

эффекта, отсутствие достаточных условий его визуального проявления, отсутствие учета особенностей его проявления в реальных условиях [2-5] и откровенно ошибочные послышки [5] не дали возможности авторам [5] создать действительно эффективный метод для решения рассматриваемой задачи. Все вышесказанное заставило авторов настоящей работы вернуться к более детальному, чем это делалось до настоящего момента в открытой печати, изучению DQ-эффекта.

Целью работы является теоретическое обоснование достаточных условий проявления эффекта двойного квантования DCT-коэффициентов ЦИ, а также достаточных условий отсутствия DQ-эффекта, что должно дать теоретическую базу для последующего создания практического метода выявления фальсифицированного ЦИ, который бы позволил не только зафиксировать факт произошедшего несанкционированного изменения изображения, но и локализовать область такого изменения.

Достаточные условия проявления и отсутствия проявления эффекта двойного квантования коэффициентов DCT. Процесс двукратного использования JPEG-сжатия (процесс двойного квантования) ЦИ, которое первоначально стандартным образом разбивается на блоки малой размерности [1], например, 8×8 , далее рассматривается как процесс, состоящий из следующих шагов [3,5], выполняемых для каждого блока:

1. Первое квантование:

а) вычисление DCT-коэффициентов (значение конкретного коэффициента обозначим u);

б) DCT-коэффициенты квантуются при помощи массива нормализации [1], элементы которого далее называются коэффициентами квантования, и округляются до ближайшего целого, становясь равными

$$u^{(1)} = [u / q^{(1)}],$$

где $q^{(1)}$ - соответствующий u коэффициент квантования;

[•] - операция округления аргумента до ближайшего целого.

2. Первое восстановление:

а) DCT-коэффициенты, полученные на предыдущем шаге, умножаются на соответствующие им коэффициенты квантования:

$$u^{-(1)} = [u / q^{(1)}] q^{(1)};$$

б) вычисление обратного DCT.

3. Второе квантование:

а) вычисление DCT-коэффициентов;

б) DCT-коэффициенты квантуются при помощи массива нормализации, элементы которого отличны от соответствующих элементов на шаге 1, и округляются до ближайшего целого. В результате получаем:

$$u^{(2)} = \left[\left[\frac{u}{q^{(1)}} \right] \frac{q^{(1)}}{q^{(2)}} \right].$$

Пусть H и $H^{(2)}$ – гистограммы DCT-коэффициентов ЦИ, отвечающих выбранной произвольным образом частоте, до первого квантования и после второго соответственно. Они, очевидно, будут отличаться друг от друга. Рассмотрим некоторые из свойств получаемых гистограмм, важные с точки зрения выявления проделанной дважды операции JPEG-сжатия изображения.

Теорема 1. Пусть $\frac{q^{(2)}}{q^{(1)}} < 1$. Если

$$\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right) \in Z, \quad (1)$$

где Z - множество целых чисел, то столбец гистограммы $H^{(2)}$, отвечающий значению $u^{(2)}$, будет пустым.

Доказательство. Если

$$\left[\left[\frac{u}{q^{(1)}} \right] \frac{q^{(1)}}{q^{(2)}} \right] = u^{(2)}, \quad (3)$$

то

$$u^{(2)} - \frac{1}{2} \leq \left[\frac{u}{q^{(1)}} \right] \frac{q^{(1)}}{q^{(2)}} < u^{(2)} + \frac{1}{2},$$

$$\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} - \frac{1}{2} \right) \leq \left[\frac{u}{q^{(1)}} \right] < \frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right), \quad (2)$$

При этом

$$\left| \frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} - \frac{1}{2} \right) - \frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right) \right| = \frac{q^{(2)}}{q^{(1)}}, \quad (3)$$

По условию теоремы $\frac{q^{(2)}}{q^{(1)}} < 1$. Если выполнено соотношение (1), то,

очевидно, не существует такого $\left[\frac{u}{q^{(1)}} \right] \in Z$, которое бы удовлетворяло (2), а значит столбец гистограммы $H^{(2)}$, отвечающий значению $u^{(2)}$, будет пустым.

Следствие. Если $q^{(1)}$, $q^{(2)}$, $u^{(2)}$ удовлетворяют условию теоремы 1, то пустыми окажутся столбцы гистограммы $H^{(2)}$, отвечающие не только значению $u^{(2)}$, но и значениям $u^{(2)} + lT$, $l \in Z$, где T – период появления пустого столбца $H^{(2)}$ за счет выполнения условия (1), определяемый в соответствии с формулой:

$$T = \frac{q^{(1)}}{\text{НОД}(q^{(1)}, q^{(2)})}, \quad (4)$$

где $\text{НОД}(q^{(1)}, q^{(2)})$ - наибольший общий делитель $q^{(1)}, q^{(2)}$.

Доказательство. Пусть $n = \text{НОД}(q^{(1)}, q^{(2)})$, тогда $q^{(1)} = nk$, $q^{(2)} = nm$, где k, m некоторые натуральные числа, причем $\text{НОД}(k, m) = 1$. В принятых обозначениях $T = k$. Рассмотрим значение

$$\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + lT + \frac{1}{2} \right) = \frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right) + \frac{q^{(2)}}{q^{(1)}} lT, \quad (5)$$

Поскольку первое слагаемое в правой части формулы (5) является целым по условию теоремы, второе –

$$\frac{q^{(2)}}{q^{(1)}} lT = \frac{m}{k} lk = ml \in Z,$$

то

$$\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + lT + \frac{1}{2} \right) \in Z,$$

а значит, соответствующие столбцы гистограммы $H^{(2)}$ будут пусты.

Возможность проявления ситуации, рассмотренной в теореме 1, никак не учтена в [5], что привело к ошибочным теоретическим расчетам и несоответствию между ними и практическими результатами.

Замечание 1. Пусть $\frac{q^{(2)}}{q^{(1)}} > 1$, тогда противоречия с существованием

$\left[\frac{u}{q^{(1)}} \right] \in Z$, удовлетворяющего (2), на данном этапе не возникает, т.к. на

полуинтервале $\left[\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} - \frac{1}{2} \right), \frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right) \right)$, длина которого больше 1 в соответствии с (3), найдется хотя бы одно целое число.

Далее предполагаем, что соотношение (2) не является противоречивым, т.е. в случае $q^{(2)} < q^{(1)}$ условие (1) не имеет места. Тогда можно показать [5], что число столбцов $n(u^{(2)})$ гистограммы H , внесших свой вклад в столбец $u^{(2)}$ гистограммы $H^{(2)}$, определяется как

$$n(u^{(2)}) = q^{(1)} \left(\left[\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} + \frac{1}{2} \right) \right] - \left[\frac{q^{(2)}}{q^{(1)}} \left(u^{(2)} - \frac{1}{2} \right) \right] + 1 \right), \quad (6)$$

где $\lceil x \rceil = \min\{n \in Z \mid x \leq n\}$, $\lfloor x \rfloor = \max\{n \in Z \mid x \geq n\}$,

$n(u^{(2)})$ - периодическая функция с периодом (4), которая не определена для $u^{(2)}$, удовлетворяющих условию (1). Периодичность функции $n(u^{(2)})$, как правило, приводит к видимым особенностям гистограмм ДСТ-коэффициентов ЦИ после двукратного использования JPEG-сжатия с разными коэффициентами квантования. Эти особенности в общем случае можно охарактеризовать как периодическое возникновение пиков и впадин, называемое DQ-эффектом (рис.1). Однако такая картина имеет место не всегда. Очевидно, упомянутую периодичность, независимо от значений $q^{(1)}, q^{(2)}$, могут нарушить особенности гистограммы ДСТ-коэффициентов исходного изображения. Возможны и другие варианты.

Замечание 2. Проявление DQ-эффекта в виде периодических пиков и впадин гистограммы $H^{(2)}$ будет отсутствовать, если $q^{(2)} / q^{(1)} = 2m$, $m \in Z$. Действительно, в этом случае в (6) $n(u^{(2)}) = q^{(1)}(2m + 1)$ для любого $u^{(2)}$.

Установим достаточные условия проявления эффекта двойного квантования, одно из которых нашло свое выражение в теореме 1.

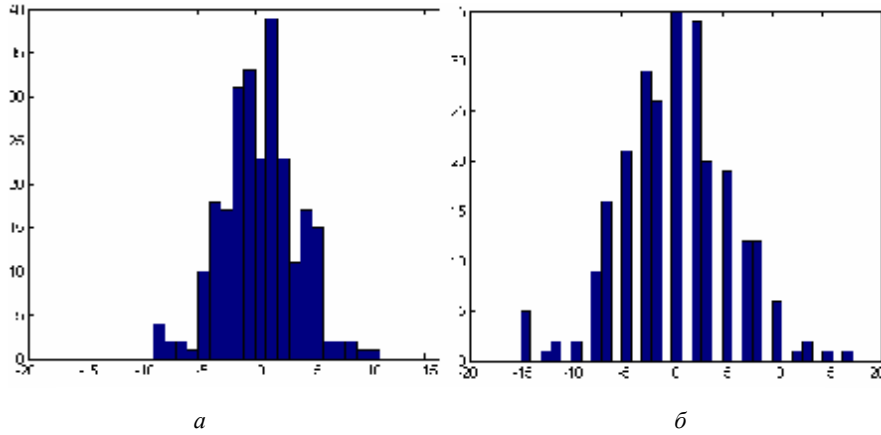


Рис. 1. Гистограммы DCT-коэффициентов частоты (2,2) для ЦИ, подвергнутого процессу двойного квантования с коэффициентами квантования:

$$a - q^{(1)} = 3, q^{(2)} = 5; \quad b - q^{(1)} = 5, q^{(2)} = 3$$

Теорема 2. Пусть $q^{(1)} = q^{(2)}$. Достаточным условием периодического возникновения незаполненных столбцов на гистограмме $H^{(2)}$ ЦИ (рис.1, б) является

$$\frac{q^{(1)}}{q^{(2)}} > \frac{3}{2}, \quad (7)$$

Доказательство. Поскольку для функции $n(u^{(2)})$ имеет место соотношение (4), то $q^{(1)}$ является ее периодом наряду с T . Рассмотрим значения $n(q^{(1)} - k)$, где $0 < k \leq q^{(1)}$, $k \in Z$:

$$\begin{aligned} n(q^{(1)} - k) &= q^{(1)} \left(\left\lfloor \frac{q^{(2)}}{q^{(1)}} \left(q^{(1)} - k + \frac{1}{2} \right) \right\rfloor - \left\lfloor \frac{q^{(2)}}{q^{(1)}} \left(q^{(1)} - k - \frac{1}{2} \right) \right\rfloor + 1 \right) = \\ &= q^{(1)} \left(\left\lfloor q^{(2)} - \frac{2k-1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor - \left\lfloor q^{(2)} - \frac{2k+1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor + 1 \right) = \\ &= q^{(1)} \left(\left\lfloor -\frac{2k-1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor - \left\lfloor -\frac{2k+1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor + 1 \right) \end{aligned} \quad (8)$$

Если выполнено условие (7), то

$$\frac{2q^{(1)} - q^{(2)}}{2q^{(2)}} > 1, \quad (9)$$

Возьмем k , удовлетворяющее неравенству

$$1 \leq k < \frac{2q^{(1)} - q^{(2)}}{2q^{(2)}}, \quad (10)$$

В силу (9) хотя бы одно такое значение k существует, при этом

$$\begin{aligned} 0 < \frac{2k+1}{2} \frac{q^{(2)}}{q^{(1)}} < 1, \quad \left\lfloor -\frac{2k+1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor = 0 \\ 0 < \frac{2k-1}{2} \frac{q^{(2)}}{q^{(1)}} < \frac{2k+1}{2} \frac{q^{(2)}}{q^{(1)}} < 1, \quad \left\lfloor -\frac{2k-1}{2} \frac{q^{(2)}}{q^{(1)}} \right\rfloor = -1. \end{aligned}$$

Таким образом, в условиях теоремы существуют такие значения $0 < k \leq q^{(1)}$, $k \in Z$, легко определяемые из соотношения (10), при которых правая часть (8) обращается в ноль, что приводит к незаполненности столбца (столбцов) $q^{(1)} - k$ в гистограмме $H^{(2)}$.

Следствие. В условиях теоремы $n(q^{(1)} - 1) = 0$ при любом значении $q^{(1)}$.

Замечание 3. Условие (7) является достаточным, но не необходимым, для того, чтобы при $q^{(1)} > q^{(2)}$ DQ-эффект проявился в наличии незаполненных столбцов в гистограмме $H^{(2)}$. Например, при $q^{(1)} = 4$, $q^{(2)} = 3$ в соответствии с (6) значение $n(2) = 0$, хотя $\frac{q^{(1)}}{q^{(2)}} < \frac{3}{2}$.

Замечание 4. Пусть $u^{(2)} = 0$. При $q^{(1)} > q^{(2)}$, независимо от их конкретных значений, $\frac{q^{(2)}}{2q^{(1)}} \notin Z$. В соответствии с (6):

$$n(0) = q^{(1)} \left(\left\lfloor \frac{q^{(2)}}{2q^{(1)}} \right\rfloor - \left\lfloor -\frac{q^{(2)}}{2q^{(1)}} \right\rfloor + 1 \right) = q^{(1)} \left(2 \left\lfloor \frac{q^{(2)}}{2q^{(1)}} \right\rfloor + 1 \right), \quad (11)$$

Из (11) вытекает, что для произвольных $q^{(1)}$, $q^{(2)}$ значение $n(0) \neq 0$, а значит, $n(lT) \neq 0$, $l \in Z$.

В любом случае, зная значения коэффициентов квантования $q^{(1)}$, $q^{(2)}$, рассмотрев функцию $n(u^{(2)})$ для аргументов $u^{(2)} = \overline{0, T-1}$ (а в силу замечания 4 достаточно рассмотреть для $u^{(2)} = \overline{1, T-1}$), конкретный вид гистограм-

мы $H^{(2)}$, т.е. наличие или отсутствие в ней периодических артефактов, можно определить априори, не производя непосредственного построения $H^{(2)}$.

Особенности проявления DQ-эффекта при реальном восстановлении изображения. Не ограничивая общность рассуждений, для простоты изложения далее рассматриваются изображения в градациях серого.

Все вышесказанное касалось проявления DQ-эффекта в «идеальных» условиях: учитывались лишь те округления при вычислениях, которые непосредственно проводились для DCT-коэффициентов. Однако реальный процесс восстановления ЦИ после первого квантования (процесс полного восстановления) включает в себя введение значений яркости пикселей после обратного DCT в диапазон $[0, 255]$, что приводит к возникновению дополнительного шума, присутствующего в восстановленном после первого квантования изображении, а это значит, что $u^{(2)}$ реально отвечает не формуле (3), а соотношению

$$u^{(2)} = \left\lceil \frac{[u/q^{(1)}]q^{(1)} + p}{q^{(2)}} \right\rceil,$$

где p - неучтенный ранее шум, возникающий за счет восстановления изображения.

Это очевидно разрушит «идеальную» визуальную картину DQ-эффекта. Соответствующие гистограммы DCT-коэффициентов далее будем обозначать $\overline{H^{(2)}}$. В реальных условиях восстановления гистограммы, представленные на рис.1, имеют вид, изображенный на рис.2.

Замечание 5. После обратного DCT при восстановлении изображения после первого квантования возможны следующие ситуации:

значения яркости пикселей попадут в границы сегмента $[0, 255]$;

значения яркости большей части пикселей окажутся отрицательными или превосходящими 255.

В первом случае полное восстановление ЦИ потребует дополнительно лишь округлений этих значений до целых, что приведет к возникновению незначительного шума округлений в пространственной области и, как следствие, величина соответствующего шума p в частотной области также окажется небольшой.

Во втором случае процесс полного восстановления ЦИ может привести, в силу значительности p , к последствиям, когда распознавание проявления DQ-эффекта в гистограммах $\overline{H^{(2)}}$ окажется затруднительным или вообще невозможным.

Упомянутая выше проблема не затрагивалась ранее в работах, посвященных изучению и использованию эффекта двойного квантования,

доступных из открытой печати, хотя для его практического применения является чрезвычайно актуальной. Шум, возникающий за счет восстановления ЦИ, является возмущающим воздействием как для матрицы яркости в пространственной области изображения, так и для матрицы DCT-коэффициентов в частотной. Но реакция на один и тот же даже малый шум различных DCT-коэффициентов будет разной. В [6] показано, что коэффициенты, отвечающие низким частотам, будут менее чувствительными [7] к возмущающим воздействиям, чем коэффициенты, отвечающие средним и высоким частотам. Таким образом, даже в случае ситуации 1 возмущения DCT-коэффициентов, отвечающих высоким и, возможно, средним частотам могут оказаться значительными, что приведет к значительным искажениям картины DQ-эффекта в гистограммах $\overline{H^{(2)}}$ для таких частот. Заметим, что коэффициенты высоких частот в принципе являются неинформативными с точки зрения обнаружения следов двойного квантования: они практически все обнуляются еще при первом квантовании.

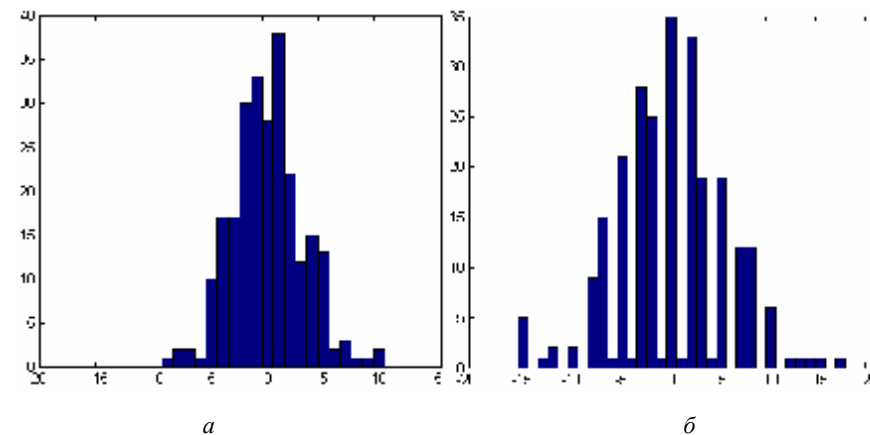


Рис. 2. Гистограммы DCT-коэффициентов частоты (2,2) для ЦИ, подвергнутого процессу двойного квантования с полным предварительным восстановлением после первого квантования: $a - q^{(1)} = 3, q^{(2)} = 5$; $b - q^{(1)} = 5, q^{(2)} = 3$

Иллюстрацией вышесказанного служит рисунок 3 (частота (3,4)): столбцы гистограммы $\overline{H^{(2)}}$, соответствующие значениям -4, 2, 7 (рис.3, б), незаполненные в гистограмме $H^{(2)}$ (рис.3, а), сравнимы по величине со столбцами, отвечающими значениям -7, 8 в $H^{(2)}$, что полностью стирает картину периодического возникновения пиков и впадин. Из вышесказанного вытекает следующий вывод.

Вывод: для обнаружения следов двойного квантования на гистограммах DCT-коэффициентов ЦИ в реальных условиях имеет смысл исследовать не

все гистограммы, а лишь те, которые соответствуют низким частотам, что значительно уменьшит вычислительные затраты для проводимого анализа.

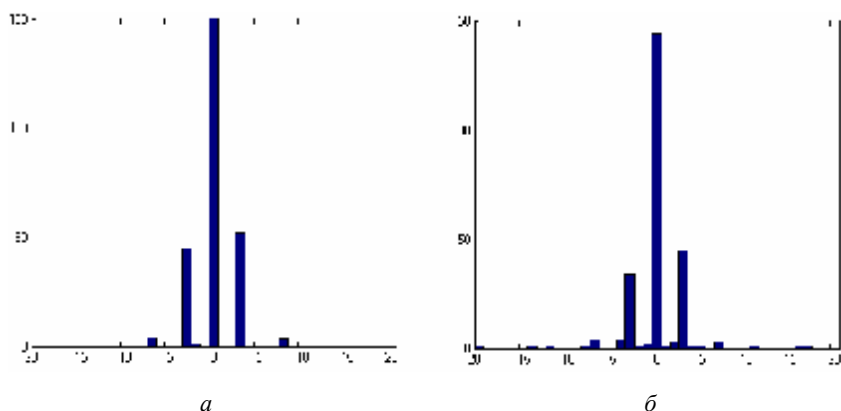


Рис. 3. Гистограммы DCT-коэффициентов ЦИ частоты (3,4): $a - H^{(2)}$; $b - \overline{H^{(2)}}$

Заключение. В работе проведено детальное исследование двукратного использовании JPEG-сжатия изображения с различными коэффициентами квантования. В результате

1. Установлены достаточные условия проявления и отсутствия проявления DQ-эффекта;

2. Получена возможность значительного сокращения вычислительной работы при анализе ЦИ с целью выявления DQ-эффекта в реальных условиях за счет учета различной чувствительности DCT-коэффициентов.

Целью дальнейшей работы авторов является создание на базе проведенных исследований нового метода для выявления фальсификации цифровых фотографий, не требующего никакой дополнительной информации о технических характеристиках использованных фотокамер, позволяющего локализовать область фальсификации.

Список литературы: 1. Гонсалес Р., Вудс Р. Цифровая обработка изображений.- М.: Техносфера, 2005.- 1072 с. 2. Lucas J., Fridrich J. Estimation of Primary Quantization Matrix in Double Compressed JPEG Images // Proc. Of DFRWS 2003, Cleveland, OH, August 5-8. 3. Popescu A. C., Farid H. Statistical Tools for Digital Forensics // Information Hiding. - 2004. - P. 128-147. 4. Popescu A. C., Farid H. Exposing digital forgeries by detecting traces of re-sampling // IEEE Trans. Signal Process. - 2005. - Vol. 53(2). - P. 758-767. 5. He J., Lin Z. u др. Detecting Doctored JPEG Images Via DCT Coefficient Analysis // ECCV(3) 2006: 423-435. 6. Кобозева А.А. Связь свойств стеганографического алгоритма и используемой им области контейнера для погружения секретной информации // Искусственный интеллект.- 2007.- №4.- С.531-538. 7. Деммель Дж. Вычислительная линейная алгебра. – М.: Мир, 2001. - 430 с.

Поступила в редколлегию 23.01.08

В.В. САЙКО, канд. техн. наук, ИП «Стратегия» (г. Желтые Воды)

УРОВНЕВАЯ МОДЕЛЬ ДАННЫХ МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЙ, ОРИЕНТИРОВАННАЯ НА ПОИСК ЗАКОНОМЕРНОСТЕЙ В РЕЗУЛЬТАТАХ ОПРОСОВ

В работе предложена модель данных маркетинговых исследований, позволяющая существенно упростить процесс проектирования опросных листов, ввод данных и их анализ.

У роботі запропонована модель даних маркетингових досліджень, яка дозволяє суттєво спростити процеси проектування опитувальних листів, введення даних та їх аналізу.

The work offers a model of marketing research data that allows to simplify substantially the processes of questionnaire designing, data input and analysis.

Введение. Для решения аналитических задач в маркетинге обычно используется электронная таблица MS Excel, пакеты SPSS и Statistica, хотя практики отмечают слабые возможности визуализации обрабатываемых результатов в этих программных продуктах. В качестве источника внутренних данных (объемы продаж, характеристики клиентов) часто используются CRM-системы (Customer Relationship Management), построенные на базе решений Microsoft.

Программные средства, ориентированные на автоматизацию проектирования маркетинговых исследований и обработку их результатов, можно разделить на три группы:

1. Компактные и доступные программные средства (модули и макросы для MS Access и MS Excel), выполняющие специфические задачи.
2. Полнофункциональные программы, в которых частично формализованы общие для всех маркетинговых исследований процессы и методы.
3. Функциональные модули в интегрированных системах автоматизации маркетинговой деятельности.

Характерными примерами программных средств первой группы являются модули для MS Access: DB1 и МКВ-2 [1]. Программа DB1 используется при проведении и обработке опросов с целью получения информации, связанной с продажами автомобилей.

Примерами второго подхода являются программы «Социолог» и ПриС. Программный комплекс «Социолог» разработан в Тюменском государственном университете с целью автоматизации создания, публикации и обработки анкет [2]. В программном комплексе реализован интерфейс формирования бланка опроса, сформированы базы данных для хранения бланков анкет и результатов опроса, имеется возможность публикации анкет