УДК 004.8

М.В. Давидов, Ю.В. Нікольський, О.В. Пасічник*
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж,
*Львівська гімназія “Сихівська”

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ РОЗПІЗНАВАННЯ У МОДЕЛЯХ ЖЕСТОВОЇ МОВИ

Проведено дослідження, яке спрямовано на пошук параметрів методів розпізнавання зображень реального часу для ідентифікації елементів мови жестів. Метою досліджень є створення програмно-комп’ютерної системи для навчання мові жестів людей, які втратили слух.

The researches made for searching the parameters of methods for the real time perception and authentication the sign language elements. The purpose of researches is to create the software for study the sign language by the people who lost an ear.

Вступ

Дослідження жестової мови, якою користуються люди з пониженим слухом, останнім часом набули значного поширення, оскільки спілкування з людьми, які користуються жестовою мовою, вимагає залучення сурдоперекладачів. Такі фахівці є дефіцитними, їх підготовка стає національним пріоритетом, що, своєю чергою, вимагає застосування до такої підготовки нових інформаційних технологій. Додатковою проблемою є вивчення жестової мови людьми, які втратили слух. Для такої категорії людей потрібні прості засоби та програмно-комп’ютерні тренажери для самостійного засвоєння мови жестів. Отже, задача розроблення нових засобів спілкування, які ґрунтуються на сучасних інформаційних технологіях, набуває значної актуальності.

Автори пропонованої статті впродовж тривалого часу займаються дослідженнями, пов’язаними із створенням програмно-комп’ютерної системи для автоматизації спілкування українською жестовою мовою [1,2]. Результати досліджень демонструвались на Всесвітніх комп’ютерних виставках CeBIT 2006 та CeBIT 2007 [3]. Для побудови систем спілкування проаналізовано систему візуальних сигналів та розроблено програмні засоби розпізнавання таких сигналів з метою їхнього подальшого перетворення у текстові або вербальні повідомлення. Важливою вимогою, яка врахована під час розроблення системи – її використання у реальному часі. В Україні відсутні аналоги проведених досліджень, а розроблення системи із врахуванням специфіки української жестової мови виконується вперше.

Для побудови системи ідентифікації жестів проаналізовано відеозображення та виділено елементи, які визначають зміст та специфіку жестів та дають змогу виявляти особливості різних способів передачі змістовної інформації. Спеціальні дослідження виконано з аналізу дактиля, в якому комбінації пальців мають зміст букв алфавіту, цифр та окремих слів [4] . У дактилі відсутні рухи всієї руки та не задіяні інші частини тіла людини, що зображає жест.

Математичне моделювання жестів, які виконують пальцями однієї руки, виконане з метою розроблення спеціалізованого програмного забезпечення як основи для створення прототипу програмно-комп’ютерної системи для підготовки сурдоперекладачів.

Аналіз останніх досліджень в галузі розпізнавання елементів жестової мови

У сфері досліджень з аналізу та моделювання жестової мови відомі такі основні підходи: за методом отримання інформації про жест, із використанням додаткових засобів отримання інформації про жест та за способами опрацювання інформації про жест.

У першому з підходів є методи, у яких використовують такі технічні засоби:

1) рукавиці із механічними давачами;
2) одну фронтальну камеру;
3) дві камери: одну фронтальну та одну верхню або бічну;
4) одну фронтальну стереокамеру.

У другому підході використано додаткові засоби отримання інформації про жест:

1) за допомогою маркерів на пальцях рук;
2) за допомогою спеціальних рукавиць;
3) без використання маркерів та рукавиць.

Третій підхід оснований на опрацювання отриманої інформації методами:

1) виділення форми долоні;
2) виділення траєкторії руху;
3) виділення міміки обличчя;
4) опрацюванням всього зображення без виділення окремих його частин.

Метод із опрацюванням всього зображення без виділення окремих його частин описано в статтях [5, 6]. За таким методом повне зображення зменшують до розмірів 32х32 піксели та обробляють різними фільтрами з метою виділення важливих його характеристик. Для порівняння жестів, зображення яких містяться у певній базі даних, використовують методи, побудовані на основі прихованої марківської моделі. Такий метод розглянуто у роботі [7], а його ефективність підвищена використанням лінгвістичної моделі, побудованої на основі триграм та визначенням траєкторії руху руки. На базі із 201 речення із трьома доповідачами було отримано 17% помилкових розпізнавань слів.

Для розпізнавання кінців пальців є декілька підходів. Один з них – метод радіальної гістограми, з допомогою якою здійснюють підрахунок зображень пальців [8]. За цим методом можна знайти кількість пальців, якщо на зображенні вони відділені від тла та не перекривають зображення долоні. Метод не застосовний тоді, коли зображення долоні перекриває обличчя.

Мета досліджень

Дослідження, результати якої наведено у цій статті, полягали в аналізі особливостей кодування зображень пальців однієї руки у реальному часі. За наведеною класифікацією досліджувані методи є такими, що отримані однією фронтальною камерою без використання маркерів та рукавиць та опрацьовані як ціле зображення без виділення окремих його частин. Тому програмне забезпечення, яке розробляється в процесі досліджень, дає змогу використовувати для розпізнавання жесту обладнання лише у складі комп’ютера та веб-камери. Досліджено різні аспекти моделювання елементів жестової мови для їх ідентифікації. Для досліджень узято відеопідручник української жестової мови, який використовують для навчання у спеціалізованих школах для дітей з вадами слуху.

Словник української жестової мови налічує близько двох тисяч жестів, кожний з яких означає букву, слово або словосполучення. Жест складається з фіксованих положень пальців рук, долоні, кисті руки, всієї руки, двох рук; його виконують у русі, але їх зображення можна розглядати як послідовності окремих фіксованих картинок. До елементів жесту додається міміка обличчя, артикуляція слів губами, рухи частин тіла.

Для розпізнавання жесту застосовано нейронну мережу, побудовану за схемою багатошарового персептрона.

Постановка задачі дослідження елементів мови жестів

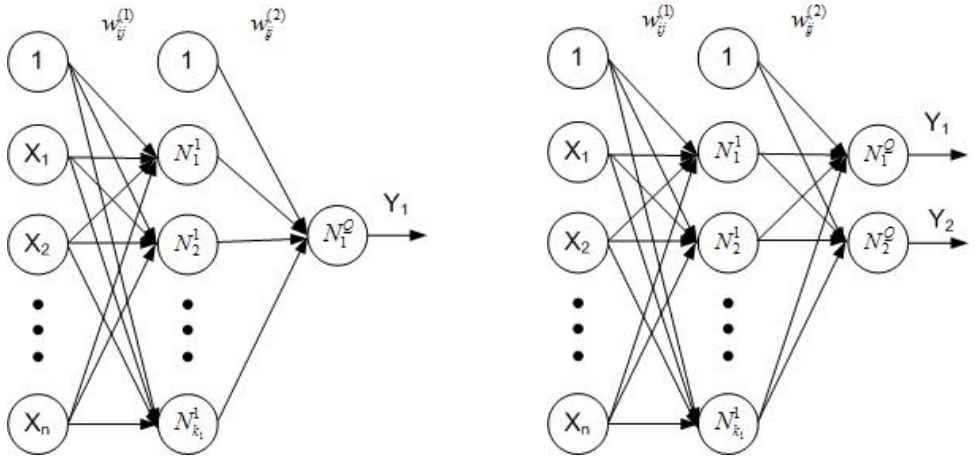
Розглянемо підходи, які використано до вирішення задачі аналізу методів кодування зображень з метою розпізнавання на зображенні комбінацій пальців однієї руки. Її вирішення дає змогу класифікувати такі комбінації та ідентифікувати елементи дактиля.

Основною складністю розпізнавання елементів жестової мови та, зокрема, кінців пальців є перекриття зображень долоні з частиною обличчя та неоднорідності тла. У разі визначення країв перекриття зображення руки та елементів обличчя призводить до помилкового розпізнавання кінців пальців.

Результати досліджень із розпізнавання кінців пальців стосуються методу еталона для створення множини навчальних прикладів, модифікованого методу найшвидшого спуску [9] та методу спряжених градієнтів навчання нейронних мереж. Метод еталона та модифікований метод найшвидшого спуску для навчання нейромереж запропоновано авторами цього дослідження.

Порівняно методи навчання з використанням нейронних мереж та інші методи, які використано для знаходження кінців пальців руки на зображенні.

Для навчання нейронної мережі за модифікованим методом найшвидшого спуску використана мережа з одним виходом, зображена на рис.1,а. Тут значення $Y_1 \geq 0{,}5$ вважаємо ознакою належності досліджуваної області кінцю пальця. Для навчання нейронних мереж за методом спряжених градієнтів побудовано мережу з двома виходами, зображену на рис.1,б. Для її навчання використано програмну реалізацію методу спряжених градієнтів з бібліотеки LTI-Lib (http://ltilib.sourceforge.net/). Якщо $Y_1 \geq Y_2$, то вважаємо, що досліджувана область зображення належить кінцю пальця, інакше – не належить. Така нейронна мережа виявилась нестійкою до шумів на зображенні. Тому додатково виконано згладжування результату усередненням за чотирма сусідніми пікселами. Зауважимо, що не вдалось навчити нейромережу за методом найшвидшого спуску, реалізованим у бібліотеці LTI-lib.



а *б*

Рис. 2. Мережі, які використано в експериментах:
а – мережа, яку навчено за модифікованим методом найшвидшого спуску;
б – мережа, яку навчено за методом спряжених градієнтів

Для відділення зображення кінців пальців від тла та визначення ступеня належності піксела зображенню кінця пальця побудовано метод чорно-білого еталона, який полягає у наступному. Нехай C – матриця розмірів $M \times N$ з елементами c_{ij}, $c_{ij} \in [0,1]$, $1 \leq i \leq M$, $1 \leq j \leq N$ – яскравостями пікселів еталона. Значення c_{ij} задає ступінь належності точки зображення із координатами (i, j) об'єкта, а значення $1 - c_{ij}$ – ступінь належності точки (i, j) тлу. Нехай матриця D з елементами d_{ij}, $1 \leq i \leq M$, $1 \leq j \leq N$ містить область зображення розмірів $M \times N$ пікселів у кольоровому просторі зображення Q. Тут об'єктом є кінці пальців. Накладанням зображення та еталона обчислено середньозваженій колір та середньоквадратичне відхилення кольору в кольоровому просторі для відповідних пікселів в околі об'єкта $(\overline{D}_1, \overline{S}_1)$ та тла $(\overline{D}_2, \overline{S}_2)$ за формулами

$$\overline{D}_1 = \frac{\sum_{i,j} d_{ij} \cdot c_{ij}}{\sum_{i,j} c_{ij}},\ \overline{D}_2 = \frac{\sum_{i,j} d_{ij} \cdot (1 - c_{ij})}{\sum_{i,j} (1 - c_{ij})},\ \overline{S}_1 = \frac{\sum_{i,j} \left| D_{ij} - \overline{D}_1 \right|^2 \cdot C_{ij}}{\sum_{i,j} C_{ij}},\ \overline{S}_2 = \frac{\sum_{i,j} \left| d_{ij} - \overline{D}_2 \right|^2 \cdot (1 - c_{ij})}{\sum_{i,j} (1 - c_{ij})}.$$

Ступінь відмінності околу зображення від еталона визначатимемо за формулою

$$S_{12} = \frac{\left|\overline{D}_1 - \overline{D}_2\right|^2 + \left|\overline{S}_1 - \overline{S}_2\right|}{1 + \overline{S}_1 + \overline{S}_2}.$$

Що ближче значення S_{12} до нуля, то менше статистичні показники околу об'єкта відрізняються від показників тла та менша імовірність того, що в досліджуваному околі знаходиться шуканий об'єкт.

Для пришвидшення навчання нейромережі на основі методу найшвидшого спуску [9] побудовано модифікацію методу зворотного поширення похибки, яка полягає у введенні множника K_{back} та обчисленні множника $\delta_j^{(q)}$ за формулою $\delta_j^{(q)} = K_{back} F'\left(s_j^{(q)}\right) \cdot \sum_{r=1}^{k_{q+1}} \delta_r^{(q+1)} w_{jr}^{(q+1)}$, $q = Q-1, Q-2, ..., 1$, $j = 1,2,..., k_q$. Тут Q – кількість шарів нейронів; q – номер шару нейронів, $q = 1,2,..., Q$; k_q – кількість нейронів у шарі q; $x_j^{(q)}$, $j = 1,2,..., k_{q-1}$ – входи нейронів шару з номером q, $x_0^{(q)} = 1$; $x_j^{(1)} = x_j$ – входи мережі; $w_{ij}^{(q)}$ – вага синаптичного зв'язку нейронів з номерами $N_i^{(q-1)}$ та $N_j^{(q)}$; $s_j^{(q)} = \sum_{i=0}^{k_{q-1}} w_{ij}^{(q)} x_i^{(q)}$, $q = 1,2,..., Q$, $j = 1,2,..., k_q$ – зважена сума входів нейрона з номером $N_j^{(q)}$; $y_j^{(q)}$ – значення, обчислене нейроном $N_j^{(q)}$, $y_j^{(q)} = F\left(s_j^{(q)}\right)$; $x_j^{(q)} = y_j^{(q-1)}$, $j = 1,2,..., k_{q-1}$, $q = 2,3,..., Q$ – входи нейронів шару q, які дорівнюють виходам нейронів шару $q-1$; $F'\left(s_j^{(q)}\right)$ – похідна активаційної функції нейрона $F(x)$ у точці $s_j^{(q)}$. Усі параметри алгоритму обчислені у певний момент часу t. Навчання нейромережі виконано на прикладах, кожний з яких є числовим кортежем з компонентами – характеристиками пікселів зображення. Кожному прикладу поставлено у відповідність одиниця, якщо він містить зображення кінця пальця, та нуль – в іншому випадку. Для побудови прикладів попередньо опрацьовано зображення з метою зменшення кількості елементів кожного піксела. Експериментально встановлено, що жодна з компонент формату RGB, у якому зображення надходить з відеокамери, не дає змоги відрізняти руку від деталей тла. Тому кольори подано у форматі YCbCr. Запропоновано формувати приклади з околу кожного піксела, який має форму мальтійського хреста. Розмір околу задано параметром R – кількістю пікселів на боці хреста. Для $R = 2$ кожний приклад складається з 9 елементів із значеннями Cr-компоненти кольору пікселів. Якість розпізнавання оцінено за якістю розпізнавання, успішністю та часом навчання у секундах. Якість розпізнавання визначаємо за відсотком помилок розпізнавання, який обчислюємо за формулою

$$Err\% = \frac{N_{notfound} + N_{spurious}}{N_{fingertips}} \cdot 100\%,$$

де $N_{not\,found}$ – кількість нерозпізнаних кінців пальців, N_{spurious} – кількість неправильних розпізнавань, $N_{fingertips}$ – загальна кількість кінців пальців на навчальному зображенні (прийнято $N_{fingertips} = 19$). Неправильною вважаємо ідентифікацію мережею елемента зображення як кінця пальця, який насправді ним не був. Спробу навчання вважатимемо успішною, якщо для неї отримано Err%<50%. Для вибору оптимального значення кількості нейронів у прихованому шарі проведено низку експериментів. Їх результати показано на рис. 2. Найкращі результати отримано для 5 та 6 нейронів. Активаційною функцією нейронів обрано непарну сигмоїдальну функцію $F(x) = \left(1 + e^{-x}\right)^{-1} - 0.5$.

Приклади для навчання нейронної мережі обрано з різних частин навчального зображення так, щоб кількості зображень, які містять кінці пальців, та зображень тла були однаковими. Для вибору найкращого методу навчання нейронну мережу навчали за методами найшвидшого спуску, запропонованою його модифікацією, а також методом спряжених градієнтів. Навчання виконано за 200 циклів, а критерієм завершення є граничне значення похибки, яке дорівнює 10%.

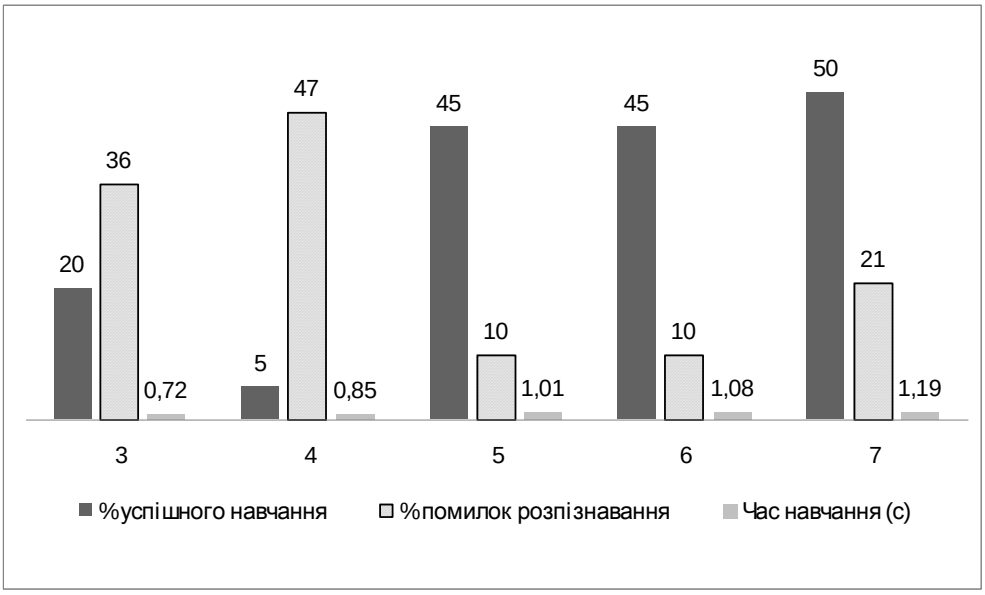


Рис. 2. Залежність якісних характеристик навчання від кількості нейронів у прихованому шарі

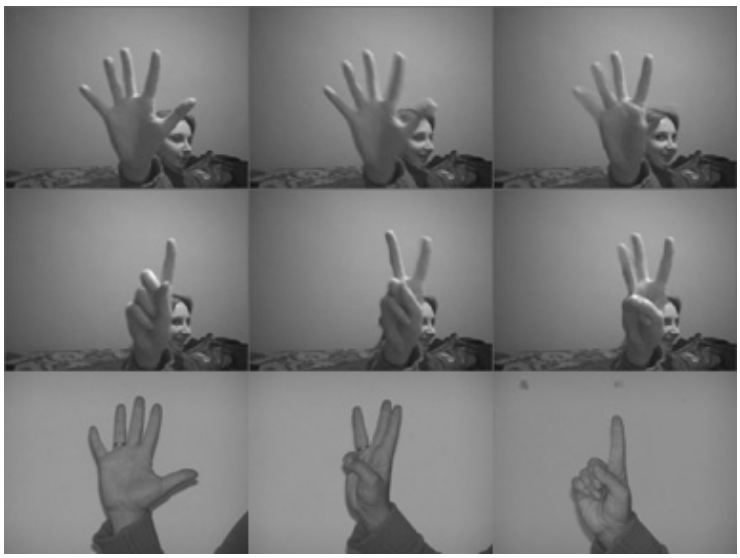
Знаходження кінців пальців за допомогою навченої нейронної мережі виконане на усьому кадрі разом із тлом, на якому відбувалась відеозйомка. Кисть руки на ньому не відокремлювалась від тла жодним способом. Якість розпізнавання оцінено на чотирьох зображеннях. Перше з них складалося з кадрів, використаних для навчання нейромережі; друге (рис.3,а) та третє сформовані з кадрів руки тієї самої людини за тих самих умов освітлення, що і на першому зображенні. Четверте зображення складалось з кадрів руки іншої людини, відзнятої як в тих самих, так і в інших умовах освітлення (рис. 3, б). На другому зображенні було видалене тло, яке заважало розпізнаванню пальців, на третьому тло було залишено. Для кожного набору параметрів здійснено 20 спроб навчання для $N = 80$ навчальних прикладів. На рис.4 показано розподіл сумарного часу опрацювання одного кадру зображення нейромережею, навченою різними методами. Нейронна мережа, навчена модифікованим методом найшвидшого спуску, значно швидше опрацювала зображення, ніж мережа, навчена програмами з бібліотеки LTI-lib. Це можна пояснити тим, що цей метод виконує обчислення лише для одного вихідного нейрона, а обчислення оптимізовані під набір команд процесора SSE2, на якому вони виконувались.

Висновки

Проведені дослідження із розпізнавання елементів жестової мови спрямовані на вибір методів навчання нейронної мережі для опрацювання зображень. Зображення надходять з відеокамери, а опрацьовуються у реальному часі. Людина демонструє рукою жест, а задачею розпізнавання є виділення на отриманих кадрах кінців пальців, за якими можна ідентифікувати знак, показаний відповідним жестом. За результатами досліджень встановлено, що для розв’язання поставленої задачі спочатку необхідно виконати перетворення способу кодування. Це дало змогу відділити

зображення руки від тла. Розпізнавання пальців руки виконане розв'язуванням відповідно сформульованої задачі машинного навчання. Для цього побудована нейромережа прямого поширення та розроблена модифікація алгоритму її навчання, основаного на методі найшвидшого спуску. Порівняно результати застосування різних методів навчання нейромережі. Введено поняття прикладу, розроблено алгоритм їх формування та методику утворення навчальної множини для уникнення перенавчання мережі. Навчена нейромережа використана для опрацювання нових прикладів. Якість розпізнавання зображень оцінено низкою параметрів. Обрано значення параметрів мережі та алгоритмів навчання, які забезпечили найкраще розпізнавання елементів жестової мови на нових зображеннях у різних умовах освітлення та для різного тла.

а



б

Рис. 3. Зображення, які використано для оцінювання якості розпізнавання: а – зображення із тлом; б – зображення руки іншої людини, отримане в інших умовах освітлення

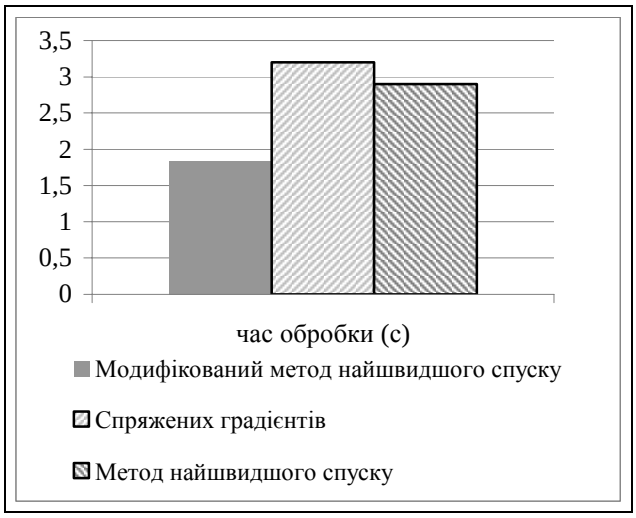


Рис. 4. Час опрацювання одного кадру нейромережею, навченою різними методами

1. Davydov M., Nikolski I., Pasichnyk O. System of finger movement identification for sign language recognition // Abstracts of First Central European Student Conference in Linguistics, 29-31 May 2006. Budapest, Hungary. – P.23–25. [2]. Давидов М.В., Нікольський Ю.В., Пасічник В.В. Математичне моделювання та програмна реалізація елементів тренажеру для навчання жестовій мові людей, що втратили слух // Сборник трудов седьмой международной конференции „Интеллектуальный аналіз информации (ИАИ-2007)", Киев, 15–18 мая 2007 г. – С.56–66. [3]. Давидов М.В., Нікольський Ю.В., Пасічник В.В. Програмний тренажер для навчання мові жестів // Праці міжнародної наукової конференції „Розвиток інформаційно-комунікаційних технологій та розбудова інформаційного суспільства в Україні". (м. Ганновер, Німеччина, CeBIT-2007). Спеціалізований тематичний додаток до загальногалузевого науково-виробничого журналу „Зв'язок". – К., 2007. – С.98–106. 4. Давидов М.В., Нікольський Ю.В. Нейромережний класифікатор елементів відеозображень реального часу // Вісник Нац. ун-ту „Львівська політехніка". – 2006. – №564. – С. 18–25. 5. Morteza Zahedi, Daniel Keysers, and Hermann Ney. Appearance-Based Recognition of Words in American Sign Language. Pattern Recognition and Image Analysis, pp. 511–519, 2005. 6. Morteza Zahedi, Daniel Keysers , Thomas Deselaers and Hermann Ney. Combination of Tangent Distance and an Image Distortion Model for Appearance-Based Sign Language Recognition. Pattern Recognition, Springer Berlin / Heidelberg. Pages 401–408, 2005. 7. P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. Speech Recognition Techniques for a Sign Language Recognition System. In Interspeech 2007, pages 2513–2516, Antwerp, Belgium, August, 2007. ISCA best student paper award Interspeech 2007. 8. Moritz Storring, Thomas B. Moeslund, Yong Liu, and Erik Granum. Computer vision-based gesture recognition for an augmented reality interface, In 4th IASTED International Conference on visualization, imaging, and image processing, pages 766–771, Marbella, Spain, Sep 2004. 9. М.В.Давидов, Ю.В.Нікольський. Класифікація елементів відеозображень реального часу з допомогою нейромережі. Вісник Національного університету "Львівська політехніка", Інформаційні системи та мережі, Львів, №549, 2005. – С.82–92.