении критерия  $\chi^2$ . Для небольшой по числу отсчетов выборки точное значение дисперсии оценить трудно, что уменьшает эффективность использования критериев, подобных критерию согласия  $\chi^2$ , но никак не влияет на точность вычисления предложенного ККС.

Еще одним из преимуществ вычисления ККС является простота и однозначность такого вычисления. Выражение (1) не содержит никаких параметров, которые нужно выбирать эмпирическим путем. В то же время при вычислении критерия согласия  $\chi^2$  требуется выбирать число и размеры интервалов, причем не существует математически обоснованного оптимального метода их выбора, что затрудняет практическое использование критерия  $\chi^2$  [12].

Для практического использования ККС нужно найти значения уровней значимости для различных сочетаний  $n$  и значения ККС. Это можно сделать табличным путем. В данной работе использовался следующий алгоритм:

1. Для каждого из значений длины выборки  $n$  от 5 до 256 генерировалось по 10000 выборок нормального шума. Для каждой выборки вычислялось значение ККС.

2. 10000 значений ККС для каждого значения  $n$  упорядочивались по возрастанию. Таким образом, для каждого значения  $n$  было получено 10000 значений ККС, соответствующих 10000 уровням значимости (от 0 до 1 с шагом 0,0001).

3. Полученная таблица сохранена в файл, который занимает на диске около 20 Мбайт (8 байт на одно значение ККС).

Подобная таблица может быть легко получена для любых других значений  $n$  (описанная процедура выполняется на персональном компьютере за несколько секунд). В табл. 1 приведены значения ККС для некоторых  $n$  и уровней значимости.

Таблица 1

Пороговые значения ККС для некоторых длин выборок и уровней значимости

n	Уровень значимости						
	0,01	0,05	0,1	0,5	0,9	0,95	0,99
5	0,8266	0,8806	0,9029	0,9596	0,9847	0,9881	0,9932
9	0,8715	0,9137	0,9298	0,9681	0,9859	0,9890	0,9939
16	0,9146	0,9441	0,9538	0,9777	0,9893	0,9913	0,9941
25	0,9414	0,9604	0,9675	0,9837	0,9917	0,9931	0,9953
49	0,9692	0,9776	0,9811	0,9902	0,9948	0,9957	0,9969
64	0,9753	0,9824	0,9854	0,9922	0,9957	0,9964	0,9973
256	0,9935	0,9951	0,9959	0,9977	0,9986	0,9988	0,9991

### 2. Оценка гауссовости коэффициентов ДКП блоков изображений

Известно, что коэффициенты ДКП для естественных фотографических изображений (кроме ко-

эффицентив с индексом  $[0,0]$ , соответствующих средним уровням блоков) распределены по обобщенному Гауссовскому закону распределения:

$$f(x, \mu, \sigma, \alpha) = \exp\left(-\left|\frac{x-\mu}{b}\right|^\alpha\right) a, \quad (2)$$

где  $a = \frac{1}{2\Gamma(1+1/\alpha)b}$ ;  $b = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$ ;  $\mu$  – математическое ожидание;  $\sigma$  – среднеквадратическое отклонение;  $\Gamma(\cdot)$  – гамма-функция;  $x$  – значение коэффициента ДКП. Значение  $\alpha=2$  соответствует нормальному закону распределения,  $\alpha=1$  соответствует Лапласовскому закону распределения.

В [7] показано, что для коэффициентов ДКП блоков изображений типичным является  $\alpha=0,5$ . На рис. 2 приведено обобщенное гауссовское распределение для  $\alpha=0,5$ , а на рис. 3 – гистограмма распределения коэффициентов ДКП реального тестового изображения Barbara.

В [7] показано, что для коэффициентов ДКП блоков изображений типичным является  $\alpha=0,5$ . На рис. 2 приведено обобщенное гауссовское распределение для  $\alpha=0,5$ , а на рис. 3 – гистограмма распределения коэффициентов ДКП реального тестового изображения Barbara.

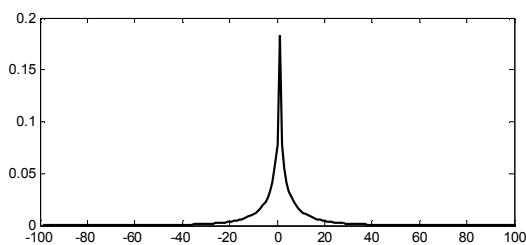


Рис. 2. Обобщенное Гауссовское распределение,  $\alpha=0,5, \mu=0, \sigma=15$

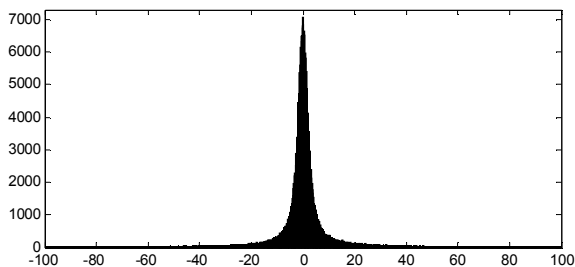


Рис. 3. Гистограмма значений коэффициентов ДКП блоков  $8 \times 8$  пикселей изображения Barbara

ДКП способно хорошо декоррелировать значения пикселей в локальных блоках изображения. При этом информационная составляющая обычно сосредотачивается в области низких коэффициентов ДКП (здесь может превалировать обобщенное гауссовское распределение с  $\alpha=0,5$ ), а в области высоких частот обычно сосредотачивается шумовая составляющая (с нормальным распределением) [5]. Существует множество способов предварительно отделить выборки данных, в которых превалирует шумовая составляющая от выборок с существенной примесью информационной составляющей [6], однако полностью исключить присутствие информационной составляющей невозможно.

В общем виде распределение значений коэффициентов ДКП в отобранных выборках с превали-

рующей шумовой составляющей можно представить как сумму гауссовского распределения с дисперсией  $\sigma_1^2$ , и обобщенного гауссовского распределения с  $\alpha=0,5$  и с дисперсией  $\sigma_2^2$ . Методы слепой оценки дисперсии шума в данном случае дадут результат, равный  $\sigma_1^2 + \sigma_2^2$ , что приемлемо только в случае, если  $\sigma_2^2$  не превышает 0,1 от  $\sigma_1^2$  [13]. Следовательно, критерий согласия должен позволять эффективно отличать смеси с  $\sigma_1^2 + \sigma_2^2 > 1,1 \sigma_1^2$  от нормального гауссовского распределения. Исследуем поведение критерия согласия  $\chi^2$  и ККС на таких смесях.

На рис. 4 приведен график зависимости среднего уровня значимости принятия гипотезы нормальности распределения от размера выборки для случая, когда выборка является обобщенным гауссовским распределением с  $\alpha=0,5$  ( $\sigma_1^2 = 0$ ).

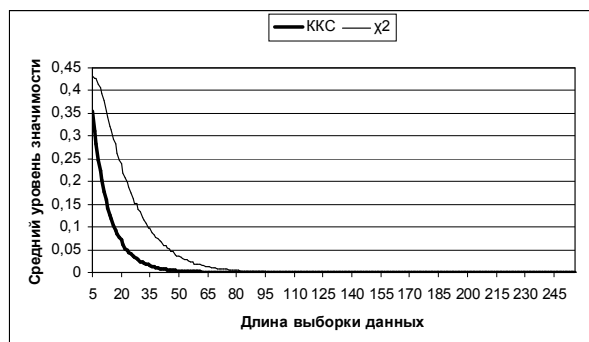


Рис. 4. Эффективность критериев согласия  $\chi^2$  и ККС для обобщенного гауссовского распределения

Как видно из рис. 4, оба критерия при длинах выборки больших 100 уверенно отвергают гипотезу о нормальности распределения. В то же время при меньших длинах выборки критерий ККС оказывается гораздо более эффективным, чем  $\chi^2$ .

На рис. 5 приведен график той же зависимости для более сложного случая смеси шумовой и информационной составляющей, когда  $\sigma_2^2 = \sigma_1^2$ .

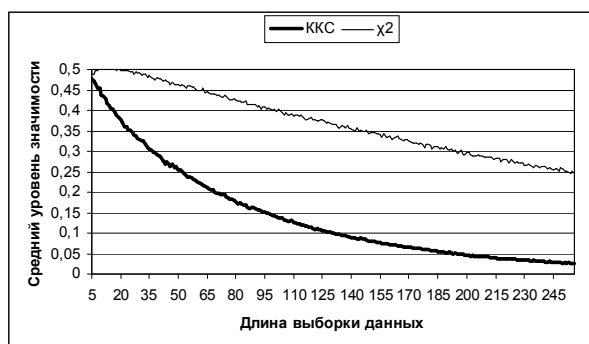


Рис. 5. Эффективность критериев согласия  $\chi^2$  и ККС для смеси  $\sigma_2^2 = \sigma_1^2$

При наборе статистики формировалось по 100000 выборок данных для каждого значения дли-

ны выборки. Присутствующая, несмотря на это, изрезанность графиков свидетельствует о существенной нестабильности значений критериев согласия для подобных смесей.

Как видно из графика на рис. 5, такая смесь информационной и шумовой составляющих является сложной для обоих критериев согласия, но, если критерий согласия  $\chi^2$  для всего диапазона длин выборок принимает гипотезу о нормальности распределения с уровнем значимости в среднем не ниже, чем 0,25, то ККС для больших длин выборок (>200) уже достаточно уверенно отвергает эту гипотезу.

На рис. 6 и 7 приведены соответственно графики для случаев с  $\sigma_2^2 = 0,5 \sigma_1^2$  и  $\sigma_2^2 = 0,2 \sigma_1^2$ .

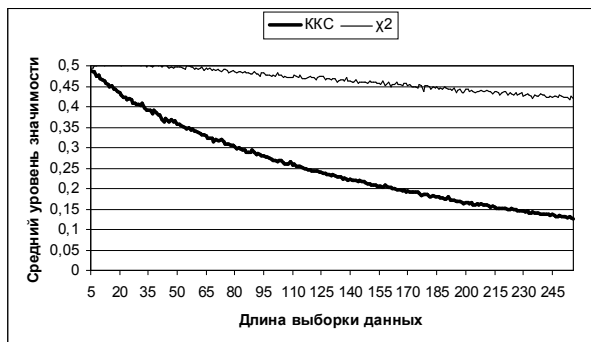


Рис. 6. Эффективность критериев согласия  $\chi^2$  и ККС для смеси  $\sigma_2^2 = 0,5 \sigma_1^2$

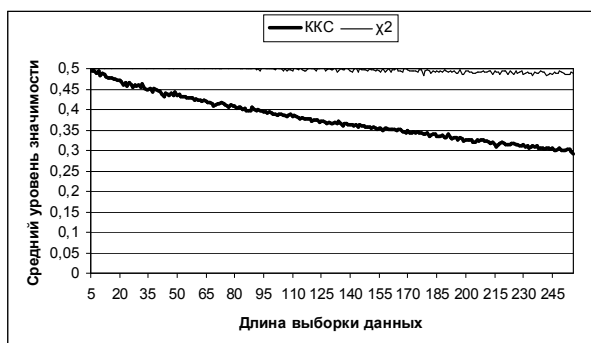


Рис. 7. Эффективность критериев согласия  $\chi^2$  и ККС для смеси  $\sigma_2^2 = 0,2 \sigma_1^2$

Видно, что в данных ситуациях, особенно для  $\sigma_2^2 = 0,2 \sigma_1^2$ , оба критерия не способны уверенно отвергнуть гипотезу о нормальности распределения, но ККС показывает существенно более высокую эффективность, чем критерий согласия  $\chi^2$ .

### 3. Пример использования ККС в задаче слепой оценки параметров шума

Покажем на простом примере, что использование ККС в задачах обработки изображений может обеспечивать лучший результат, чем использование критерия  $\chi^2$ .

Для заданного изображения в скользящем окне  $8 \times 8$  пикселей будем выполнять ДКП. Для каждого положения скользящего окна вычислим оценку  $\sigma_{ij}^2$ , где  $i, j$  – координаты верхнего левого угла окна, а также значение критерия согласия (отбросив коэффициенты с индексом  $[0,0]$ ). Далее, усредним те  $\sigma_{ij}^2$ , для которых значение критерия согласия превысит заданный порог (данные в данном окне можно считать нормальным шумом) и получим оценку дисперсии шума на изображении.

На рис. 8 приведено изображение Barbara, искаженное аддитивным гауссовским шумом с  $\sigma^2 = 100$ , а также карты значений критериев согласия  $\chi^2$  и ККС для него (более яркие пиксели соответствуют более высоким уровням значимости принятия гипотезы нормальности распределения коэффициентов ДКП в блоке изображения).

Из рис. 8 видно, что ККС позволяет отсеять по порогу гораздо больше информационных участков, чем критерий согласия  $\chi^2$ . Также на рис. 8, б и 8, в хорошо виден характер распределения значений критериев согласия на однородных участках изображения – это равновероятное случайное число в диапазоне от 0 до 1 (средняя вероятность – 0,5). Это обуславливает целесообразность использования какой-либо пространственной постобработки карт значений критериев, например одним из сохраняющих детали медианных фильтров [14].

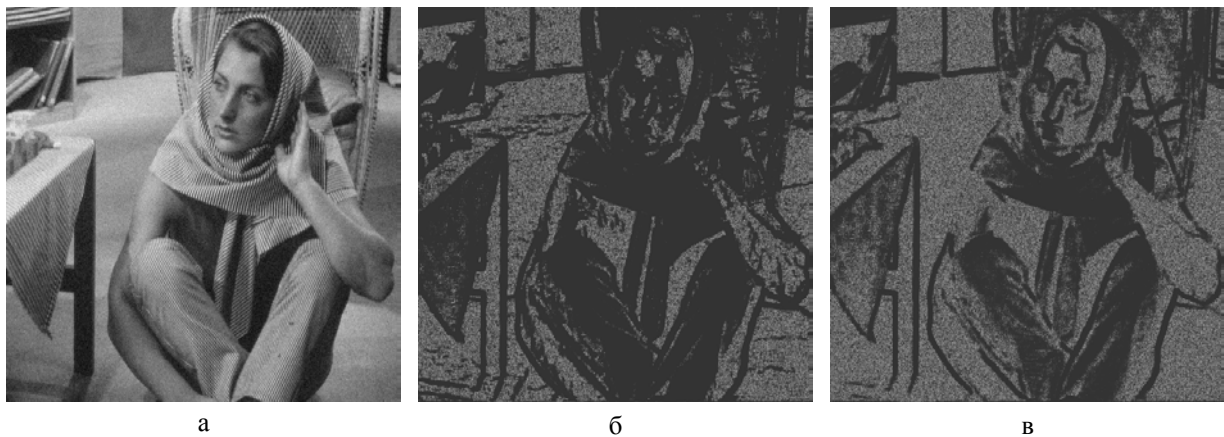


Рис. 8. Тестовое изображение Barbara и карты критериев согласия: а – Barbara,  $\sigma^2 = 100$ , б – карта значений ККС, в – карта значений  $\chi^2$

На рис. 9 приведены графики зависимости ошибки оценки дисперсии на изображении рис. 8, а от порога уровня значимости для обоих сравниваемых критериев согласия.

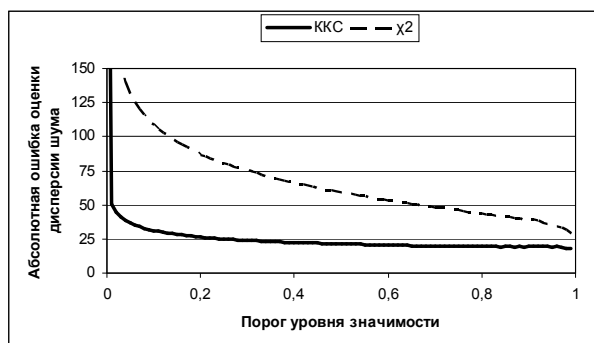


Рис. 9. Зависимость ошибки оценки дисперсии шума для изображения Barbara от уровня значимости

Видно, что использование ККС обеспечивает в полтора раза меньшую ошибку оценки дисперсии при близких к единице порогах уровня значимости, и в 2–3 раза меньшую ошибку при уровнях значимости 0,2..0,6.

На рис. 10 и 11 приведены аналогичные графики для изображений Lena и Baboon, искаженных шумом с такими же параметрами.

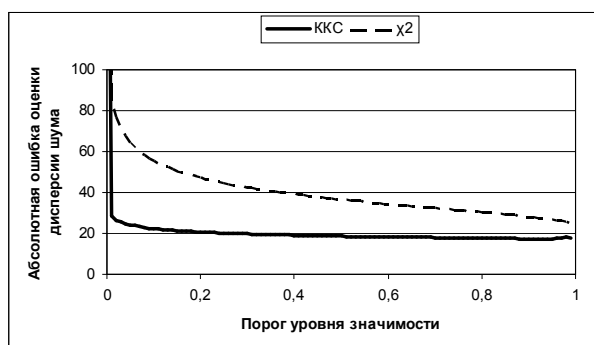


Рис. 10. Зависимость ошибки оценки дисперсии шума для изображения Lena от уровня значимости

Приведенные графики демонстрируют существенно более высокую эффективность предложенного ККС при анализе выборок коэффициентов ДКП блоков изображений.

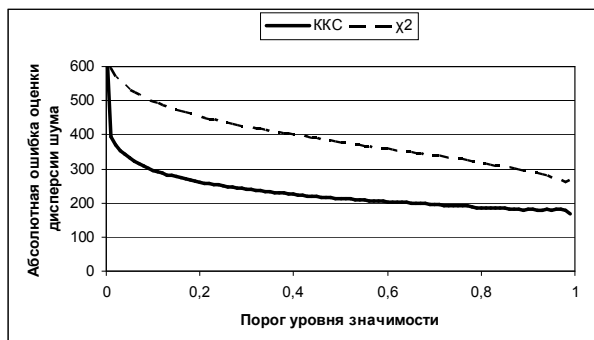


Рис. 11. Зависимость ошибки оценки дисперсии шума для изображения Baboon от уровня значимости

Набор использованных тестовых изображений Baboon, Barbara и Lena (все 512×512 пикселей) доступен для скачивания на [15].

## Заключение

В работе предложен новый корреляционный критерий согласия для оценки гауссовости выборок данных. Показана существенно большая эффективность ККС по сравнению с широко используемым критерием  $\chi^2$  при работе с выборками коэффициентов ДКП, что характерно для задач обработки изображений.

Предложенный критерий прост в использовании, так как не требует при его вычислении эмпирического выбора каких либо параметров. Кроме того, значение ККС никак не зависит от дисперсии выборки, что также облегчает его вычисление.

## Список литературы

1. Бендат Дж. Прикладной анализ случайных данных / Дж. Бендат, А. Пирсол. – М.: Мир, 1989. – 540 с.
2. Ponomarenko N. ADCTC: A new high quality DCT based coder for lossy image compression / N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola // Proceedings of LNLA. – Switzerland. – August 2008. – P. 33-40.
3. Lukin V.V. Image filtering based on discrete cosine transform / V.V. Lukin, R. Oktem, N. Ponomarenko, K. Egiazarian // Telecommunications and Radio Engineering. – 2007. – Vol. 66, No. 18. – P. 1685-1701.
4. Dabov K. Image denoising by sparse 3-D transform-domain collaborative filtering / K. Dabov, A. Foi, V. Katkovich, K. Egiazarian // IEEE Transactions On Image Processing. – 2007. – Vol. 16, issue 8. – P. 2080-2095.
5. An Automatic Approach to Lossy Compression of AVIRIS Images / N.N. Ponomarenko, V.V. Lukin, M.S. Zriakhov, A. Kaarna, J.T. Astola // IEEE International Geoscience and Remote Sensing Symposium, Barcelona Spain, 23-27 July 2007. – P. 472-475.
6. Пономаренко Н.Н. Метод слепой оценки на изображениях характеристик аддитивного Гауссова шума с неравномерным спектром / Н.Н. Пономаренко // Авиацонно-космическая техника и технология. – 2009. – № 2 (59). – С. 82-90.
7. Реконструкция значений утраченных точек изображений по энтропии коэффициентов ДКП. Информационные, телекоммуникационные и программные средства цифровой обработки сигналов / Н.Н. Пономаренко, В.В. Лукин, А.А. Зеленский, К. Егизарян, Я. Астола // Монография под ред. В.И. Марчука. – Шахты: Изд-во ЮРГУЭС. – 2008. – С. 5-24.
8. On between-coefficient contrast masking of DCT basis functions [Электронный ресурс] / N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin // Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics. – Scottsdale, USA, 25-26 January 2007. – 4 p. 1 электрон. опт. диск (CD-ROM); 12 см.
9. Hyvarinen A. Natural Image Statistics: A Probabilistic Approach to Early Computational Vision / A. Hyvarinen, J. Hurri, P.O. Hoyer // New York: Springer. – Computational Imaging and Vision. – 2009. – Vol. 39. – 448 p.
10. Большев Л.Н. Таблицы математической статистики / Л.Н. Большев, Н.В. Смирнов. – М.: Наука, 1983. – 416 с.
11. Кобзарь А.И. Прикладная математическая статистика / А.И. Кобзарь. – М.: Физматлит, 2006. – 816 с.

12. P 50.1.033-2001. Рекомендації по стандартизації. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа хи-квадрат. – М.: Изд-во стандартов, 2002. – 87 с.

13. Automatic Robust Procedure for Radar Image Preliminary Analysis and Filtering / V.V. Lukin, N.N. Ponomarenko, S.K. Abramov, A.A. Zelensky, K.O. Egiazarian, J.T. Astola // *Telecommunication and Radioengineering*. – Begell House (NY). – 2002. – Vol.58, no. 5-6. – P. 66-87.

14. Astola J. *Fundamentals of nonlinear digital filtering /*

J.Astola, P.Kuosmanen. – Boca Raton (USA): CRC Press LLC. – 1997. – 273 p.

15. Набор тестовых изображений [Электронный ресурс]. – 2009. – Режим доступа к ресурсу: <http://ponomarenko.info/testset1.zip>.

Поступила в редколлегию 16.02.2010

**Рецензент:** д-р техн. наук, проф. В.К. Волосюк, Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Харьков.

#### КОРЕЛЯЦІЙНИЙ КРИТЕРІЙ ЗГОДИ ДЛЯ ОЦІНЮВАННЯ ГАУСОВОСТІ ВИБІРОК ДАНИХ В ЗАДАЧАХ ОБРОБКИ ЗОБРАЖЕНЬ

М.М. Пономаренко

Запропоновано новий критерій згоди, який базується на обчислюванні коефіцієнту кореляції Пірсона між упорядкованою вибіркою даних та впорядкованою усередненою вибіркою даних для нормального закону розподілу. Показано, що для задач обробки зображень, у яких аналіз даних здійснюється у області спектральних коефіцієнтів дискретного косинусного перетворення, запропонований критерій суттєво перевищує за ефективністю критерій згоди  $\chi^2$ , який найбільш часто застосовується на практиці. Позитивною якістю запропонованого критерію є також інваріантність до дисперсії шуму у вибірці.

**Ключові слова:** критерій згоди, обробка зображень, дискретне косинусне перетворення.

#### CORRELATION FITTING CRITERION FOR ESTIMATE OF GAUSSIANTY OF DATA SAMPLE IN TASKS OF IMAGE PROCESSING

N.N. Ponomarenko

A new fitting criterion based on calculation of Pearson correlation between ordered data sample and ordered idealized data sample for normal distribution is proposed. It is shown that for image processing tasks where analyzing of data is carrying out in space of spectral coefficients of discrete cosine domain the proposed criterion considerably outperforms the fitting criterion  $\chi^2$  that is most often used in practice. An advantage of the proposed criterion is invariance to value of noise variance in data sample.

**Keywords:** fitting criterion, image processing, discrete cosine transform.