

В.І. СОЛОВЙОВ

Компанія «Силентіум Систем», Ванкувер, Канада, e-mail: edemsvi@gmail.com.

О.В. РИБАЛЬСЬКИЙ

Національна академія внутрішніх справ, Київ, Україна.

В.В. ЖУРАВЕЛЬ

Київський науково-дослідний експертно-криміналістичний центр МВС України, Київ, Україна, e-mail: fonoscopia@ukr.net.

О.М. ШАБЛЯ

Одеський науково-дослідний інститут судових експертиз Міністерства юстиції України, Одеса, Україна, e-mail: alik.shablya@gmail.com.

Є.В. ТИМКО

Київський науково-дослідний інститут судових експертиз Міністерства юстиції України, Київ, Україна, e-mail: e.tymko@kndise.gov.ua.

МЕТОД ПОБУДОВИ СИСТЕМИ ВИЯВЛЕННЯ ТА ЛОКАЛІЗАЦІЇ ТОЧОК МОНТАЖУ ЦИФРОВИХ ВІДЕОГРАМ

Анотація. Розглянуто та запропоновано базовий підхід до створення системи виявлення монтажу у цифрових відеограмах та відповідний метод. Запропонований підхід ґрунтується на дослідженнях з ідентифікації апаратури запису цифрових зображень за її власними шумами, зафіксованими на цифрових носіях. Встановлено, що основою методу виявлення та локалізації точок монтажу у відеограмах має бути використання функцій, які описують динаміку помилок ідентифікації сусідніх кадрів та динаміку функції модуля різниці рівнів сигналів кольорів двох кадрів перевірюваної відеограми. Для отримання цих функцій запропоновано застосувати декомпозицію сигналів з використанням вейвлета Хаара. Показано, що реалізацію системи слід здійснювати на нейронних мережах глибокого навчання, що забезпечить високу достовірність експертизи.

Ключові слова: апаратура запису цифрових зображень, вейвлети, відеограма, декомпозиція сигналу, ідентифікація, матриця, нейронні мережі глибокого навчання, власні шуми, криміналістика.

ВСТУП

На сьогодні створення надійних методів виявлення слідів модифікації цифрових зображень (ЦЗ), зафіксованих на цифрових носіях будь-якого типу, має принаймні два аспекти. З одного боку, їх можна розглядати як складову частину численного експертного інструментарію, використовуваного під час проведення експертизи матеріалів та засобів відео- та звукозапису [1–5]. З іншого боку, вони мають стати елементом систем кібербезпеки держави, який забезпечує розв’язання деяких задач, пов’язаних із протидією підробкам, зокрема, створеним із застосуванням технології Deepfake [6–16]. Тому цей інструментарій має забезпечувати виявлення та локалізацію точок монтажу як в окремих ЦЗ, тобто у зображеннях, відзнятих на цифрових фотоапаратах (ЦФА), так і в ЦЗ, знятих і зафіксованих на цифрових відеограмах (ЦВ). У цьому разі слід враховувати, що кожна ЦВ є часовою послідовністю ЦЗ, тобто послідовністю кадрів. Зазначена система має виявляти сліди монтажу, що має вигляд вирізки, вставки або перестановки кадрів із ЦЗ, що входять до змонтованої ЦВ. На думку авторів цієї роботи, такий підхід до постановки задачі можна вважати найбільш універсальним і таким, що забезпечує реалізацію будь-якого з аспектів його застосування. При цьому слід виходити з результатів аналізу та класифікації методів оброблення та протидії такому обробленню ЦЗ з використанням технології Deepfakes, наведених у роботі [17]. Відповідно до класифікації, запропонованої в цій роботі, два з трьох виділених методів протидії побудовані на відмінностях у фоно-

вих та шумових складових ЦЗ, які зазнали такого оброблення. Проте система, запропонована у [17], ґрунтується на тому, що як окремі ЦЗ, так і ЦВ, створені з використанням цієї технології, містять матеріал, знятий на різній апаратурі запису цифрових зображень (АЗЦЗ). Зауважимо, що це не відповідає вимогам судової експертизи, оскільки системи, розроблені для судової експертизи, повинні виявляти сліди монтажу в ЦЗ, записаних не тільки на різній, а й на одній і тій самій апаратурі, зокрема сліди монтажу, виконаного шляхом вирізування та перестановки кадрів або їхніх фрагментів в одній і тій самій ЦВ. Це стосується і модифікацій ЦЗ, знятих одним і тим самим ЦФА.

Мета цієї статті — дослідження підходу до розроблення методу створення системи для виявлення та локалізації точок монтажу в ЦЗ, записаних на будь-якій АЗЦЗ, та виконаного будь-яким відомим способом їхньої модифікації.

БАЗОВИЙ ПІДХІД ДО РОЗРОБЛЕННЯ МЕТОДУ ПОБУДОВИ СИСТЕМИ ВИЯВЛЕННЯ ТА ЛОКАЛІЗАЦІЇ ТОЧОК МОНТАЖУ У ЦИФРОВИХ ВІДЕОГРАМАХ

Розглянемо базовий підхід до ідентифікаційних досліджень АЗЦЗ. На думку авторів, основою подальшого розроблення є гіпотеза щодо строгої індивідуальності власних шумів АЗЦЗ та потреби в їхньому подальшому коректному виділенні для порівняльних досліджень. Ця гіпотеза спирається на те, що статистичні характеристики шумів відеоматриць, як і статистичні характеристики шумів аналого-цифрових перетворювачів (АЦП), мають індивідуальні особливості для кожного конкретного записувального пристрою. Справедливість цієї гіпотези неодноразово підтверджено як на теоретичному та експериментальному рівнях, так і всім попереднім досвідом побудови експертного інструментарію. Сигнали власних шумів АЗЦЗ є сумішшю шумів цих джерел. Розділити їх на окремі складові шумів від відеоматриць та АЦП неможливо. Але це й не потрібно, оскільки завдяки строгій індивідуальності кожної складової цієї суміші, яку фіксують у власних шумах ЦЗ, ця суміш також має суто індивідуальні статистичні характеристики. Потрібно лише забезпечити коректне виділення власних шумів з їхньої суміші з корисним сигналом.

Численні експерименти показали, що для цього доцільно використовувати вейвлет-декомпозицію вхідних сигналів з використанням вейвлета Хаара. Для цього вхідний сигнал кожного пікселя всього рядка розкладають на коефіцієнти апроксимації (сА) та деталізації (сD). При цьому коефіцієнти деталізації належать власним шумам АЗЦЗ.

Цю операцію для рядка червоного кольорофільтра АЗЦЗ з використанням вейвлета Хаара можна представити у загальному вигляді:

$$[cA, cD] = \text{dwt}(\text{image_Red}, 'haar'), \quad (1)$$

де сА — коефіцієнти апроксимації, сD — коефіцієнти деталізації, dwt — вейвлет-перетворення рядка частини відеосигналу певної кольоровості, image_Red — рядок відеосигналу певної кольоровості (у цьому випадку червоний, тобто Red), 'haar' — вейвлет Хаара [18], який записують так:

$$\psi = \begin{cases} -1, & \text{якщо } 0 \leq t < 1/2, \\ 1, & \text{якщо } 1/2 \leq t < 1, \\ 0, & \text{якщо } t \notin [0, 1), \end{cases}$$

де t — час.

Коефіцієнти деталізації мають стійку кореляцію з власними шумами АЗЦЗ, а коефіцієнти апроксимації корелюють з корисним сигналом.

Приклад результату деталізації під час виділення власних шумів для одного рядка ЦЗ із червоним та синім кольорофільтрами наведено на рис. 1.

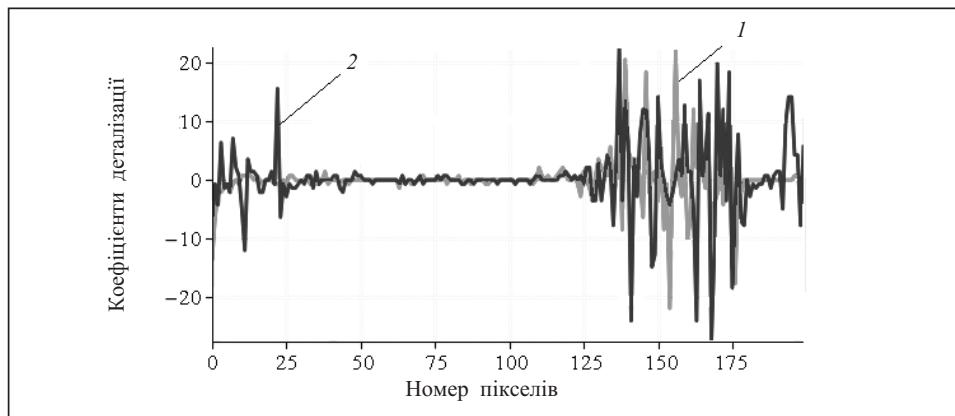


Рис. 1. Фрагмент виділених з i -го рядка ЦЗ власних шумів за коефіцієнтами деталізації (для кольорів Red (крива 1), Blue (крива 2))

Як і під час ідентифікації, операцію декомпозиції потрібно виконувати за всіма рядками та кольорофільтрами ЦЗ. Для кожного кольорофільтра ЦЗ за всіма рядками будують нормований розподіл величини власних шумів.

Експерименти свідчать про близькість цих розподілів для ЦЗ, зроблених одним і тим самим ЦФА, та їхню істотну різницю для ЦЗ, отриманих з використанням різних ЦФА [19].

Оскільки будь-яка ЦВ (відеофайл) є послідовністю ЦЗ, цей підхід цілком придатний і для застосування до відеофайлів, записаних будь-якою АЗЦЗ. Цей підхід реалізовано авторами у системі ідентифікації АЗЦЗ на основі моделі, отриманої під час навчання нейронної мережі на величезному обсязі статистичного матеріалу [19]. Ця модель дає змогу визначити (з певною ймовірністю помилки прийняття рішення), як було створено дві різні ЦВ — на одному й тому самому пристрої запису чи на різних. Порівняльне дослідження виконано на основі статистичного аналізу параметрів відеокадрів у двох ЦВ (без участі EXIF інформації з відеограми). Модель реалізовано у програмі FIVA.

Для побудови системи виявлення точок монтажу ЦВ дослідимо динаміку поведінки власних шумів АЗЦЗ, зафіксованих в одному відеофайлі. Як вихідну величину для кожного відеокадру розглядатимемо функцію динаміки помилок ідентифікації двох сусідніх кадрів за близькістю характеристик шумів апаратури відеозапису (ці помилки безпосередньо дає модель на нейронній мережі, побудована в попередніх дослідженнях та у програмі FIVA [19]).

Ілюстрацію динаміки помилок ідентифікації двох сусідніх ЦЗ (як отриманих з використанням однієї й тієї самої АЗЦЗ) наведено на рис. 2 і 3.

Численні експерименти свідчать про те, що у разі здійснення будь-якого виду монтажу сусідніх кадрів у ЦВ на графіку динаміки помилок ідентифікації практично завжди є присутній максимум, який відповідає межі між змонтованим та незмонтованим кадрами. Цей максимум показано на рис. 2. За відсутності монтажу на реальних графіках динаміки помилок ідентифікації двох сусідніх кадрів також є локальні максимуми (див. рис. 3), однак поведінки графіків динаміки помилок ідентифікації поряд з локальними максимумами у разі наявності та відсутності монтажу суттєво різняться.

Відмінності в поведінці цих графіків можна наочно оцінити за допомогою моделі нейронної мережі, в якій входом є динаміка помилок ідентифікації сусідніх кадрів, а виходом — можлива наявність монтажу на межі двох сусідніх кадрів. Цю модель реалізовано на основі великого статистичного матеріалу щодо динаміки параметрів шумів, як за наявності монтажу, так і за його відсутності. Ефективність навчання нейронної мережі на тестових масивах даних становить 0.975. Множину варіантів відеомонтажу з різними параметрами згенеро-

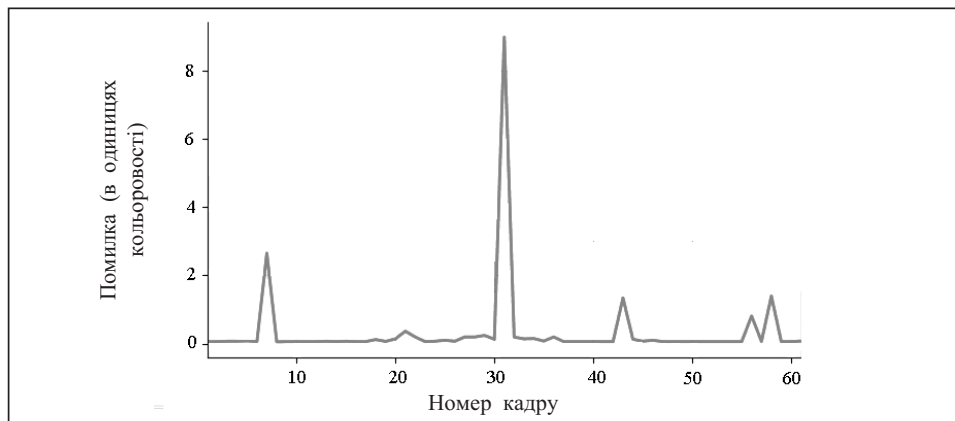


Рис. 2. Динаміка помилок ідентифікації сусідніх кадрів ЦВ за наявності слідів монтажу

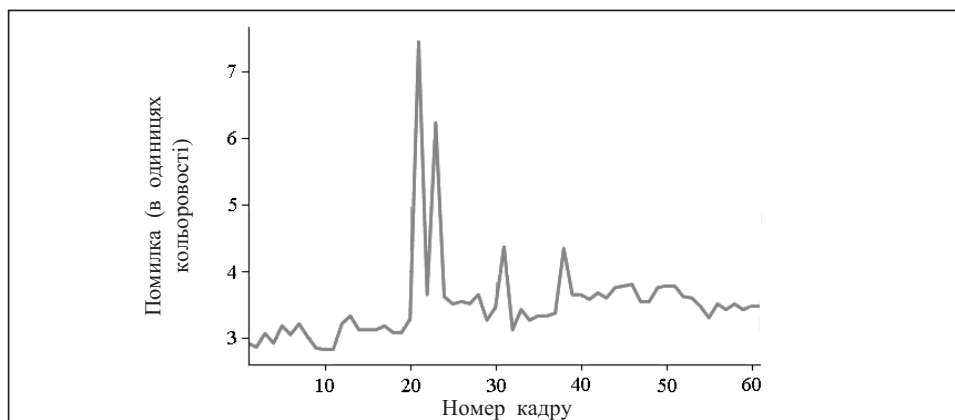


Рис. 3. Динаміка помилок ідентифікації сусідніх кадрів ЦВ за відсутності слідів монтажу

вано у випадковий спосіб спеціальною програмою. Експерименти проведено для різних відеофайлів з різною роздільною здатністю та форматом кадрів.

Як свідчать експерименти, у середньому на кожні 20–30 с ЦВ ця модель дає 3–4 хибні точки монтажу. Цей результат можна вважати прийнятним для практичних цілей, оскільки експерт під час перевірки конкретної ЦВ, маючи точки потенційного монтажу, може проаналізувати зміст суміжних кадрів та прийняти рішення на основі свого візуального сприйняття та експертного досвіду. Проте завдяки використанню додаткової інформації та застосуванню спеціальної технології цей результат можна значно покращити. Для цього можна розглянути додаткову модель динаміки зміни параметрів сусідніх відеокадрів на основі інших факторів, які не пов'язані (або слабо взаємопов'язані) у разі нелінійного зв'язку з шумами АЗЦЗ. На основі двох незалежних факторів можна більш ефективно виявити точки потенційного монтажу. Далі, враховуючи відносну незалежність цих факторів, можна приймати рішення щодо наявності або відсутності монтажу зі значно меншою ймовірністю помилки, оскільки в разі незалежності факторів (подій), за якими приймають рішення, ймовірність помилкового рішення дорівнюватиме добутку ймовірностей помилок прийняття рішення щодо кожного фактора.

Як другий фактор у цьому методі використано суму модулів різниці рівнів сигналів кольоровості для кожної пари пікселів двох суміжних кадрів. Фактично, другий фактор еквівалентний функції відеосигналу (без шумів апаратури, оскільки їх відсікають коефіцієнти апроксимації).

У загальному вигляді функцію декомпозиції суміші власних шумів та корисного сигналу можна записати так:

$$DcAs = DcAr + DcAb + DcAg, \quad (2)$$

де $DcAs$ — середні значення нормованих різниць коефіцієнтів апроксимації для суми трьох кольорів двох суміжних кадрів, $DcAr$, $DcAb$, $DcAg$ — значення нормованих різниць коефіцієнтів апроксимації для трьох кольоровостей двох суміжних кадрів.

Цю різницю можна записати за окремими кольорами у такому вигляді:

$$DvAr = \sum \text{abs} (cAr_{j+1} - cAr_j) / N, \quad (3)$$

$$DvAb = \sum \text{abs} (cAb_{j+1} - cAb_j) / N, \quad (4)$$

$$DvAg = \sum \text{abs} (cAg_{j+1} - cAg_j) / N, \quad (5)$$

де N — кількість пікселів зображення (залежить від роздільної здатності матриці АЗЦЗ), abs — операція обчислення модуля різниці.

У формулах (3)–(5) підсумовування здійснюють за всіма ідентичними пікселями двох суміжних кадрів.

Наведена на рис. 4 функція динаміки зміни нормованих різниць кольоровості сусідніх кадрів свідчить про те, що у точці монтажу обов'язково наявний екстремум різниць кольоровості.

У разі відсутності монтажу на графіку функції динаміки різниці кольорів суміжних кадрів також є мінімуми і максимуми. Але так само, як і для першого фактора, за наявності та відсутності монтажу поведінка наведених на рис. 4 та 5 графіків у точках, які межують з точками екстремумів, істотно відрізняється. Після отримання зазначеної функції кожного сусіднього кадру здійснено навчання нейронної мережі для моделі за другим фактором. Входом для моделі є масиви динаміки зміни різниці кольорів з ознаками та без ознак монтажу, а виходом — можливість ідентифікації точок монтажу.

У результаті застосування розробленого методу до відеофайлів отримано середні значення ймовірності помилок першого та другого роду виявлення та локалізації точок монтажу у межах 0.0005–0.00001. У перерахунку на тривалість відеофайлу це означає, що, наприклад, у разі тривалості відео 3 хв. та частоті кадрів 30 Гц маємо приблизно одну-дві помилки як першого, так і другого роду.

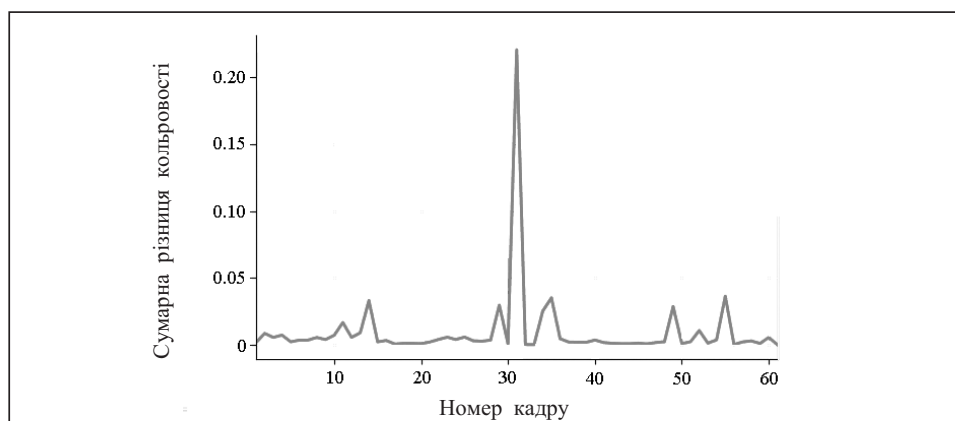


Рис. 4. Динаміка функції модуля різниці рівнів сигналів кольорів двох кадрів за наявності слідів монтажу

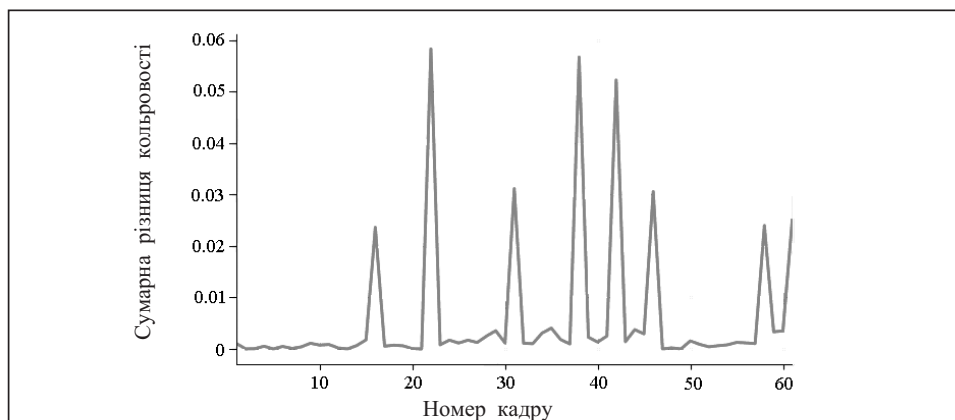


Рис. 5. Динаміка функції модуля різниці рівнів сигналів кольорів двох кадрів за відсутності слідів монтажу

ВИСНОВКИ

1. Запропоновано базовий підхід та метод створення систем виявлення монтажу цифрових відеограм. Базовий підхід ґрунтується на дослідженнях з ідентифікації апаратури цифрового запису зображень за її власними шумами, зафіксованими на цифрових носіях.

2. Розроблено метод виявлення слідів та точок монтажу у відеограмах. В основу методу покладено використання функцій динаміки помилок ідентифікації сусідніх кадрів та динаміки функції модуля різниці рівнів сигналів кольорів двох кадрів перевірюваної відеограми. Для отримання цих функцій застосовано декомпозицію сигналів з використанням вейвлета Хаара.

3. Запропоновано здійснювати реалізацію системи на основі нейронних мереж глибокого навчання, що забезпечить високу достовірність експертизи.

СПИСОК ЛІТЕРАТУРИ

1. Рыбальский О.В. О возможности создания метода проверки подлинности неподвижных изображений, записанных на цифровых носителях. *Спеціальна техніка у правоохоронній діяльності. Матеріали III Міжнародної конференції КНУВС* (25–26 жовтня 2007, Київ, Україна). Київ, 2007. С. 13–14.
2. Рыбальский О.В., Соловьев В.И. К развитию теории, методов и средств проведения экспертизы материалов цифровых фото-, видео- и звукозаписи. *Сучасна спеціальна техніка*. 2012. № 3 (30). С. 119–121.
3. Рыбальский О.В., Соловьев В.И., Белозеров Е.В. Программа идентификации цифровой фото- и видеоаппаратуры и проверки оригинальности цифровых изображений. *Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка*. 2013. Вип. 41. С. 77–80.
4. Рыбальский О.В., Соловйов В.І., Белозьоров Є.В. Методологія розробки методики та програми проведення ідентифікаційних досліджень цифрової апаратури запису зображень. *Сучасна спеціальна техніка*. 2013. № 2 (33). С. 3–7.
5. Solovyov V.I., Rybalskiy O.V., Zhuravel V.V., Semenova N.V. Analyzing the models of speech recognition on the basis of neural networks of deep learning for examination of digital phonograms. *Cybernetics and Systems Analysis*. 2021. Vol. 57. N 1. P. 133–138. <https://doi.org/10.1007/s10559-021-00336-y>.
6. Zhang K., Zuo W., Chen Y., Meng D., Zhang L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*. 2017. Vol. 26, N 7. P. 3142–3155.
7. Cozzolino D., Verdoliva L. Noiseprint: A CNN-based camera model fingerprint. *IEEE Trans. Inf. Forensics Secur.* 2020. Vol. 15. P. 144–159.

8. Li Y., Lyu S. Exposing deepfake videos by detecting face warping artifacts. *Proc. IEEE 14 Conference on Computer Vision and Pattern Recognition Workshops*. 2019. P. 46–52. URL: https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Li_Exposing_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.pdf. <https://arxiv.org/pdf/1811.00656.pdf>.
9. Li L., Bao J., Zhang T., Yang H., Chen D., Wen F., Guo B. Face X-ray for more general face forgery detection. *Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (13–19 June 2020, Seattle, USA). Seattle, 2020. P. 5000–5009.
10. Lee G., Kim M. Deepfake detection using the rate of change between frames based on computer vision. *Sensors*. 2021. Vol. 21, Iss. 21. 7367. <https://doi.org/10.3390/s21217367>.
11. Karandikar A., Deshpande V., Singh S., Nagbhidkar S., Agrawal S. Deepfake video detection using convolutional neural network. *International Journal of Advanced Trends in Computer Science and Engineering*. 2020. Vol. 9, N 2. P. 1311–1315. <https://doi.org/10.30534/ijatcse/2020/62922020>.
12. Li Y., Chang M.-C., Lyu S. Inictu oculi: exposing AI created fake videos by detecting eye blinking. *Proc. 2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (11–13 December 2018, Hong Kong, China). Hong Kong, 2018. P. 1–7. URL: <https://arxiv.org/pdf/1806.02877.pdf>.
13. Koopman M., Rodriguez A.M., Geradts Z. Detection of deepfake video manipulation. *Proc. Irish Machine Vision and Image Processing Conference 2018 (IMVIP 2018)* (29–31 August 2018, Belfast, Northern Ireland). Belfast, 2018. P. 133–136.
14. Zhu W., Ma Y., Zhou Y., Benton M., Romagnoli J. Deep learning based soft sensor and its application on a pyrolysis reactor for compositions predictions of gas phase components. In: *Computer Aided Chemical Engineering*. Eden M.R., Ierapetritou M.G., Towler G.P. (Eds.). Elsevier, 2018. Vol. 44. P. 2245–2250.
15. Shcherbyna Y., Kazakova N., Frazee-Frazenko O., Parchuts L., Schneider S. Analysis of attacks in modern cyberphysical systems. *Proc. 1st International Conference on Intellectual Systems and Information Technologies (ISIT 2019)* (19–24 August 2019 Odesa, Ukraine). Odesa, 2019. CEUR Workshop Proceedings. 2019. Vol. 2683. P. 12–14.
16. Schisterman E.F., Perkins N.J., Liu A., Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005. Vol. 16, Iss. 1. P. 73–81. <https://doi.org/10.1097/01.ede.0000147512.81966.ba>.
17. Дудикевич В.Б., Микитин Г.В., Руда Х.С. Застосування глибокого навчання для виявлення Deepfakes модифікацій біометричного зображення. *Сучасна спеціальна техніка*. 2022. № 1 (68). С. 13–32.
18. Дьяконов В.П., Малла С. Вейвлеты в обработке сигналов. Москва: Мир, 2005. 670 с.
19. Соловйов В.І., Рибальський О.В., Журавель В.В., Шабля О.М., Тимко Є.В. Можливість ідентифікації апаратури цифрового відеозапису та цифрових фотоапаратів. *Кібернетика та системний аналіз*. 2022. Т. 58, № 6. С. 31–37.

V.I. Solovyov, O.V. Rybalskiy, V.V. Zhuravel, O.M. Shablya, Y.V. Tymko
A METHOD FOR CONSTRUCTING A SYSTEM FOR DETECTING
AND LOCALIZING THE PLACES FOR EDITING DIGITAL VIDEOGRAMS

Abstract. The basic approach and method for creating a system for detecting montage in digital videograms are considered and proposed. The proposed approach is based on the identification studies of digital image recording equipment by its own noise recorded on digital media. It has been established that the method for detecting traces and localizing edit points in videograms should be based on the use of the functions of the dynamics of errors in the identification of adjacent frames and the dynamics of the change in the function of the modulus of the difference in the levels of chrominance signals of two frames of the videogram being checked. To obtain these functions, it is proposed to apply signal decomposition using the Haar wavelet. It is shown that the implementation of the system should be carried out on deep learning neural networks, which will ensure high reliability of the examination.

Keywords: digital image recording equipment, wavelets, videogram, signal decomposition, identification, matrix, deep learning neural networks, intrinsic noise, forensics.

Надійшла до редакції 25.09.2022