ISSN 0002-306X. Изв. НАН РА и ГИУА. Сер. ТН. 2009. Т. LXII, № 1.

УДК 004.056.53

ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

Д.Г. АСАТРЯН, Г.С. ШАХВЕРДЯН, Н.С. АСАТРЯН

УСТОЙЧИВЫЙ ЦИФРОВОЙ АЛГОРИТМ ЗАЩИТЫ ИЗОБРАЖЕНИЯ

Рассматривается задача цифровой защиты мультимедийной информации от несанкционированного доступа и использования. Предложен алгоритм, основанный на встраивании цифровых водяных знаков (ЦВЗ) в пространственную область изображения и извлечения из него ЦВЗ. Получены аналитические выражения для среднеквадратической ошибки, вызываемой процедурой встраивания при наличии случайной атаки, зависящие от характеристик изображений, процедуры встраивания и атаки. Методом моделирования исследуется устойчивость предложенного алгоритма к случайной атаке, а также к сжатию изображения по стандарту JPEG.

Ключевые слова: защита изображения, цифровой водяной знак, атака, сжатие информации, устойчивость процедуры.

Введение. Проблема защиты мультимедийной информации от несанкционированного доступа, использования и влияния на нее различных искажающих факторов в последние годы находится в центре внимания исследователей. При этом наиболее интенсивно развиваются методы защиты, основанные на современных информационных технологиях с применением математических методов цифровой обработки сигналов и изображений.

Методов встраивания ЦВЗ существует множество, в зависимости от цели и требований решаемой задачи. Отметим лишь некоторые работы, содержащие достаточно полные обзоры по этим методам. В [1-2] описаны методы, пригодные для широкого класса случаев, в [3] специально рассмотрены методы, применяемые для защиты медицинской информации. В [4] предложен адаптивный алгоритм встраивания ЦВЗ в изображение, основанный на использовании свойства визуальной системы человека [5-6]. В [7] рассмотрена задача защиты медицинской информации от несанкционированного доступа, осуществляемого с целью злонамеренного внесения изменения в информацию. На примере конкретного изображения показано, что принудительные изменения, вносимые в изображение, искажают также встроенный в него ЦВЗ, что может быть обнаружено после изъятия ЦВЗ из атакованного изображения. Показана устойчивость предложенного алгоритма к атакам типа случайных помех, накладываемых на изображение.

В настоящей работе предложен метод цифровой защиты информации, основанный на новом алгоритме встраивания ЦВЗ в пространственную область изображений. При этом показывается устойчивость алгоритма встраивания и извлечения ЦВЗ к различного рода атакам.

Алгоритм встраивания ЦВЗ. Предлагаемый алгоритм встраивания ЦВЗ основан на разбиении изображения на блоки, количество которых равно количеству пикселов ЦВЗ.

Полученные блоки ставятся во взаимно-однозначное соответствие с пикселами ЦВЗ, используя некоторый ключ Ξ для перемешивания пикселов ЦВЗ в соответствии с заданным псевдослучайным правилом. Каждый блок обрабатывается самостоятельно, исходя из значений интенсивностей пикселов блока и ЦВЗ.

Рассмотрим изображение-контейнер I и ЦВЗ W в формате Gray Scale (8 bit). Пусть изображение I имеет размеры $M \times N$ и $I = (a_{mn})$, $m = 0,1,\ldots,M-1$, $n = 0,1,\ldots,N-1$, а ЦВЗ W представляет собой изображение с размерами $K \times L$ и $W = (b_{kl}),\ k = 0,1,\ldots,K-1;\ l = 0,1,\ldots,L-1$. Необходимо встраивать ЦВЗ W в изображение I, т.е. получить новое изображение $I_W = (a_{mn}^W)$ при помощи некоторого оператора встраивания $E_+(I, W, \Xi) = I_W$, где Ξ - секретный ключ. При этом оператор встраивания присваивает новые значения пикселам каждого блока в соответствии с фиксированным правилом.

Встраивание одного байта ЦВЗ в изображение. Рассмотрим сначала простейший случай, когда $K = L = 1$, т.е. изображение состоит из одного блока, а ЦВЗ состоит из одного пиксела с интенсивностью b_{11}. Пусть $M, N > 1$, тогда имеем некоторую избыточную возможность помещения ЦВЗ в пикселы изображения, и при разумном подборе оператора встраивания $E_+(I, W, \Xi)$ можно повысить устойчивость всей процедуры защиты информации в целом. Предлагается оператор встраивания, основанный на следующей формуле:

$$a_{mn}^W = (1-\alpha)a_{mn} + \alpha b_{11},$$

где $m = 0,1,\ldots,M-1;\ n = 0,1,\ldots,N-1$, а $\alpha > 0$ - константа, определяющая степень (силу) встраивания ЦВЗ.

Извлечение ЦВЗ в отсутствие каких-либо атак производится по формуле

$$b_{11} = \frac{\mu^W - (1-\alpha)\mu}{\alpha},$$

где

$$\mu = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} a_{mn},\ \mu^W = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} a_{mn}^W.$$

Рассмотрим теперь ситуацию, когда на изображение I_W воздействует атака X, в результате чего получается новое изображение $I_{W,X}$. Предположим, что атака приводит к изменению пикселов изображения в соответствии с некоторым случайным законом. Примем, для определенности, следующую аддитивную модель изменения значений пикселов:

$$I_{W,X} = I_W + X.$$

Предполагается, что значения x_{mn} случайной величины X накладываются на пикселы независимо и с одинаковой функцией распределения. Тогда

$$a_{mn}^{W,X} = a_{mn}^{W} + x_{mn} = (1-\alpha)a_{mn} + \alpha b_{11} + x_{mn},$$

и задача извлечения ЦВЗ сводится к оцениванию параметра b_{11} по выборке $a_{mn}^{W,X}$, $m = 0,1,\ldots,M-1;\ n = 0,1,\ldots,N-1$. Можно показать, что оценка наименьших квадратов при этом имеет вид

$$\hat{b}_{11} = \frac{\mu^{W,X} - (1-\alpha)\mu}{\alpha},$$

где

$$\mu^{W,X} = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} a_{mn}^{W,X}. \qquad (1)$$

Если принять, что случайная величина X распределена со средним ноль и с дисперсией σ_X^2, то дисперсия оценки (1) будет равна

$$D[\hat{b}_{11}] = \frac{\sigma_X^2}{MN\alpha^2}. \qquad (2)$$

Формула (2) характеризует помехоустойчивость предложенного метода встраивания ЦВЗ к атакам со случайными характеристиками.

Представляет интерес исследование степени искажения изображения вследствие встраивания в него ЦВЗ и воздействия атаки X. Для этого вычислим среднеквадратическое отклонение интенсивностей пикселов изображений I и $I_{W,X}$. Можно показать, что

$$MSE^2 = E\left[\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(a_{mn}^{W,X} - a_{mn}\right)^2\right] = \alpha^2 S_I^2 + \alpha^2(\mu - b_{11})^2 + \sigma_X^2, \qquad (3)$$

где

$$S_I^2 = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(a_{mn} - \mu)^2. \qquad (4)$$

Первое слагаемое выражения (3) является показателем рассеяния интенсивностей пикселов изображения I относительно средней интенсивности. Чем больше этот показатель, тем больше различаются атакованное изображение со встроенным ЦВЗ и изображение-контейнер. Этот факт показывает принципиальное различие предложенного алгоритма от адаптивного алгоритма [4], в котором высокие значения этого же показателя позволяют более успешно встраивать ЦВЗ.

Второе слагаемое в (3) является мерой отклонения интенсивности встраиваемого ЦВЗ от средней интенсивности пикселов изображения. Естественно, бо'льшие отклонения приведут к соответственно бо'льшим отклонениям между изображениями I и $I_{W,X}$. При этом большое значение имеет также параметр α. Третье слагаемое

показывает непосредственное влияние рассеяния самой случайной величины X на процедуру встраивания.

Встраивание полного ЦВЗ в изображение. Предположим, что размеры изображения I кратны размерам ЦВЗ, т.е. при разбиении изображения получается KL блоков с размерами $U = M/K$ и $V = N/L$.

Встраивание ЦВЗ производится следующим образом. Сначала устанавливается взаимно-однозначное соответствие между байтами ЦВЗ и полученными блоками, используя ключ Ξ. Затем каждый пиксел ЦВЗ встраивается в соответствующий блок изображения, используя описанную выше процедуру встраивания одного байта ЦВЗ.

Можно легко получить выражения для суммарного отклонения интенсивностей пикселов изображений I и $I_{W,X}$:

$$MSE^2 = \alpha^2 \frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1} S_{I,kl}^2 + \alpha^2 \frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}(\mu_{kl} - b_{kl})^2 + \sigma_X^2, \qquad (5)$$

где $S_{I,kl}^2$ вычислены аналогично (4) для блока с номерами k и l, а слагаемое

$$\frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}(\mu_{kl} - b_{kl})^2$$

вычислено по всем KL блокам, средняя интенсивность пикселов которых обозначена через μ_{kl}, $k = 0,1,\ldots,K-1;\ l = 0,1,\ldots,L-1$. Третье слагаемое в (5) вычислено для всего изображения I.

Таким образом, степень (среднеквадратического) искажения изображения вследствие встраивания ЦВЗ и при наличии случайных атак по модели (4) задается выражением (5). При отсутствии случайных атак $\sigma_X^2 = 0$, поэтому (5) переходит в

$$MSE^2 = \alpha^2 \frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1} S_{I,kl}^2 + \alpha^2 \frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}(\mu_{kl} - b_{kl})^2, \qquad (6)$$

что показывает степень искажений, вносимых самой процедурой встраивания ЦВЗ. Выражение (6) позволяет еще до встраивания ЦВЗ определить степень ожидаемых искажений и контролировать ее путем надлежащего выбора параметра α.

Результаты моделирования и численных расчетов. Рассмотрим численные примеры встраивания и извлечения ЦВЗ в изображение и оценивания помехоустойчивости предложенной процедуры. Расчеты удобно проводить в логарифмическом масштабе на основе расчета PSNR.

Для эксперимента выбрано изображение маммограммы пациента N, имеющее размеры 256 x 256 пикселов и ЦВЗ с фотографией пациента с размерами 64 x 64. Защита информации состоит во встраивании фотографии пациента в изображение маммограммы. Встраивание ЦВЗ проведено при $\alpha = 0{,}07$.

Визуальный анализ изображения-оригинала I и изображения со встроенным ЦВЗ I_W никаких заметных различий в них не обнаружил. Значение $PSNR = 31{,}6$ дБ также свидетельствует о достаточной близости этих изображений по среднеквадратическому критерию.

Исследование устойчивости предложенного алгоритма проведено для двух типов атак.

1. Устойчивость к случайным атакам, проявляемым в соответствии с моделью (4) с использованием численного моделирования в предположении, что случайная величина X распределена как $X \sim N(0, \sigma_X^2)$ для различных значений дисперсии σ_X^2.

Таблица 1

PSNR (*дБ*) между изображениями I и $I_{W,X}$, извлеченные ЦВЗ и PSNR между встроенным и извлеченным ЦВЗ при различных значениях σ_X

$X \sim N(0, \sigma_X^2)$	$\sigma_X = 0$	$\sigma_X = 2$	$\sigma_X = 4$	$\sigma_X = 6$
PSNR между изображениями I и $I_{W,X}$, *дБ*	31,60	31,09	30,13	28,98
Извлеченные ЦВЗ				
PSNR между встроенным и извлеченным ЦВЗ	∞	28,00	24,26	21,20

В табл. 1 приведены полученные результаты для одной реализации случайной величины X. Видно, что с увеличением σ_X ухудшается качество извлеченного ЦВЗ, оставаясь, впрочем, на уровне, достаточном для узнавания образа пациента вплоть до значения $\sigma_X = 6$. Качество же изображения со встроенным ЦВЗ остается в районе 30 *дБ*, что никак не отражается на диагностической ценности деталей данной маммограммы.

2. ***Устойчивость к сжатию по стандарту JPEG.*** В табл. 2 приведены результаты по исследованию устойчивости предложенной процедуры к атакам в виде сжатия при помощи стандарта JPEG для различных значений параметра качества Q. Эксперименты проводились на тех же изображении-контейнере и ЦВЗ.

Выводы, которые можно сделать из данных табл. 2, во многом схожи с выводами, сделанными относительно табл. 1. Отметим, однако, неожиданно высокую устойчивость предложенной процедуры к достаточно сильному сжатию изображений: ведь при параметре качества, равном всего 40, размер сжатого JPEG-файла для рассматриваемого

изображения составляет около 6% от исходного при сохранении качества на уровне около 31 *дБ* и вполне различимом образе пациента по извлеченному ЦВЗ.

Таблица 2

PSNR (*дБ*) между изображениями I и $I_{W,X}$, извлеченные ЦВЗ и PSNR между встроенным и извлеченным ЦВЗ при различных значениях параметра качества Q

Параметр качества по стандарту JPEG	100	80	60	40
PSNR между изображениями I и $I_{W,Q}$, *дБ*	31,60	31,27	31,04	30,73
Извлеченные ЦВЗ				
PSNR между встроенным и извлеченным ЦВЗ, *дБ*	∞	33,04	27,42	23,99

СПИСОК ЛИТЕРАТУРЫ

1. **Cox J., Miller M. L.** and **Bloom J.A.** Digital Watermarking. - Morgan Kaufmann, 2001.
2. **Hartung M., Kutter M.** Multimedia Watermarking Techniques //Proceedings of the IEEE. – July, 1999. - Vol. 87, № 7. - P. 1079-1086.
3. **Anand D., Niranjan U.C.** Watermarking Medical Images with Patient Information //Proc. IEEE/EMSB Conference. - Hong Kong, China, Oct. 1998. – P. 703-706.
4. **Асатрян Д.Г., Ланина Н.С.** Адаптивный алгоритм встраивания цифровых водяных знаков в изображение //Труды научной годичной конференции Российско-Армянского (Славянского) государственного университета. - Ереван, 2006. - Т.1. - С. 87-90.
5. **Podilchuk C I., Zeng W.** Image-adaptive watermarking using visual models //IEEE Journal of Selected Areas in Communication. – 1998. - 16(4). – P. 525–539.
6. **Chang-Hsing Lee, Yeuan-Kuen Lee.** An Adaptive Digital Image Watermarking Technique for Copyright Protection // IEEE Transactions on Consumer Electronics. – 1999. - V. 45, № 4. - P. 1005-1015.
7. **Asatryan D.G., Lanina N.S., Shahverdyan H.S.** Adaptive Robust Algorithm for Digital Watermarking of Medical Images //Proc. of 6th Int. Conf. on Computer Science and Information Technologies - CSIT'2007. - Yerevan, 2007. - P. 161-164.

Институт проблем информатики и автоматизации НАН РА, Государственный инженерный университет Армении, Российско-Армянский (Славянский) университет.

Материал поступил в редакцию 11.09.2008.

Դ.Գ. ԱՍԱՏՐՅԱՆ, Հ.Ս. ՇԱՀՎԵՐԴՅԱՆ, Ն.Ս. ԱՍԱՏՐՅԱՆ

ՊԱՏԿԵՐԻ ՊԱՇՏՊԱՆՈՒԹՅԱՆ ԹՎԱՅԻՆ ԿԱՅՈՒՆ ԱԼԳՈՐԻԹՄ

Դիտարկվել է բազմամիջավայր ինֆորմացիան չարտոնված ձեռքբերումից և օգտագործումից թվային մեթոդներով պաշտպանելու խնդիր: Առաջարկված է ալգորիթմ՝ հիմնված պատկերի տարածական տիրույթում թվային ջրանիշի ներդրման և արտածման վրա: Ստացվել է պատահական աղմուկի առկայությամբ ներդրման գործառման հետևանքով առաջացած միջին քառակուսային սխալի անալիտիկ արտահայտություն՝ կախված պատկերների, ներդրման ալգորիթմի և աղմուկի բնութագրերից: Մոդելավորման եղանակով հետազոտվել է առաջարկված ալգորիթմի կայունությունը պատահական բնույթի հարձակման դեպքում, ինչպես նաև JPEG ստանդարտով պատկերի սեղմման արդյունքում:

Առանցքային բառեր. պատկերի պաշտպանություն, թվային ջրանիշ, հարձակում, կայունություն, ինֆորմացիայի սեղմում:

D.G. ASATRYAN, H.S. SHAHVERDYAN, N.S. ASATRYAN

DIGITAL ROBUST ALGORITHM FOR IMAGE PROTECTON

A problem of multimedia information protection on illegal access and use is considered. Algorithm for embedding and extracting a watermark in the space-domain of an image is proposed. A formula for mean-square-error connected with the watermark embedding procedure depending on characteristics of the images, procedure, and attack is obtained. The robustness of proposed algorithm to a random attack and compression of an image by JPEG standard is investigated.

Keywords: information protection, digital watermark, attack, information compression, robustness.