УДК 004.932

С.Г. Антощук, д-р техн. наук,
О.С. Яровенко

ЛЮДИНО-МАШИННИЙ ІНТЕРФЕЙС НА ОСНОВІ РОЗПІЗНАВАННЯ ЖЕСТІВ

Проаналізовано і змодельовано методи реалізації людино-машинних інтерфейсів на основі жестів, а саме: методи захвату зображення з камери, сегментації зображень і розпізнавання жестів. Розроблено інтерфейс на основі аналізу жестів рук.

Проанализированы и смоделированы методы реализации человеко-машинных интерфейсов на основе жестов, а именно: методы захвата изображения с камеры, сегментации изображений и распознавания жестов. Разработан интерфейс на основе анализа жестов рук.

In this work were analyzed and modeled methods of realization of human-computer interfaces on the basis of gestures, namely: methods of capture of image from a chamber, segmentation of images and recognition of gestures. An interface is developed on the basis of analysis of gestures of hands.

У теперішній час завдяки розвитку комп'ютерного зору та поширенню використання комп'ютерних систем у різних сферах життя постійно зростає інтерес до інтуїтивно зрозумілих людино-машинних інтерфейсів. Бурхливий розвиток інформаційних технологій і цифрових пристроїв вимагають нових підходів до людино-комп'ютерної взаємодії, щоб полегшити адаптацію, навчання і процес роботи користувачів.

Ще у 70-х роках минулого століття ученими була запропонована концепція графічного інтерфейсу користувача (GUI - Graphical user interface), яка замінила собою консольні застосування, значно спростивши роботу користувачів. Проте GUI не здатний охопити усе різноманіття людських почуттів і вмінь, які люди набули впродовж взаємодії з матеріальним світом.

Нова хвиля розвитку людино-машинної взаємодії заснована на сприйнятті (Perceptual User Interface) та містить матеріальні, голосові (Voice User Interface), сенсорні (Touch User Interfase) і засновані на комп'ютерному зорі (Vision Based Interface) інтерфейси, які використовують природні для людини способи введення інформації: мову, письмове введення, жести рук, напрям погляду, мімі-ку, рухи голови і тіла людини.

Рухи рук – одна з найбільш природних форм спілкування серед людей. Інтерфейси на їх основі привернули особливу увагу останнім часом, особливо в області комп'ютерного зору. Такий тип спілкування надає виразний, природний і інтуїтивно-зрозумілий спосіб людино-машинної взаємодії. На сьогодні управління цифровими пристроями за допомогою жестів є дуже актуальною і перспективною областю досліджень, яка містить безліч невирішених задач. Управління жестами викликає інтерес у виробників побутової техніки, особливо це актуально у разі управління мультимедійними системами.

Вбудована система управління жестами за допомогою мультимедійних пристроїв може замінити собою пульт дистанційного керування. Така система може включати набір з 10 жестів (табл. 1) для реалізації основних функції управління.

Загальний алгоритм роботи системи управління жестами містить ряд основних етапів:

захоплення зображення – здійснюється за допомогою веб-камери, яка підключена до ПК і безперервно передає в програму послідовність кадрів (кольорових зображень) в реальному часі (25 кадрів/ сек.);

локалізація руки на зображенні – на зображенні, одержаному за допомогою веб-камери, знаходиться область руки. Процедура локалізації містить, як правило, два етапи: сегментацію (виділення) за кольором та аналіз отриманих областей, щоб позбавитися

© Антощук С.Г., Яровенко О.С., 2010

1. Набір жестів, що розпізнаються системою

№	Жест	Позначення	№	Жест	Позначення
1		«Start»	6		«Volume Up»
2		«Stop»	7		«Volume Down»
3		«Pause»	8		«Close»
4		«Previous»	9		«Forward»
5		«Next»	10		«Full Screen»

невірно сегментованих часток зображення і знайти область на зображенні, що відповідає області руки;

розпізнавання жестів – на отриманому зображенні знаходимо контур руки і його характеристики. Після цього можна класифікувати жест;

формування програмної команди на основі розпізнаного жесту.

Існує декілька засобів одержання зображень жестів:

тривимірні камери (Depth-aware камери) – використовуються спеціальні TOF-камери (Time-of-flight scanning camera), які здатні вимірювати відстань до кожної з точок об'єкта, що потрапляє в поле її зору. Відстань вимірюється за зміною фази відбитого світла. Зображення подається в тривимірному просторі. Не дивлячись на низьке розділення (не більше за 320x240 пікселів), така камера здатна захопити 100 зображень в секунду. Завдяки тому, що на відміну від звичайних камер, тривимірні не чутливі до змін світла, руху і зашумленності фону, вони часто використовуються для розпізнавання жестів;

стереокамери – застосовуються дві (і більше) камери, взаємозв'язок між якими відомий заздалегідь. Тривимірне подання зображення апроксимується з отриманих вихідних зображень камер. Для визначення взаємозв'язку камер може використовуватися відношення їх позицій;

одиночні камери – використовуються для розпізнавання жестів там, де ресурси або середовище не дозволяють використовувати інші форми розпізнавання зображень. Слід відзначити, що такі камери не так ефективні як стерео- обо TOF-камери, але мають безумовну перевагу, оскільки вони доступні широкому загалу користувачів і прості в налаштуванні і застосуванні. Згідно з вимогами щодо забезпечення роботи системи у реальному часі (для цього система має обробляти приблизно 15 – 25 кадрів/с) розмір оброблюваного зображення обмежений розділенням 320x240 пікселів.

Мета процедури сегментації – зменшити обсяг оброблюваної інформації. Існуючі методи сегментації засновані на урахуванні властивостей однорідності або неоднорідності обраної ознаки, наприклад, інтенсивності. Найпоширенішим завдяки простоті реалізації та високій швидкодії з методів, що засновані на першому підході, залишається сегментація за порогом. Результат сегментації початкового зображення $f(x, y)$ визначається так:

$$g(x, y) = \begin{cases} 1, \textit{якщо} & f(x, y) > T \\ 0, \textit{якщо} & f(x, y) \le T \end{cases},$$

де T – значення порогу.

Але, як показали дослідження, результати цієї сегментації чутливі до таких факторів: освітленості приміщення; розташування

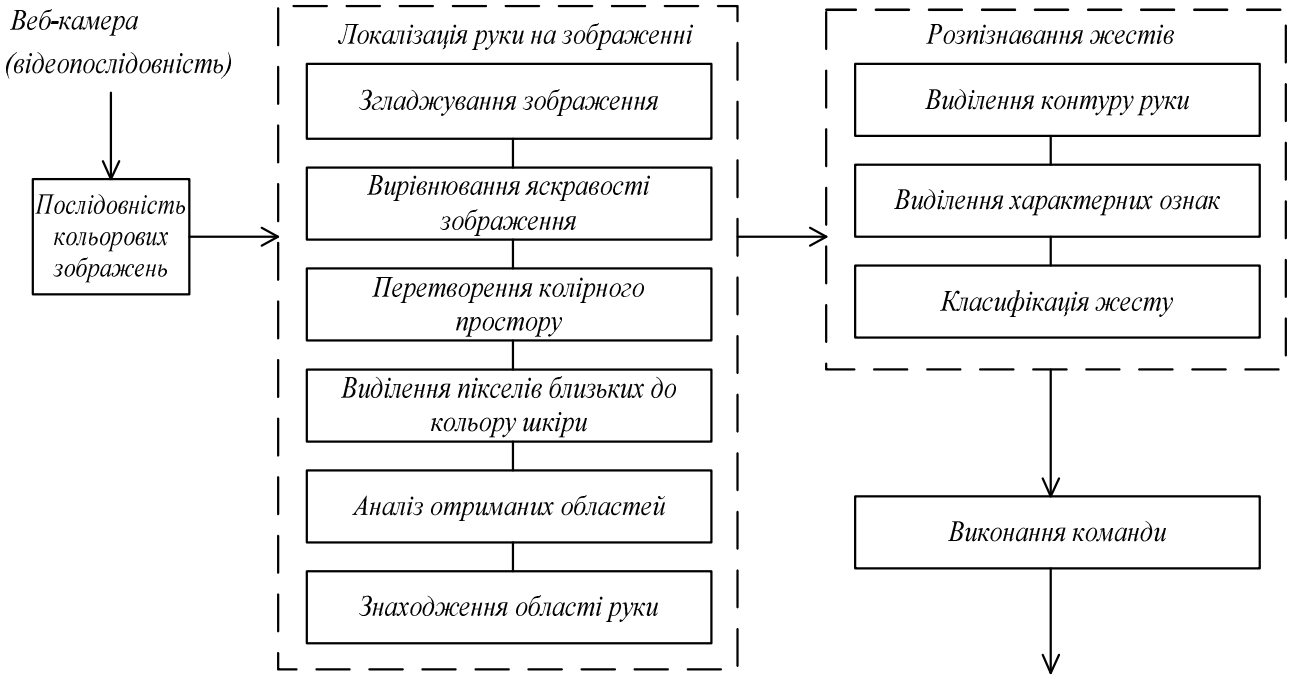


Рис.1. Структурна схема системи

джерел світла (знаходження за об'єктом, перед об'єктом або збоку викликає різну нерівномірну освітленість руки і, як результат, підвищує вірогідність втрати частини пікселів, які належать області, що цікавить нас); фону (наявність областей фону, колір яких такий же, як і колір шкіри); колірний баланс камери або переважання якого-небудь відтінку (жовтого або червоного – при штучному освітленні, синього або білого – при природному освітленні або лампах денного світла у разі захоплення зображення з веб-камери).

Дію цих факторів можна усунути або частково знизити при використанні різних колірних просторів. Для того, щоб обрати колірний простір, який дозволяє отримати найкращі результати, було проведено моделювання сегментації шкіри у різних колірних просторах (RGB, HSV, нормалізований RGB).

Перетворення моделі RGB в HSV відбувається за наступними виразами:

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)\cdot(G-B))}},$$

$$S = 1 - 3\cdot\frac{\min(R,G,B)}{R+G+B},$$

$$V = \frac{1}{3}\cdot(R+G+B),$$

де *R, G і B* – значення відповідно червоної, зеленої і синьої складових.

Перехід до нормалізованого RGB здійснювався процедурою нормалізації відповідних кольорових складових:

$$r = \frac{R}{R+G+B},$$

$$g = \frac{G}{R+G+B},$$

$$b = \frac{B}{R+G+B},$$

де *r, g, b* – нормалізовані колірні канали.

Нормалізований RGB простір має властивість

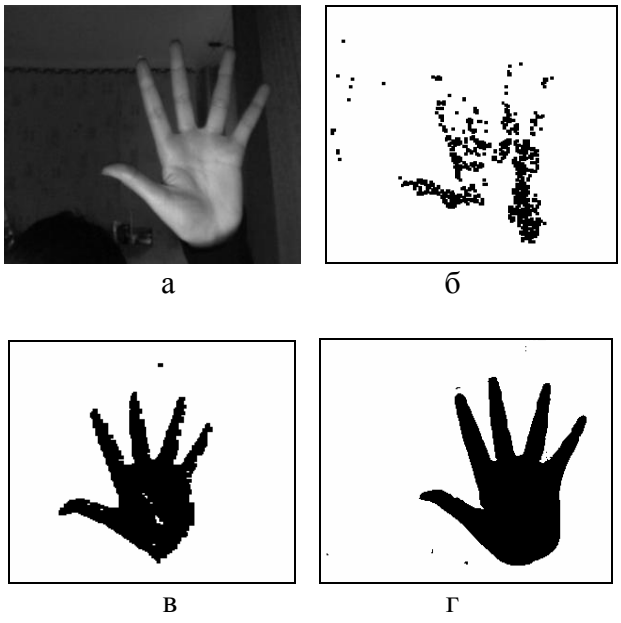
$$r + g + b = 1.$$

В роботі були досліджені алгоритми сегментації в різних колірних просторах (рис. 2). Колірний простір RGB виявився самим нестійким до змін освітлення, оскільки в цій моделі представлення компонента яскравості не відокремлена від колірної компоненти. Нормалізований RGB [1] простір частково допомагає уникнути залежності від освітлення. У просторі HSV [2] компонента *V* відповідає за яскравість зображення і, якщо проводити аналіз тільки *H* і *S* каналу, то ре-

зультати будуть менш залежні від освітлення, але повністю позбавитися від нього, особливо при роботі з камерою, неможливо.

Слід відзначити, якщо фон містить області однакові з кольором шкіри, то результати сегментації містять хибні сегменти. Останні можна вибракувати шляхом процедури верифікації, яка передбачає аналіз площі одержаних областей та їх форми [3].

Як з'ясувалося в результаті моделювання, в просторі *HSV* самим інформативним є канал *S*, а в нормалізованому *RGB* канали *R* і *G* несуть приблизно рівну кількість інформації і, якщо поєднати два канали *S* і *R*, то можна отримати стійкіші результати сегментації (рис. 2).

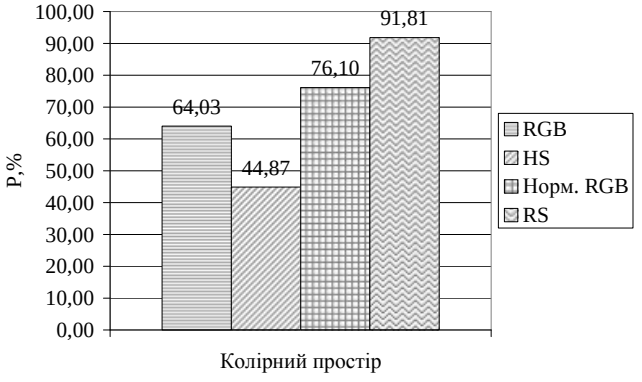


а б в г

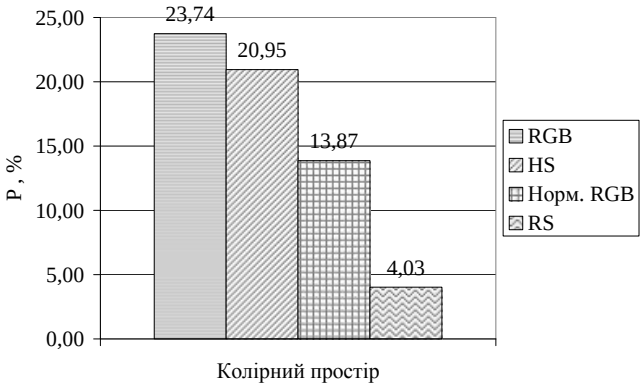
Рис.2. Сегментація при недостатній освітленості: початкове зображення (а) та результат його сегментації в *HSV* (б), нормалізованому *RGB* (в) просторі, результат сегментації з урахуванням *S* і *R* компонентів (г)

На рис. 3 показані усереднені результати моделювання процедури сегментації у різних колірних просторах при різних умовах освітлення (джерело освітлення спрямовано на об’єкт, знаходиться поза об’єктом і в умовах недостатнього освітлення – по 30 зображень для кожного типу освітлення).

Для розпізнавання жестів використовується метод розпізнавання з урахуванням ознак контуру.



а



б

Рис.3. Усереднені результати сегментації:
а – відсоток вірних рішень;
б – відсоток хибних рішень
(помилок II роду)

Доцільність використання контурних ознак пояснюється наступним:

зменшенням обсягу даних – кількість пікселів, які належать контуру об’єкта, набагато менше загальної кількості пікселів зображення;

можливістю забезпечити інваріантність освітленості, оскільки границі виявляються приблизно в одних і тих же місцях, незалежно від характеристик освітлення;

однозначним описом відповідного жесту на основі властивостей контуру руки, оскільки долоня і пальці людини мають специфічну форму.

В результаті аналізу існуючих підходів був розроблений алгоритм розпізнавання, який використовує характеристики контуру (області опуклості, які обчислюються за алгоритмом Sklansky [4] та їх дефекти) і співвідношення осей еліпса (або прямокутника), в який вписана долоня (рис.4). Для реалізації алгоритму необхідно на контурі долоні знай-

ти точки опуклостей контуру, які відповідають кінчикам пальців (A1, B1, C1, D1, E1) та точки дефектів опуклостей (A, B, C, D). Характерні ознаки обчислювались як відстань від прямої, що з'єднує дві сусідні точки опуклості, до найглибшої западини (рис. 4) контуру руки [5].

За допомогою наведених характерних ознак контуру руки можна легко визначити кількість пальців у полі зору. Однак для розпізнавання усіх десяти жестів (табл. 1) цього недостатньо. Тому використана додаткова ознака, яка дозволяє проводити нормування зображення: відношення більшої осі до меншої еліпса (або прямокутника), в який вписана долоня. Для різних людей таке відношення знаходиться приблизно в однакових діапазонах.

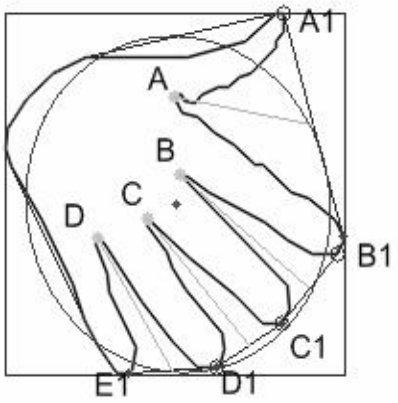


Рис.4. Розрахунок характерних ознак: точки опуклості (A1, B1, C1, D1, E1) і точки дефектів опуклості (A, B, C, D)

Реалізований алгоритм розпізнавання жестів було протестоване на 75 зображеннях, одержаних при різних умовах освітлення та на різному фоні. Якість розпізнавання жестів оцінювалась за процентним співвідношенням вірно розпізнаних жесті та поминок I та II роду. Результати тестування показали, що процент вірно розпізнаних жесті для розробленої системи розпізнавання жестів не менш 92%. Помилки I та II роду не перевищували у середньому 4% (табл. 2). Таким чином, перевагою системи є висока якість при високій швидкодії, що забезпечує її роботу у реальному часі.

2. Результати розпізнавання жестів

Позначення	Вірне розпізнавання, %	Помилки, %	
		I роду	II роду
«Start»	97,33	1,33	1,33
«Stop»	94,66	1,33	4,01
«Pause»	94,66	2,66	2,66
«Previous»	92,0	1,33	6,66
«Next»	94,66	0,0	5,33
«Volume Up»	93,33	4,0	2,66
«Volume Down»	96,0	0,0	4,0
«Close»	97,33	1,33	1,33
«Forward»	96,0	2,66	1,33
«Full Screen»	97,33	2,66	0,0

Висновок. Останнім часом інтерфейси на основі жестів рук поступово впроваджуються в реальні цифрові системи і користуються великим інтересом як у розробників, так і у користувачів. У статті запропоновано структуру людино-машинного інтерфейсу на основі жестів рук, розроблено програмне забезпечення основних процедур (захоплення зображення з веб-камери, локалізації області руки на зображенні, розпізнавання жестів та керування мультимедійним програмним забезпеченням) та проведено його дослідження. На основі розробленого людино-машинного інтерфейсу створено систему управління мультимедійними пристроями на основі жестів. Програмне забезпечення розроблено на мові C# із використанням бібліотеки комп'ютерного зору Open CV. Проведене тестування розробленої системи, яке показало такі результати: середня ймовірність вірного розпізнавання складає 95%, а помилки 1-го та 2-го роду не перевищують, у середньому, 3 %.

Дана розробка людино-машинного інтерфейсу на основі жестів здатна підняти взаємодію людини і комп'ютера на новий

рівень, тобто наблизити взаємодію людини і машини до рівня спілкування людей між собою у реальному житті. Її можна рекомендувати для застосування у побутовій техніці, медичному обладнанні, мобільних пристроях, робототехніці, тренажерах, ігрових консолях та ін.

Список використаних джерел

1. Skin detection in video under changing illumination conditions / Maricor Soriano, Birgitta Martinkauppi, Sami Huovinen // Proc. of International Conference on Pattern Recognition – vol.1. – 2000. – С. 839-842.

2. Fast and Accurate Hand Pose Detection for Human-Robot Interaction / Luis Antón-Canalís, Elena Sánchez-Nielsen, Modesto Castrillón Santana // IbPRIA – 2005. – С.553-560.

3. Skin Detection using Neighborhood Information / Ruiz-del-Solar J., Verschae R. // IEEE International Conf. on Image Processing – vol. 1 – 2004. – С.207-210.

4. Graham R., Yao F. Finding the convex hull of a simple polygon : J. Algorithms 4 / R. Graham, F. Yao. 1983. – 324 с.

5. Bradsky G., Kaehler A. Learning OpenCV – O'Reilly. 2008. – 555 с.

Отримано 30.06.2010

Антощук
Светлана Григорівна,
д-р техн.наук,
професор, Одеск. нац.
політехн. ун-та
e-mail:asg@ics.opu.ua

Яровенко
Ольга Сергіївна,
мол. наук. співробітник
ННЦ «ІТ-інкубатор»,
Одес. нац. політехн. ун-та