УДК 006.9:004.415.2(047)(476)

Е. Н. Савкова,
Е. Н. Карпиевич

МЕТРОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ РЕЗУЛЬТАТОВ МНОГОПАРАМЕТРИЧЕСКИХ ИЗМЕРЕНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Учитывая современные потребности в обеспечении надежности, сопоставимости и прослеживаемости результатов измерений, а также многообразие и различную природу величин, их корректное комплексирование в процессе измерений требует систематизации. Усложнение объектов измерений, в частности, увеличение количества и качества параметров для их корректного описания, приводит к необходимости разработки иерархии моделей от одномерных к многомерным с целью комплексирования данных параметров. С учетом развития цифровой техники и информационных технологий возникла необходимость в разработке информационной модели измерения, которая позволяет рационально использовать временные, финансовые ресурсы, исключить влияние человеческого фактора и повысить точность получаемой измерительной информации.

Given the current need to ensure the reliability, comparability and traceability of measurements, as well as the diversity and the different nature of the quantities of the correct aggregation in the measurement process requires systematization. The increasing complexity of the measurement object, in particular an increase in the quantity and quality parameters for a correct description of them, leads to the need to develop a hierarchy of models of one-dimensional to multidimensional data with the aim of interconnecting options. In view of the development of digital technology and information technology, it was necessary to develop information model of measurement that allows efficient use of time, financial resources, eliminate the human factor and increase the accuracy of the measurement information.

Введение

Создание, усложнение и повышение потенциальной опасности технических объектов обуславливают необходимость контроля их характеристик на всех стадиях жизненного цикла. С развитием цифровой техники и информационных технологий стало возможным применение перспективных методов контроля, основанных на цифровой регистрации обработки цифровых изображений. Однако в настоящее время такие методы основаны в основном на качественном анализе свойств объектов и оценивании прецизионности. Это связано с тем, что обработка данных в информационных системах имеет ряд особенностей:

1. Многообразие элементов технического и программного обеспечения, формирующих информационный канал, и, соответственно, многообразие их возможных сочетаний предполагает решение комбинаторных задач при отслеживании потерь точности.
2. Технические средства – сканеры, цифровые камеры, мониторы и т. д. – не являются средствами измерений и не имеют метрологических характеристик, характеризуются ограничениями разрешающей способности, динамических диапазонов и гамутов, что снижает достоверность результатов измерений.
3. Преобразования измерительной информации в информационном канале осуществляются с помощью встроенного программного обеспечения по различным стандартам, что затрудняет процесс поиска и анализа источников неопределенности.

Системы управления цветом позволяют оценивать фотометрические и колориметрические свойства объектов с помощью встроенных ранговых шкал, не имеющих «фиксированной» единицы. Измерение же как экспериментальное количественное определение свойств объектов предполагает наличие «основы для сравнения» [1] – ссылки на эталоны или стандартные образцы путем иерархии поверок или калибровок. Следовательно, для корректной интерпретации и использования количественных данных необходимо решить вопросы единства измерений за счет обеспечения метрологической прослеживаемости условных шкал, как предложено в работе [2], и оценивания методической составляющей неопределенности. Очевидно, что доказательной основой результатов измерений является их корректное описание, включающее оценивание неопределенности.

Величины, вовлеченные в процесс измерения

Применительно к измерениям авторами, как и в работе [5], предложено их условно разделить на три класса – неархимедовы, скалярные и многомерные (рис. 1).

Множество «неархимедовых» величин характеризуется тем, что для них не выполняется аксиома Архимеда-Евдокса. Преобразовать «неархимедовы» величины в другие виды величин невозможно [6]. Скалярные величины являются основным видом величин для количественного описания моделей свойств объектов в рамках основного уравнения измерений. Эти величины можно разделить на счетные, пропорциональные, аддитивные, интервальные и относительные. Многомерные величины могут быть двумерными, трехмерными и различной другой мерности. Для многомерных величин логическое соотношение «больше – меньше» в общем случае не имеет смысла. Операции сложения и умножения носят для них специфический характер. Так, сумма нескольких ненулевых векторов может быть равна нулю, а произведение векторов бывает скалярным и векторным [5]. Тензор определяется как геометрический объект, который описывается многомерным массивом, то есть набором чисел, занумерованных несколькими индексами (n-мерной таблицей, где n – валентность тензора). Так, вектор (тензор первого ранга) задается одномерным массивом (строкой или лучше – столбцом), а такие объекты, как линейный оператор и квадратичная форма, – двумерной матрицей. Скаляр же (тензор нулевого ранга) задается одним числом (которое можно рассматривать как нульмерный массив с единственным элементом). Скаляры и векторы удобно рассматривать в качестве частных случаев тензоров, так как все тензорные определения и теоремы для них в силе и векторы со скалярами можно при общем рассмотрении не упоминать отдельно [7].

Для каждого вида величин используют соответствующие шкалы. Многомерные шкалы могут быть образованы сочетанием шкал различных типов и в случае сложных свойств должны представлять собой сочетания информационных характеристик (размер, функция, функционал, оператор и др.) в соответствующих функциональных пространствах [8]. Согласно [1] величина – свойство явления, тела или вещества, которое может быть выражено количественно в виде числа с указанием отличительного признака как основы для сравнения. Согласно [4] многомерная случайная величина – упорядоченный набор (вектор) фиксированного числа N одномерных случайных величин. Основными числовыми характеристиками многомерной случайной величины являются вектор средних и ковариационная матрица. Для n-мерной величины X модель математических ожиданий (вектор средних) выглядит следующим образом:

$$X = (x_1, x_2, \ldots, x_N)^T, \qquad (1)$$

где x_1, x_2, …, x_N – оценки входных величин.

Таким образом, входные величины, вовлеченные в процесс измерения, могут иметь различную при-

Рис. 1. *Виды величин в измерениях*

роду, и важнейшим вопросом с точки зрения разработки метрологического обеспечения является нахождение их интервалов варьирования и удельных вкладов в изменчивость выходных величин.

В математической и графической интерпретации результат измерения трансформируется от точки, накрываемой интервалом охвата, к точке и плоскости охвата для двумерных величин, а для трехмерных – к точке и области охвата; для сложных моделей с неограниченным числом входных и выходных величин – к совокупности вектор-столбцов и ковариационных матриц, состоящих из ковариационных пар возрастающих как $2n$ к числу входных величин. При увеличении числа входных величин (размерности вектора) увеличивается размерность ковариационной матрицы, а, следовательно, увеличивается количество ковариационных пар. Так, например, для одномерной величины количество ковариаций равно единице, для двумерной – четырем, для трехмерной – девяти, для четырехмерной – шестнадцати и т. д. Таким образом, при увеличении мерности вектора, количество ковариаций увеличивается пропорционально квадрату согласно выражению n^2, где n – мерность. Определитель ковариационной матрицы является обобщенной дисперсией случайного вектора, который характеризует меру рассеивания случайного n-мерного вектора. При переходе от более простых к более сложным моделям усложняется соответственно и описание моделей результата измерения, в том числе и с точки зрения точностных характеристик мер положения и мер рассеяния. Так, для одномерных скалярных величин при прямых измерениях мера положения представляет собой сумму математического ожидания и поправок на погрешности, обусловленных различными источниками возникновения. При косвенных измерениях, являющихся следующей ступенью иерархии, математическое ожидание определяется исходя из функциональной зависимости входных и выходных величин. Многомерные же модели математических ожиданий включают в себя результаты прямых и косвенных измерений и описываются вектором средних. Что же касается моделей рассеяния, то они аналогично усложняются как по методу расчета, так и по их геометрическому представлению: от отрезка для скалярных величин до геометрического тела в пространстве (например, эллипсоид для трехмерной величины). Графическая интерпретация формирования результата измерения многомерной величины представлена на рисунке 2.

Таким образом, процесс моделирования результата измерения многомерной величины можно представить как технологию поэтапного перехода от прямых и косвенных измерений к многопараметрическим с неограниченным числом входных и выходных величин. Подмодель определенного уровня может быть разложена на более простые подмодели нижних уровней и, в свою очередь, является подмножеством моделей более высокого уровня.

Базовая модель. В качестве базовой модели результата измерения авторами предложено рассматривать механизм формирования интервала

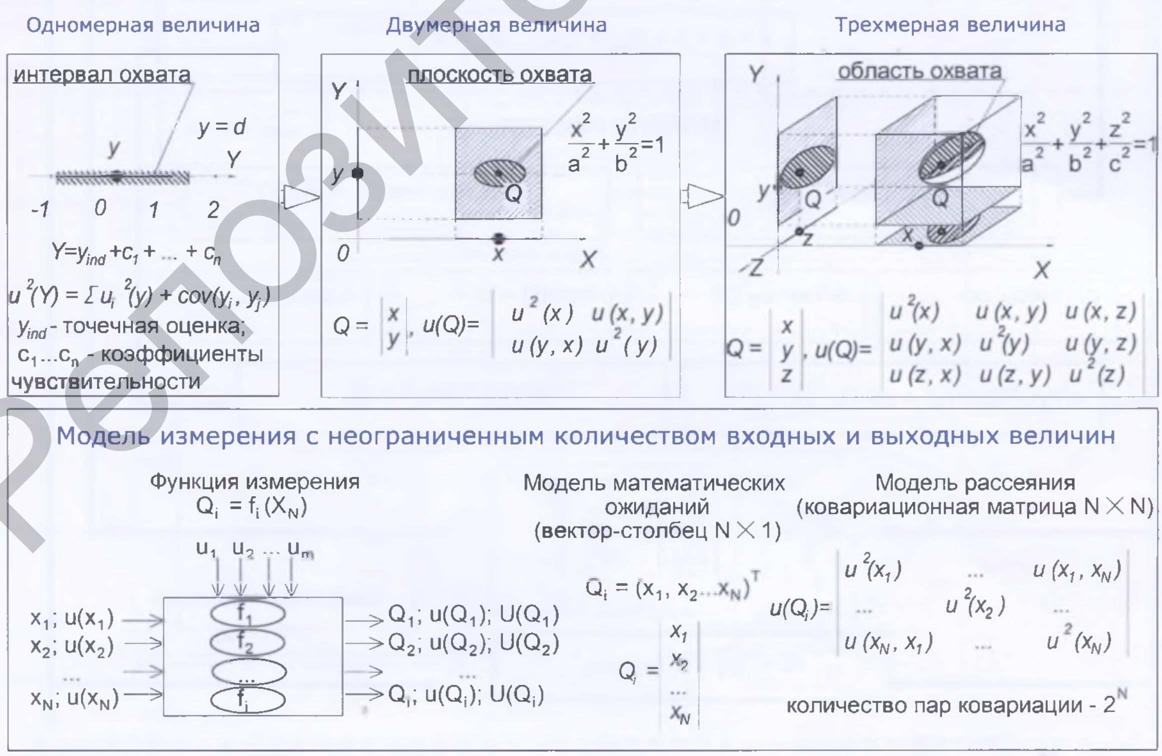

Рис. 2. *Принцип формирования результата измерения многомерной величины*

охвата скалярной величины, схематически показанный на рисунке 3. Учитывая, что в общем случае результат измерения представляет собой набор показателей: точечную оценку (среднее арифметическое, моду, медиану), суммарную или расширенную неопределенность, вероятность охвата, – он может быть описан моделями математических ожиданий (для определения точечной оценки выходной величины) и рассеяния (суммарной или расширенной неопределенностями). Предложенная базовая модель объединяет данные модели и основана на подходах, изложенных

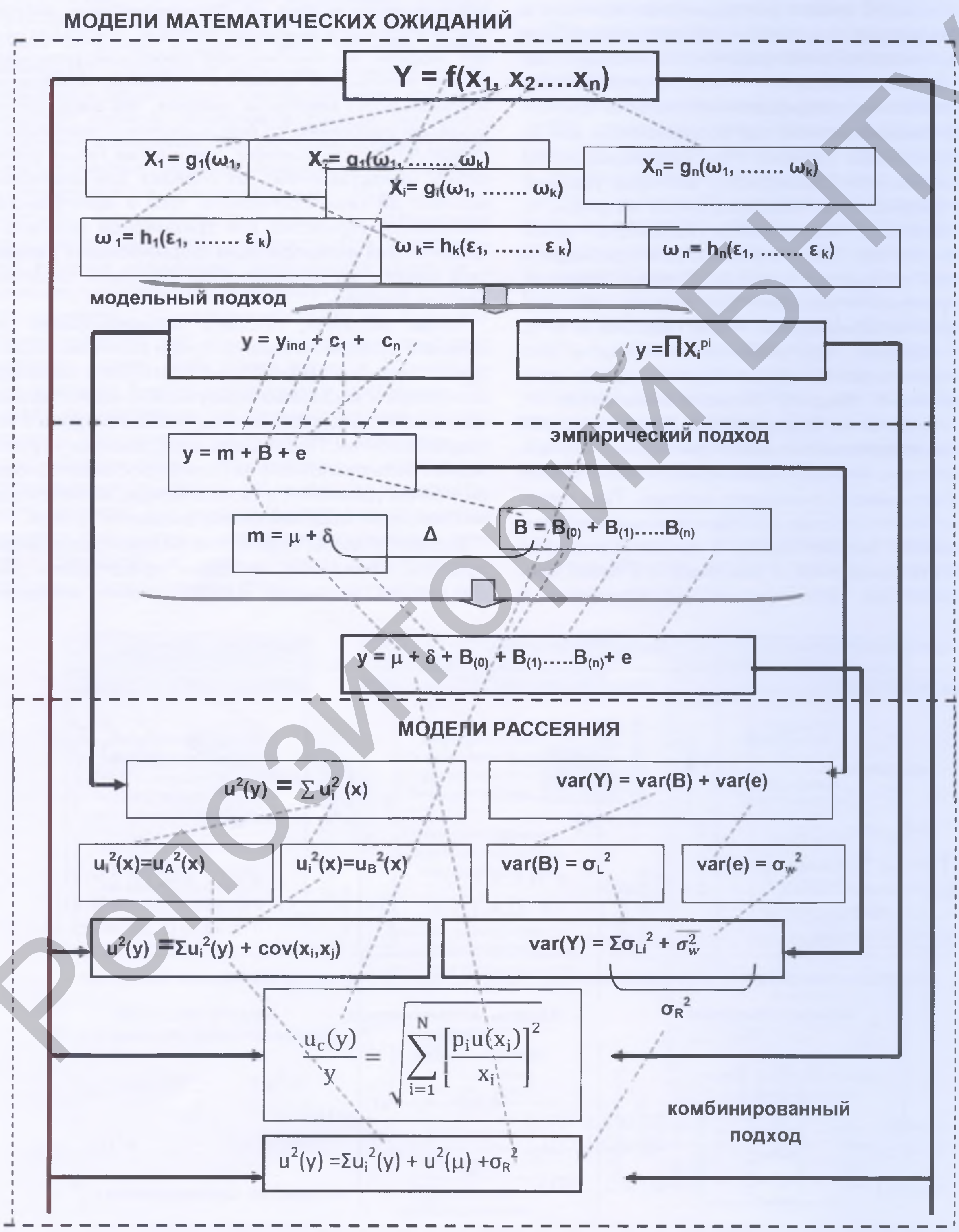

Рис. 3. *Базовая модель результата измерения скалярной величины*

в Руководстве по выражению неопределенности в измерениях [3] и ISO 5725. Она может строиться как по восходящему (комплексирование входных величин и их неопределенностей), так и по нисходящему (приведение косвенных измерений к прямым) принципам и упорядочивает процесс формирования интервала охвата результата измерения по заданному уровню достоверности.

«Фрактальная» модель. Модели, показанные на рисунках 2 и 3, являются частными случаями и могут рассматриваться как подмодели первого, второго и более высоких уровней при формировании результатов измерений в современных информационных системах. Для описания систем с неограниченным количеством входных и выходных величин, которые приводятся в ISO 17450-2 [9], авторами предлагается использовать так называемую «фрактальную» модель, графическая интерпретация которой приведена на рисунке 4. В основу модели положена идея фрактала: можно осуществлять разложение более «сложных» подмоделей на более «простые» (нисходящий подход) и комплексировать их в модели более высоких уровней практически без ограничений (восходящий подход).

При этом каждая величина может быть измеряемой (выходы $[Z_N; U(Z_N)]$; $[V_N; U(V_N)]$; $[T_N; U(T_N)]$; $[Q_N; U(Q_N)]$), а также может комплексироваться с другими величинами, поступая вместе со своими неопределенностями и ковариациями на входы подмоделей более высоких уровней (входы $[Z_N; u(Z_N)]$; $[V_N; u(V_N)]$; $[T_N; u(T_N)]$; $[Q_N; u(Q_N)]$).

Модель рассеяния многомерной величины описывается ковариационной матрицей [4]:

$$u_x = \begin{vmatrix} u(x_1, x_1) \ldots u(x_1, x_N) \\ \vdots \quad \ddots \quad \vdots \\ u(x_N, x_1) \ldots u(x_N, x_N) \end{vmatrix}, \qquad (2)$$

где $u^2(x_1)$ – дисперсия оценки x_1; $u(x_N, x_N)$ – ковариация.

Для многомерного пространства, помимо вектора средних и ковариационной матрицы, определяется и область охвата, которая содержит значение измеряемой величины с вероятностью охвата p.

Моделирование результата измерения в дискретных информационных системах

Энтропийная модель. Исследования вопросов обработки данных в дискретных передающих системах [10] позволили сделать вывод о том, что цифровое

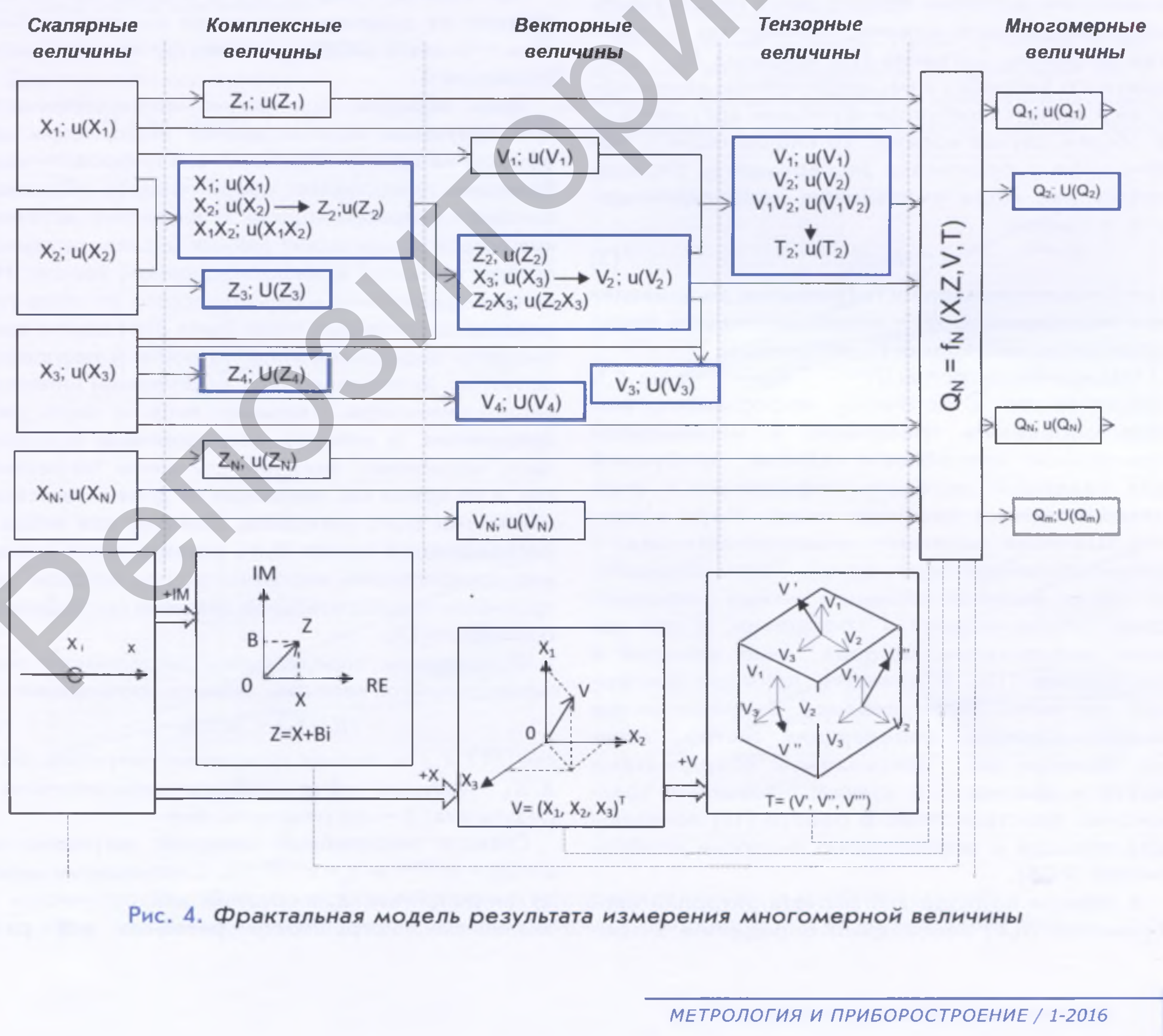

Рис. 4. Фрактальная модель результата измерения многомерной величины

изображение представляет собой информационную модель и строится на следующих положениях:

1. В общем случае цифровое изображение представляет собой модель реального или моделируемого объекта – двумерный первичный излучатель, состоящий из дискретных элементов (пикселей), каждый из которых имеет геометрические, яркостные и цветовые характеристики.
2. Каждая контрольная точка регистрируемого объекта представляет собой первичный или вторичный равнояркий излучатель и отображается на цифровом изображении группой пикселей размерностью $N \times M$.
3. Любой пиксель, принадлежащий выделенной области $N \times M$ цифрового изображения, имеет фотометрические и колориметрические характеристики, аналогичные другим элементам, принадлежащим данному массиву; и в пределах выделенной области яркость и цветность представляют собой многократно воспроизводимые величины.

Учитывая особенности обработки графических данных в информационных системах (необходимость выполнения операций фильтрации, дискретизации, квантования, кодирования и др.), для корректного описания результатов измерений авторами предлагается использовать информационную модель, основанную на понятии энтропии. Согласно [11] энтропия $H(p)$ дискретного ансамбля с распределением вероятностей p является выпуклой функцией аргумента p. В общем случае количество информации I, получаемое в результате эксперимента, определяется как мера уменьшения неопределенности, а именно:

$$I = H - H_0, \tag{3}$$

где H – неопределенность (энтропия) до проведения эксперимента, H_0 – неопределенность после проведения эксперимента (остаточная).

Измерение энтропии ($H = -p_1 \log_2 p_1 - \dots - p_n \log_2 p_n$), применяемое к источнику информации, может определить требования к минимальной пропускной способности канала, требуемой для надежной передачи информации в виде закодированных двоичных чисел. Мера энтропии Шеннона выражает неуверенность реализации случайной переменной. Таким образом, энтропия является разницей между информацией, содержащейся в сообщении, и той частью информации, которая точно известна в сообщении [11]. В качестве примера векторной величины, определяемой с применением информационных передающих систем, можно привести цвет, описываемый координатами цвета и цветности в стандартизованных трехмерных пространствах. В работе [12] показаны два подхода к определению энтропии изображений (RGB).

В первом подходе для расчета энтропии изображения $H(X)$ необходимо определить энтропию каждого из каналов изображения. Пусть вектор C – канал изображения X, $C = \{R, G, B\}$. Тогда энтропия канала изображения определяется по формуле Шеннона

$$H(C) = \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right), \tag{4}$$

где C – канал изображения X; p – вероятность, определяемая как частное от деления количества появлений i-го байта ($i = 0 \dots 255$) в канале изображения C к числу байт канала C изображения X.

Так как энтропия независимых источников равна сумме энтропий источников, то энтропия всего изображения $H(X)$ определяется как сумма энтропий каналов изображения

$$H(X) = \sum H(C). \tag{5}$$

При этом $C = \{R, G, B\}$.

В соответствии со вторым подходом энтропия изображения вычисляется по формуле Шеннона, однако вероятности определяются иным образом:

$$H(X) = \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right). \tag{6}$$

При этом p_i – вероятность, определяемая как частное от деления количества вхождений пикселя i-го цвета (RGB) к количеству пикселей изображения X.

Таким образом, оценивание неопределенности при получении количественной информации об объекте на основе обработки его цифрового изображения представляет собой процесс «распаковывания» «черного ящика» – выявление источников потерь графических данных (а следовательно, потерь точности) в информационном канале. На основе применения математического аппарата генераторов случайных чисел была определена взаимосвязь информационной энтропии и неопределенности. Установлено, что энтропийный интервал неопределенности охватывает лишь ту часть распределения, в которой сосредоточена основная часть возможных значений случайной погрешности, в то время как некоторая их доля остается за границами этого интервала. Поэтому для любого распределения может быть указано такое значение доверительной вероятности, при котором энтропийное и доверительное значения погрешности совпадают [13].

Формальным определением энтропийного значения случайной величины является соотношение

$$H(X/X_N) = \ln(2\Delta), \tag{7}$$

где $H(X/X_N)$ – полная остаточная энтропия, бит; X, X_N – текущие значения измеряемой величины и результата; Δ – погрешность, бит.

Отсюда энтропийный интервал погрешности: $d = 2\Delta = e^{H(X/X_N)}$ и $\Delta_j = e^{(H(X/X_N))}/2$. Соотношение между энтропийным Δ_j и средним квадратическим σ значениями погрешности различно для разных законов распределения, и его удобно характеризовать значением энтропийного коэффициента $k = \Delta j/\sigma$ данного закона распределения. Так, для равномерного распределения $\Delta_j = \Delta_m = \sqrt{3}\sigma \approx 1{,}73\sigma$ и, следовательно, $k = 1{,}73$. Для нормального распределения $\Delta_j = \sigma\sqrt{2\pi e}\ /2 = 2{,}066$ и $k = 2{,}066$. Для треугольного распределения Симпсона $k = 2{,}02$, для распределения Лапласа $k = 1{,}93$, для арксинусоидального распределения $k = 1{,}11$ и т. д.

Представленные выражения были расширены и адаптированы к понятию «неопределенность». Выражение, устанавливающее взаимосвязь между энтропией и неопределенностью, имеет вид

$$H = 2\ln U, U(H) = e^{H}/2, \qquad (7)$$

где H – полная остаточная энтропия, бит; $U(H)$ – энтропийный интервал неопределенности, бит. На основе данных о распределении авторами построены зависимости, отражающие взаимосвязь информационной энтропии и неопределенности для различных распределений вероятностей (рис. 5). Установлено, что наиболее точным является равномерное распределение:

$$H = \ln(2\sqrt{3}\sigma), U(H) = e^{2\sqrt{3}\sigma}/2. \qquad (8)$$

Распределение вероятностей по закону Симпсона описываются выражением

$$H = \ln(\sqrt{6e}\sigma), U(H) = e^{\sqrt{6e}\sigma}/2. \qquad (9)$$

Для нормального закона

$$H = \ln(\sqrt{2\pi e}\sigma),\ U(H) = e^{\sqrt{2\pi e}\sigma}/2. \qquad (10)$$

Применение теории информации с использованием энтропийного подхода является общим принципом, способом описания и оценки неопределенности результата измерений, пригодным для использования в равной степени как в метрических, так и неметрических шкалах. Авторами предложена графическая интерпретация энтропийной модели результата измерения в информационных системах, показанная на рисунке 6.

В данном случае на исходное изображение оказывает влияние внешняя среда (например, освещение), которая искажает исходное изображение. В дальнейшем эти искажения, являясь частью информации об изображении, проходят все стадии преобразования (предфильтрация, дискретизация, квантование, кодирование и т. д.). Кроме этого, потери в точности результата измерения вызывает и внутренняя среда (операции преобразования измерительной информации). Таким образом, процесс измерения цвета (изображения) с использованием информационных систем одновременно с изменением неотъемлемых параметров объекта (исходная энтропия H_0, координаты цветности $R_0G_0B_0$, яркость I_0) вводит и дополнительные параметры, идентификация и учет которых повышает точность измерительной информации.

Комплексная модель. Применительно к дискретным информационным передающим системам для описания результата измерения авторами предлагается ввести комплексную модель, объе-

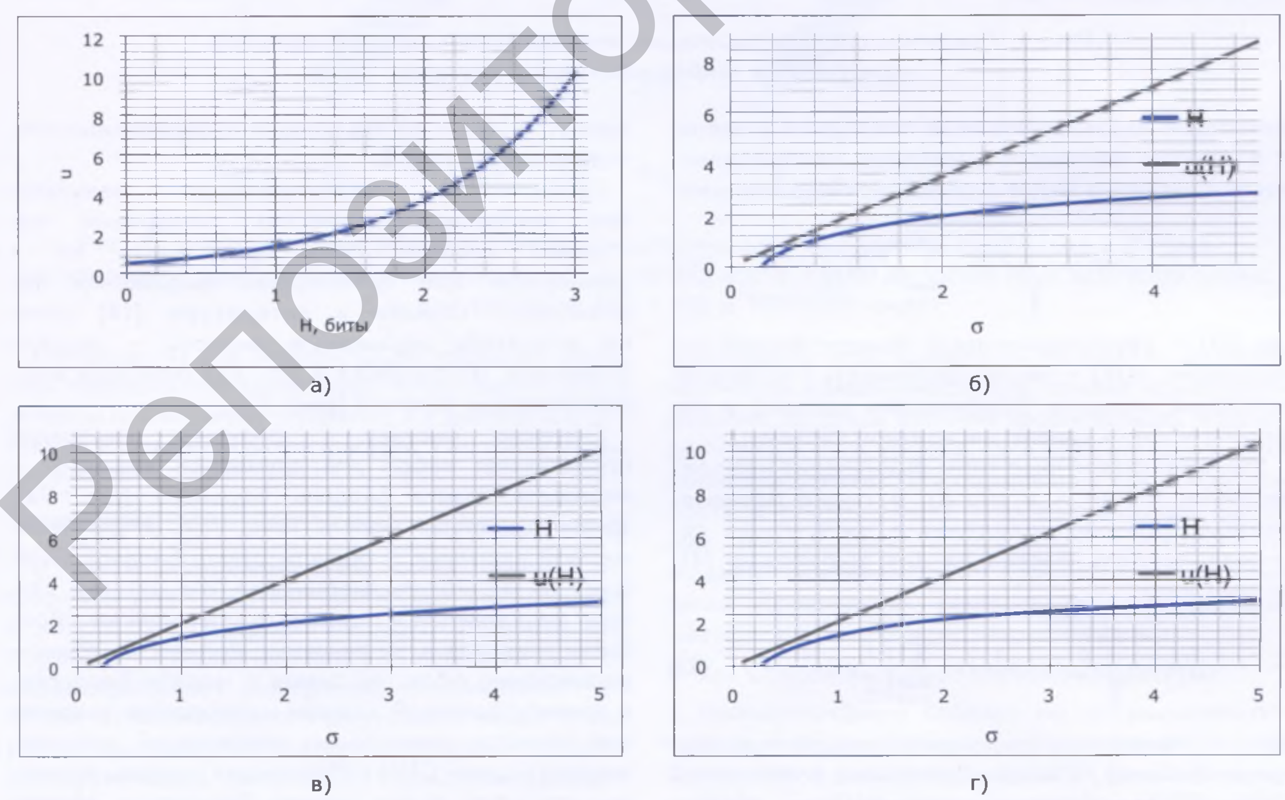

Рис. 5. Взаимосвязь энтропии и неопределенности: а — общий случай; б — для равномерного закона; в — для закона Симпсона; г — для нормального закона

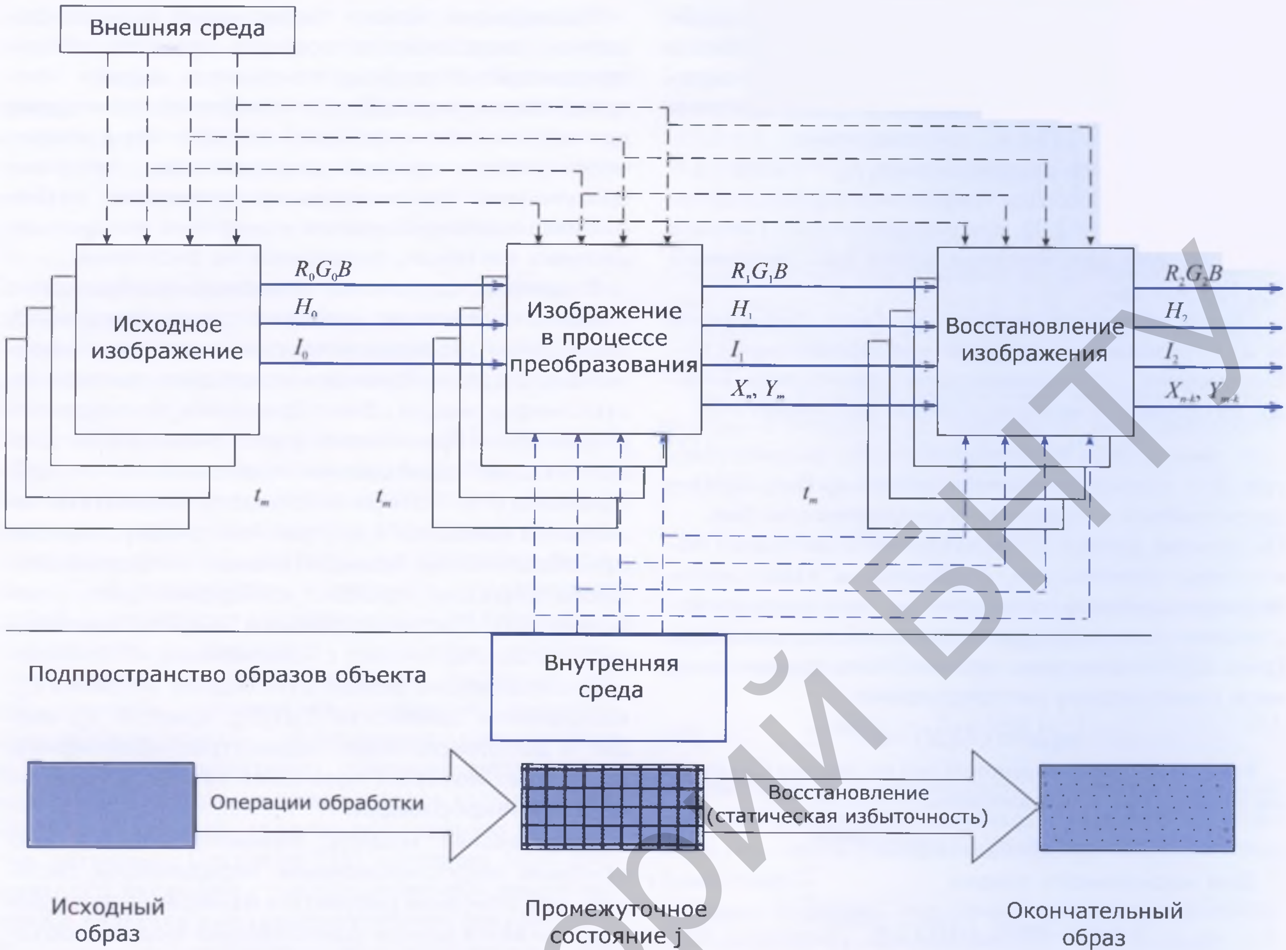

Рис. 6. *Графическая интерпретация энтропийной модели результата измерения в информационных системах*

диняющую вышеизложенные подходы и модели. И в данном контексте уравнения неопределенность измерения будет оцениваться выражением

$$u(H_0)=\sqrt{u^2(I)+\sum_{j=1}^{m}u^2(H_j)}\ , \qquad (11)$$

где $u(I)$ – неопределенность измерительной информации; $u(H_j)$ – неопределенность j-го преобразования сигналов (графических данных) в информационном канале дискретной системы.

Учитывая различные типы оценивания неопределенности (тип А и тип В), а также равновероятный вид распределения в дискретных системах, модель рассеяния для выражения (11) примет вид:

$$u(H_0)=\sqrt{u_A^2(I)+u_B^2(I)+\sum_{j=1}^{m}\left(\frac{1}{2}e^{2\sqrt{3}\sigma}\right)^2}\ , \qquad (12)$$

где σ – стандартное отклонение (стандартная неопределенность) j-й входной величины, вовлеченной в процесс преобразования информации; $u_A^2(I)$ – составляющая неопределенности, оцененная по типу А; $u_B^2(I)$ – составляющая неопределенности, оцененная по типу В.

Графические интерпретации результатов многопараметрических измерений настолько сложны, что представить их в двумерном или трехмерном выражении невозможно. Однако в литературе [14] можно встретить двумерные аналоги – модели Вороного, Фершильда и др. для описания коррелирующих величин (рис. 7).

Согласно рисунку 7 результат измерения представляет собой для двумерной величины – плоскость охвата (модель Вороного (рис. 7а), триангуляционная модель (рис. 7б)), а для трехмерной величины – области охвата (модель GUM (рис. 7в) [4], модель Фершильда (рис. 7г) [16]). Потери информации в результате операций ее обработки приводят к уменьшению количества точек и расширению областей охвата в модели Вороного, в триангуляционной модели наблюдается увеличение диаметра описываемых окружностей, для трехмерной модели GUM наблюдаются провалы отдельных секторов, а для модели Фершильда области неопределенности накладываются друг на друга.

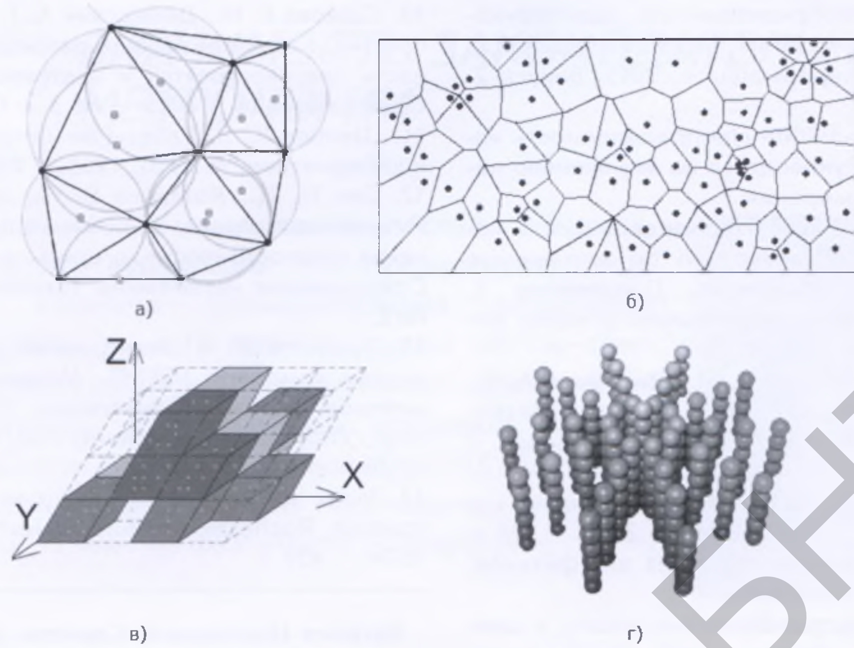

Рис. 7. *Графическая интерпретация энтропийной модели результата измерения в информационных системах: а — модель Вороного; б — триангуляционная модель; в — модель GUM; г — модель Фершильда*

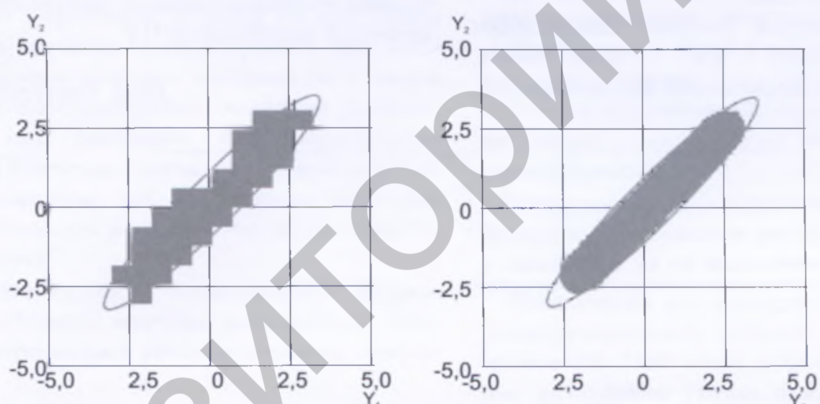

Рис. 8. *Эллиптические области охвата, построенные по процедуре разбиения в форме эллипсоида и приближения наименьших областей охвата: а — для выборки 10 × 10 и 1000 точек; б — для выборки 100 × 100 и 1000000 точек*

Согласно [4] для модели GUM проекции областей охвата на плоскости будут представлять собой эллипсы или параллелограммы, а в дискретных системах – более сложные фигуры, как показано на рисунке 8.

Заключение

В связи с применением современных информационных систем возникают потребности в разработке метрологического аппарата для обработки и моделирования результатов многопараметрических измерений многомерных величин. Установлено, что результаты таких измерений описываются вектор-матрицами (модели математических ожиданий) и областями охвата – набором ковариационных матриц (модели рассеяния). Комплексный подход к моделированию результатов измерений позволяет выявлять источники изменчивости, на основе комплексирования входных величин формировать подмодели различных уровней для корректной оценки областей охвата. Для оценивания неопределенности в современных информационных системах предложено использовать энтропийную модель, основанную на применении условных шкал измеряемых величин, обеспечивающих метрологическую прослеживаемость.

Список использованной литературы

1. Международный словарь по метрологии: основные и общие понятия и соответствующие термины: пер. с англ. и фр. /ВНИИМ им. Д. И. Менделеева, БелГИМ. Изд. 2-е, испр. – СПб.: НПО «Профессионал», 2010. – 82 с.

2. Савкова Е. Н. Неопределенность идентификации цвета в колориметрии высокого разрешения. / Системи обробки інформації. – 2015, Випуск 2. С. 96–99.
3. ISO/IECGuide 98-3:2008 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения.
4. ГОСТ Р 54500.3.2-2013 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения. Дополнение 2. Обобщение на случай произвольного числа выходных величин.
5. Артемьев Б. Г., Взоров В. И., Дмитриев А. В., Красивская М. И., Юрин А. И. О научном и техническом понятии величины// Главный метролог. – 2014. – № 2.
6. Хренников А. Ю. Неархимедов анализ и его приложения. – М.: ФИЗМАТЛИТ. – 2003. – 216 с.
7. Схоутен Я. А. Тензорный анализ для физиков. М.: Наука. – 1965 – 455 с.
8. Многомерные метрологические шкалы и операции измерений, контроля, испытаний, идентификации. Технические науки/12. Автоматизированные системы управления на производстве. – http://www.nesnauka.com.
9. ISO 17450-2 Geometrical product specifications (GPS) – General concepts – Part 2: Basic tenets, specifications, operators, uncertainties and ambiguities.
10. Савкова Е. Н., Демидович А. Г. Стандартизация понятий, характеризующих разрешающую способность цветовосприятий и цветовоспроизведения// Стандартизация. – 2015. – № 3. – С. 53–58.
11. Шеннон К. Э. Работы по теории информации и кибернетике. М.: ИЛ, 1963. – 829 с.
12. Сен Н. Д., Котляров В. П., Григорьев Я. Ю. Применение оценок на основе энтропии для сравнения криптостойкости алгоритмов шифрования.// Современные наукоемкие технологии. – 2013. – № 2.
13. Лукьянов В. Г. Электронный учебно-методический комплекс ТОИИТ. Модуль 2. Информационное описание измерения. Режим доступа: http://it.fitib.altstu.ru/neud/toiit/index.php?doc=teor&module=2
14. Марк Д. Фершильд. Модели цветового восприятия. Rochester Institute of Technology, USA. – 2004. – 439 с.

Евгения Николаевна Савкова, кандидат технических наук, доцент кафедры «Стандартизация, метрология и информационные системы» БНТУ;
Елена Николаевна Карпиевич, магистрант кафедры «Стандартизация, метрология и информационные системы» БНТУ

Дата поступления 18.11.2015