

А.А. ЯРОВИЙ, С.Г. КАШУБІН, О.О. КУЛИК, І.М. ЛИПКАНЬ
Вінницький національний технічний університет

РОЗПІЗНАВАННЯ МІМІЧНИХ МІКРОВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ОСНОВІ TIME DELAY NEURAL NETWORK

В роботі досліджуються окремі підходи до нейромережевого розпізнавання мімичних мікроекспресій обличчя людини. На основі їх аналізу запропоновано систему розпізнавання динамічних зображень за допомогою Time Delay Neural Network (TDNN) для підвищення точності розпізнавання мімичних мікроекспресій обличчя людини. На основі проведених експериментів досліджено вплив неоднорідності кількості зображень класу на точність розпізнавання та запропоновано спосіб його усунення, що призвело до підвищення точності розпізнавання мімичних мікроекспресій обличчя людини як на основі динамічних, так і на основі статичних зображень.

Ключові слова: нейронні мережі, обробка зображень, розпізнавання образів, мімичні мікроекспресії обличчя людини, time delay neural network.

A.A. YAROVYY, S.G. KASHUBIN, O.O. KULYK, I.M. LYPKAN
Vinnytsia National Technical University

PARALLEL-HIERARCHICAL TRANSFORMATION OF SPOT IMAGES BASED ON THE GPU-ORIENTED ARCHITECTURE

The particular approaches to neural network recognition of human facial microexpression are investigated. On basis of its analysis the TDNN-based system of dynamic images recognition is developed for the purpose of accuracy increment of human facial microexpression recognition. On basis of experimental results the influence of no uniformity of images quantity in collection on accuracy of recognition is investigated. The method for eliminating of above-mentioned influence is proposed. Its ensured to accuracy increment of human facial microexpression recognition both dynamic and static images.

Keywords: neural networks, image processing; pattern recognition, human facial microexpression, time delay neural network.

Вступ

Мімичні мікроекспресії є короткими мимовільними виразами, що виникають на обличчі людини, яка свідомо чи несвідомо приховує або подавляє емоцію. Мімичні мікроекспресії не піддаються свідомому контролю, тому відображають справжнє ставлення людини до того, що відбувається і можуть слугувати ознакою обману при невідповідності до слів чи макроекспресій або індикаторами майбутньої поведінки людини (наприклад, людина в гніві є більш схильною до насильницьких дій, ніж у стані спокою). В той же час мімичні мікроекспресії характеризуються малою тривалістю – до п'ятої частини секунди, тому для недосвідченої людини залишаються зазвичай непоміченими [1]. Вказаний фактор є одним із визначальних для доцільності автоматичного розпізнавання мікроекспресій, що дозволить ідентифікувати мікроекспресії навіть при спостереженні за великою кількістю людей одночасно. Такого роду інтелектуальні системи можуть застосовуватися в медицині, психології, слідчій та охоронній справах. Особливу користь розпізнавання мікроекспресій надає для виявлення неправдивих свідчень, що також визначає такі області застосування, як політика, торгівля і журналістика [2]. Окрім того, система розпізнавання мімичних виразів та мікроекспресій людини є важливим елементом взаємодії людини і комп'ютера. Враховуючи тотальну комп'ютеризацію, розуміння людських емоцій комп'ютером відкриває нові можливості в багатьох галузях, таких як навчальна та ігрова індустрія. У зв'язку з вищевказаним актуалізується необхідність створення інтелектуальної системи, що виконує розпізнавання мімичних мікроекспресій обличчя людини.

Посилення інтересу в сучасному науковому суспільстві до використання нейронних та паралельно-ієрархічних мереж та нещодавні досягнення у підвищенні їх ефективності вказують на можливість їх застосування для створення вищевказаної інтелектуальної системи [3–6].

Представлені дослідження присвячено розпізнаванню мімичних мікроекспресій обличчя людини з використанням технологій нейронних мереж. Задача розпізнавання зображень мімичних мікроекспресій обличчя людини трактується в роботі в контексті задачі класифікації зображень обличчя людини у відповідності до емоції, що відображена мікроекспресією. Вхідними даними системи є цифрові зображення обличчя людини в анфас у кольоровій моделі "Grayscale". У табл. 1 наведено окремі результати проведеного аналізу існуючих методів нейромережевого розпізнавання мімичних мікроекспресій обличчя людини із зазначенням досягнутої точності розпізнавання на невідомих для нейромережі обличчях (тестова вибірка).

Усі вказані методи були використані авторами для розпізнавання виразів обличчя на основі статичних зображень. У загальному випадку розпізнавання виразів обличчя на основі динамічних зображень (відеорядів) можна звести до розпізнавання на статичних зображеннях кожного кадру. Однак такий спосіб не враховує різницю між окремими кадрами динамічного зображення, тому не використовує частину наявної інформації про вираз обличчя, що підлягає розпізнаванню.

В даній роботі запропоновано розпізнавання динамічних зображень за допомогою Time Delay Neural Network – нейромережі, яка пристосована для роботи з часовими послідовностями вхідних даних і враховує вищевказану інформацію [13]. Це дозволяє підвищити точність розпізнавання порівняно із покадровим розпізнаванням виразів обличчя на статичних зображеннях.

Таблиця 1

Існуючі методи нейромережевого розпізнавання мимічних мікровиразів обличчя людини

Метод розпізнавання	Автор(и)	Рік	Точність
Нейронні мережі на основі радіально-базисних функцій	M. Rosenblum, L. Davis, Y. Yacoob [7]	1994	73%
Нейромережі Елмана	J.-Y. Chang, J.-L. Chen [8]	2001	84.7%
Квантові нейромережі	P. Li, J. Li [9]	2008	96.5%
Згорточні нейронні мережі	B. Fasel [10]	2009	30%
Глибокі нейромережі з попереднім навчанням без учителя	T. McLaughlin, G. E. Hinton [11, 12]	2010	67%

В якості тренувальних та тестових зображень обрано зображення із міжнародної бази даних Extended Cohn-Kanade Facial Expression Dataset, University of Pittsburgh, USA (327 динамічних зображень) [14], оскільки ця база даних містить набори кадрів для кожного динамічного зображення (відеоряду). Існуючий підхід до нейромережевого розпізнавання мимічних мікровиразів обличчя людини на статичних зображеннях, що також використовує цю базу, описаний Т. McLaughlin [11] і полягає у застосуванні глибоких нейронних мереж. Найкраща досягнута точність розпізнавання – 67%. Інші вищезгадані підходи застосовують набори окремих статичних зображень замість кадрів із відеоряду, тому такі набори не можуть бути застосовані для тестування методу розпізнавання динамічних зображень.

Метою дослідження є збільшення точності розпізнавання динамічних зображень шляхом використання нейронних мереж типу Time Delay Neural Network. Для цього було використано середовище програмування MATLAB й експериментальним шляхом визначено найкращі параметри мережі та алгоритм навчання.

Задачами дослідження є: визначення показників точності існуючих підходів до задачі нейромережевого розпізнавання мимічних мікровиразів обличчя людини, реалізація системи нейромережевого розпізнавання мимічних мікровиразів обличчя людини на основі Time Delay Neural Network, аналіз одержаних в ході експериментальних досліджень показників точності розпізнавання мимічних мікровиразів обличчя людини на основі Time Delay Neural Network.

Розпізнавання динамічних зображень мимічних мікровиразів обличчя людини на основі Time Delay Neural Network

Time Delay Neural Network (TDNN) – тип нейронних мереж, що застосовується до послідовностей вхідних даних, диференційованих у часі [13]. Їх основною особливістю є те, що вхід кожного нейрона прихованого шару має зв'язки не лише із безпосередньо вхідними даними поточного моменту часу, але й з вхідними даними K попередніх моментів. Зв'язок із i -м елементом вхідних даних має однакову вагу для кожного з K моментів часу, що спрощує структуру нейромережі та, водночас, забезпечує можливість відокремлення незалежних та залежних від часу ознак. Параметр K – затримка прихованого шару TDNN, оскільки мережа починає працювати лише після проходження K квантів часу від початку надходження вхідних даних. Варто зауважити, що затримка може бути введена й на глибших шарах нейромережі. Структуру Time Delay Neural Network зображено на рис. 1.

Для реалізації системи було обране середовище програмування MATLAB, оскільки воно має вбудовану бібліотеку для роботи з Time Delay Neural Network і дозволяє здійснювати моніторинг процесу їх навчання та тестування.

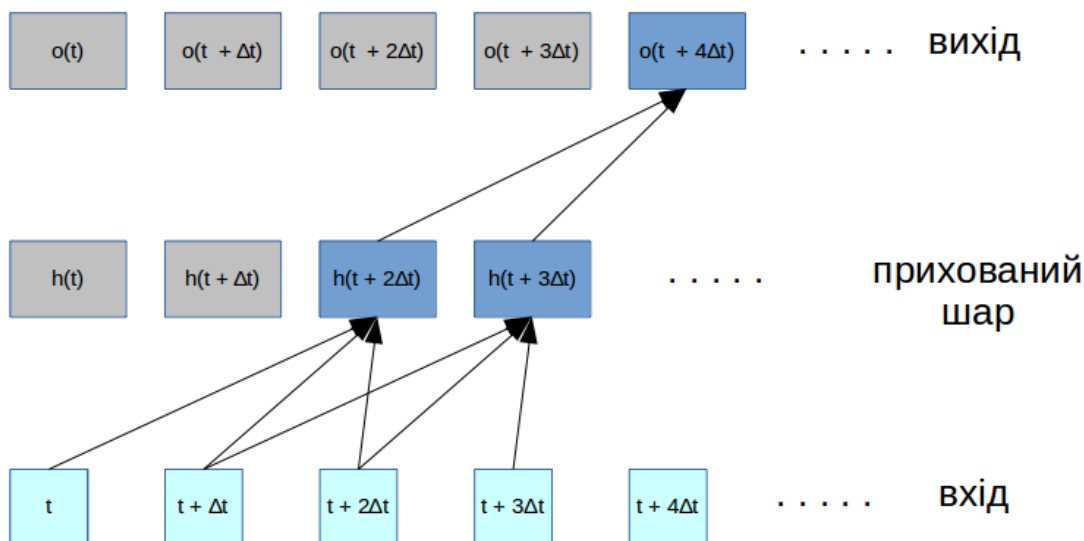


Рис. 1. Структура Time Delay Neural Network із затримкою на вхідному шарі (рівною 2) та затримкою на прихованому шарі (рівною 1)

Кожен відеоряд із бази даних представляє появу на обличчі виразу певної емоції із нейтрального виразу [14]. Оскільки тривалість мімічного мікровиразу складає 0.02 – 0.5 с [1], то відеоряди було штучно пришвидшено, для того щоб імітувати швидку появу емоції. Із кожного відеоряду було обрано 5 зображень – перше, останнє та третє (між ними) на рівній відстані між собою. Окрім адаптації під швидкість появи мікровиразу, це дозволило уникнути необхідності аналізу емоцій на кожному кадрі, а вважати натомість, що на трьох останніх кадрах емоція проявляється, а на двох перших – ні.

В якості тренувального набору було обрано випадковим чином 70% від загальної кількості відеорядів. Відеоряди, що залишилися (30%), було використано в якості тестового набору.

Для передобробки зображення використано: для виділення точок – Active Appearance Model [15], а також нормалізацію значень їх координат до відрізка [-1;1]. Active Appearance Model (AAM) є статистичною моделлю форми досліджуваного об'єкту й розподілу відтінків у моделі "Grayscale" на його зображенні [15]. Під час навчання ААМ виявляє зв'язок між зміщеннями параметрів моделі та різницями між навчальним зображенням і синтезованим за допомогою моделі зображенням. Для узгодження параметрів поточної моделі із зображенням, що обробляється, вимірюються вищевказані різниці й параметри змінюються на значення, отримані за допомогою статистичної моделі, що приводить до більшої відповідності поточної моделі до зображення. Зазвичай прийнятне для подальшого застосування співпадіння досягається вже після кількох таких ітерацій. Для навчання ААМ використовується набір зображень із виділеними точками орієнтуру на кожному зображенні. Точки орієнтуру описують форму досліджуваного об'єкту на зображенні. Приклад такого зображення наведений на рис. 2 [15].



Рис. 2. Зображення із виділеними точками орієнтуру для ААМ [15]

При цьому було використано можливість задавання таких параметрів навчання: об'єм тренувальної й контрольної вибірки, максимальна кількість ітерацій навчання (1000), максимальна кількість ітерацій збільшення похибки розпізнавання контрольної вибірки для передчасної зупинки навчання (75), затримка прихованого шару та кількість нейронів у прихованому шарі. Значення останніх двох параметрів, що дозволили досягти покращення точності, було встановлено експериментальним шляхом для кожного окремого алгоритму навчання.

Для кожного окремого алгоритму навчання із визначеними значеннями всіх параметрів було проведено незалежні експериментальні дослідження, що містили процеси створення та навчання нової мережі. Оскільки початкові ваги зв'язків ініціалізуються випадковими числами, результати навчання двох мереж, створених незалежно, можуть різнитися, незважаючи на однакові архітектури та тренувальні набори зображень.

Середнє значення точності із десяти експериментів розпізнавання динамічних зображень за допомогою Time Delay Neural Network склало 88%, найкраще значення точності – 91%.

Як видно, досягнуто покращення точності на 24% порівняно з існуючим підходом. Це дозволяє зробити висновок, що використання Time Delay Neural Network, а отже – врахування інформації про різницю між сусідніми кадрами – збільшує точність розпізнавання мімічних мікровиразів обличчя людини на динамічних зображеннях.

Дослідження впливу неоднорідності кількості зображень у різних класах на точність розпізнавання

Використаний набір зображень із міжнародної бази даних Extended Cohn-Kanade Facial Expression Dataset, University of Pittsburgh, USA (327 динамічних зображень) [14] містить зображення мікровиразів, що відображають 7 різних емоцій, однак у різних пропорціях. Так, емоції "відраза" відповідає лише 18 зображень, у той час як емоції "радість" – 86. Для дослідження впливу неоднорідності у кількості прикладів кожного класу в тренувальних даних на точність розпізнавання було проведено навчання 7 різних систем бінарної класифікації, які відрізняли одну емоцію від усіх інших та проведено оцінювання точності розпізнавання за метрикою F1-score [16].

Залежність F1-score від кількості зображень у базі, що відповідають класу певної емоції, зображено на рис. 3. Значення 0 відповідає виродженій системі, що класифікувала всі зображення в один клас.

Як видно, зазначена неоднорідність сильно впливає на точність розпізнавання. На початковому етапі було розроблено систему, яка розрізняє лише 3 види емоцій, що представлені в базі найбільшою кількістю зображень (здивування, радість та відраза). Точність розпізнавання склала 99%. Також, було розроблено систему, яка розпізнає 4 види емоцій, що представлені в базі найбільшою кількістю зображень (здивування, радість, відраза та гнів). Точність такої системи склала 94%.

Для зменшення впливу вищевказаної неоднорідності на точність розпізнавання було використано наступний підхід: зображення із класу, що має невелику порівняно з іншими кількість зображень (в даному випадку – менше 60 зображень) було дубльовано відповідну кількість разів (так, щоб в результаті клас був представлений не менш ніж 60 зображеннями) лише в тренувальному наборі. Тестовий набір зображень

залишилися без змін. При цьому співвідношення кількості тренувальних та тестових зображень змінилося до 80% і 20% відповідно за рахунок збільшення кількості тренувальних зображень.

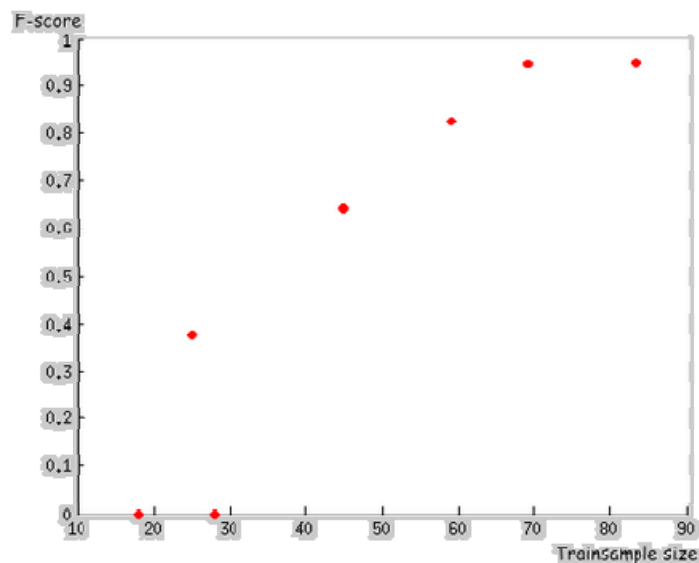


Рис. 3. Залежність метрики F1-score від кількості зображень, що відповідають емоції

Після повторного проведення навчання різних систем бінарної класифікації, які відрізняли одну емоцію від усіх інших, значення F1-score для жодного з класів не було менше за 0.8, що вказує на успішність використаного підходу усунення неоднорідності.

При використанні однієї Time Delay Neural Network для розпізнавання усіх класів при умові дублювання тренувальних зображень класів із малою кількістю представників середнє значення точності з 10 незалежних експериментів – 93%, найкраща досягнута точність – 95%.

Висновки

Проведені дослідження присвячено вирішенню задачі нейромережевого розпізнавання мімічних мікровиразів обличчя людини. На їх основі запропоновано систему розпізнавання динамічних зображень за допомогою Time Delay Neural Network для підвищення точності розпізнавання мімічних мікровиразів обличчя людини.

У роботі проаналізовано існуючі підходи до задачі розпізнавання мімічних мікровиразів обличчя людини у статичній й вказано на їх недоліки у зв'язку із додатковою інформацією, яку містять динамічні зображення. У зв'язку з цим для розпізнавання мімічних мікровиразів обличчя людини на динамічних зображеннях використано Time Delay Neural Network, що пристосована для роботи з часовими послідовностями. В якості тренувальних та тестових зображень для експериментальних досліджень розпізнавання мімічних мікровиразів обличчя людини обрано зображення із міжнародної бази даних Extended Cohn-Kanade Facial Expression Dataset, University of Pittsburgh, USA [14], оскільки ця база даних містить набори кадрів для кожного динамічного зображення (відеоряду). Точність розпізнавання за допомогою Time Delay Neural Network складає 95%, що є значним покращенням порівняно із роботою T. McLaughlin (67%) – існуючим методом розпізнавання мімічних мікровиразів на статичних зображеннях із тієї самої бази даних.

У роботі розглянуто вплив неоднорідності кількості зображень класу на точність розпізнавання та запропоновано спосіб його усунення, що у розглянутому випадку призвів до підвищення точності розпізнавання на динамічних зображеннях з 91% до 95% та на статичних зображеннях з 67% до 72%.

З цього можна зробити висновок, що використання Time Delay Neural Network для розпізнавання мімічних мікровиразів обличчя людини на динамічних зображеннях дозволяє збільшити точність розпізнавання порівняно із покадровим розпізнаванням виразів на статичних зображеннях.

Основним шляхом подальшого вдосконалення даної системи є її модифікація, яка передбачає комбінування інших методів розпізнавання на динамічних та статичних зображеннях. Зокрема, в подальших дослідженнях планується застосування комбінованого підходу із використанням Time Delay Neural Network та глибокої нейромережі для покадрової обробки з метою збільшення точності розпізнавання.

Література

1. Экман П. Психология эмоций / Экман П. – СПб : Питер, 2011. – 336 с.
2. P. Ekman, "Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage", New York: W. W. Norton & Company, 2009. – 416 p.
3. S. O. Haykin "Neural Networks and Learning Machines", Third Edition, Prentice Hall, 2009. – 936 p.

4. S. Wermter, C. Weber, W. Duch, T. Honkela, and more "Artificial Neural Networks and Machine Learning" – ICANN 2014: 24th International Conference on Artificial Neural Networks, Hamburg, Germany, Springer, 2014, 852 p.
5. Паралельно-ієрархічне перетворення як системна модель оптико-електронних засобів штучного інтелекту : [монографія.] / В.П. Кожем'яко, Ю.Ф. Кутаєв, С.В. Свєчніков, Л.І. Тимченко, А.А. Яровий – Вінниця : УНІВЕРСУМ-Вінниця, 2003. – 324 с. – ISBN 966-641-072-9.
6. Образний відео-комп'ютер око-процесорного типу : [Монографія] / Кожем'яко В.П., Лисенко Г.Л., Яровий А.А., Кожем'яко А.В. – Вінниця : Універсум-Вінниця, 2008. – 215 с. – ISBN 978-966-641-261-7.
7. M. Rosenblum, Y. Yacoob, L. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture", IEEE Workshop on Motion of Non-Rigid and Articulated Objects, 1994, pp. 43–49.
8. J.-Y. Chang, J.-L.Chen, "Automated Facial Expression Recognition System Using Neural Networks". Journal of the Chinese Institute of Engineers, 2001, vol. 24, No. 3. – pp. 345-356.
9. P. Li, J. Li, "A Facial Expression Recognition Method Based on Quantum Neural Networks", Proceedings of International Conference on Intelligent Systems and Knowledge Engineering, 2007, pp. 74–78.
10. B. Fasel, "Multiscale Facial Expression Recognition using Convolutional Neural Networks" // IDIAP Research Report 02-52. – 2002. – pp. 2–9.
11. T. McLaughlin, M. Le, N. Bayanbat, "Emotion Recognition with Deep-Belief Networks" // Stanford CS 229 Machine Learning Final Projects. – 2010. – pp. 62–66.
12. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, 2006, vol. 18, no. 7, pp. 1527–1554.
13. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks ", IEEE Transactions on Acoustics, Speech and Signal Processing, 1989, vol. 37, no. 3, pp. 328–339.
14. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference, 13-18 June 2010, pp.94-101.
15. T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active Appearance Models", IEEE Transactions on pattern analysis and machine intelligence, 2001, vol. 23, no. 6, pp. 681-685.
16. D. M. W. Powers, "Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, International Journal of Machine Learning Technology, 2011, vol. 2, no. 1, pp. 37–63.

References

1. P. Ekman, "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life", Publishing House "Piter", 2011. – 336 p.
2. P. Ekman, "Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage", New York: W. W. Norton & Company, 2009. – 416 p.
3. S. O. Haykin "Neural Networks and Learning Machines", Third Edition, Prentice Hall, 2009. – 936 p.
4. S. Wermter, C. Weber, W. Duch, T. Honkela, and more "Artificial Neural Networks and Machine Learning" – ICANN 2014: 24th International Conference on Artificial Neural Networks, Hamburg, Germany, Springer, 2014, 852 p.
5. V.P. Kozhemyako, Yu.F. Kutaev, S.V. Svechnikov, L.I. Tymchenko, A.A. Yarovy, "Parallel-Hierarchical Transformation as the System Model of Optoelectronic Artificial Intelligence Tools", Monograph, Universum-Vinnitsia, 2003. – 324 p. – ISBN 966-641-072-9.
6. V.P. Kozhemyako, G.L. Lysenko, A.A. Yarovy, A.V. Kozhemyako, "Pattern video-computer of eye-processor type", Monograph, Universum-Vinnitsia, 2008. – 215 p. – ISBN 978-966-641-261-7.
7. M. Rosenblum, Y. Yacoob, L. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture", IEEE Workshop on Motion of Non-Rigid and Articulated Objects, 1994, pp. 43–49.
8. J.-Y. Chang, J.-L.Chen, "Automated Facial Expression Recognition System Using Neural Networks". Journal of the Chinese Institute of Engineers, 2001, vol. 24, No. 3. – pp. 345-356.
9. P. Li, J. Li, "A Facial Expression Recognition Method Based on Quantum Neural Networks", Proceedings of International Conference on Intelligent Systems and Knowledge Engineering, 2007, pp. 74–78.
10. B. Fasel, "Multiscale Facial Expression Recognition using Convolutional Neural Networks" // IDIAP Research Report 02-52. – 2002. – pp. 2–9.
11. T. McLaughlin, M. Le, N. Bayanbat, "Emotion Recognition with Deep-Belief Networks" // Stanford CS 229 Machine Learning Final Projects. – 2010. – pp. 62–66.
12. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, 2006, vol. 18, no. 7, pp. 1527–1554.
13. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks ", IEEE Transactions on Acoustics, Speech and Signal Processing, 1989, vol. 37, no. 3, pp. 328–339.
14. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference, 13-18 June 2010, pp.94-101.
15. T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active Appearance Models", IEEE Transactions on pattern analysis and machine intelligence, 2001, vol. 23, no. 6, pp. 681-685.
16. D. M. W. Powers, "Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, International Journal of Machine Learning Technology, 2011, vol. 2, no. 1, pp. 37–63.

Рецензія/Peer review : 5.1.2015 р. Надрукована/Printed :25.1.2015 р.
Рецензент: д.т.н., проф. Кожем'яко В.П.