

УДК 621.307

В.Г. Абакумов, д-р техн. наук¹, О.Ю. Ломакіна¹, О.Б. Яровенко²

Застосування жестів рук при людино-машинному інтерфейсі

У статті розглядається можливість застосування жестів рук як пристрій введення інформації при людино-машинному інтерфейсі та проводиться дослідження з коротким аналізом роботи запропонованої системи керування жестами.

This article deals with the possibility of using hand gestures as an input device for human-machine interface. Investigation and a brief analysis of the proposed gesture control system are considered in detail.

Ключові слова: система управління жестами (СУЖ), інтерфейс, яскравість, колірний простір, колірна модель, сегментація.

Вступ

Перспективним напрямком розвитку інформаційних технологій є розробка нових способів забезпечення інтерфейсу людина-машина. Перед розробниками подібних інтерфейсів ставиться задача використання природних для людини способів спілкування з комп'ютерами. Враховуючи усі можливі перешкоди та наявність шумів в оточуючому середовищі, перевага надається системам на основі комп'ютерного зору. Особливо перспективними для побудови інтерфейсів управління програмним та апаратним забезпеченням комп'ютерів є жести [1]. Перевага такого інтерфейсу полягає у тому, що жести дозволяють розширити можливості інтерфейсу

для людей з вадами слуху і мови, та забезпечити дистанційне управління різними побутовими пристроями, хоча при їх створенні існує ряд не повністю вирішених задач, які розглядаються у статті.

Процес розпізнавання жестів

Розглянемо основні етапи процесу розпізнавання жестів (рис. 1).

Для реалізації основних функцій управління система управління жестами (СУЖ) може включати набір з 10 жестів (табл. 1).

- Робота системи містить ряд основних етапів:
- захоплення зображення - здійснюється за допомогою веб-камери, що підключена до ПК і безперервно передає в програму послідовність кадрів в реальному часі (25 кадрів/сек);
 - локалізація руки - полягає у знаходженні області руки на зображенні захопленому з веб-камери. Звичайні веб-камери не відзначаються якісним зображенням, у зв'язку з чим, спочатку, потрібно обробити зображення, щоб позбутися на ньому завад, які можуть бути викликані зовнішніми умовами, а саме: шуми на зображенні (щоб позбутися шумів здійснюється згладжування зображення) і перепади освітлення (для цього проводиться вирівнювання яскравості зображення). Далі здійснюється перетворення колірних

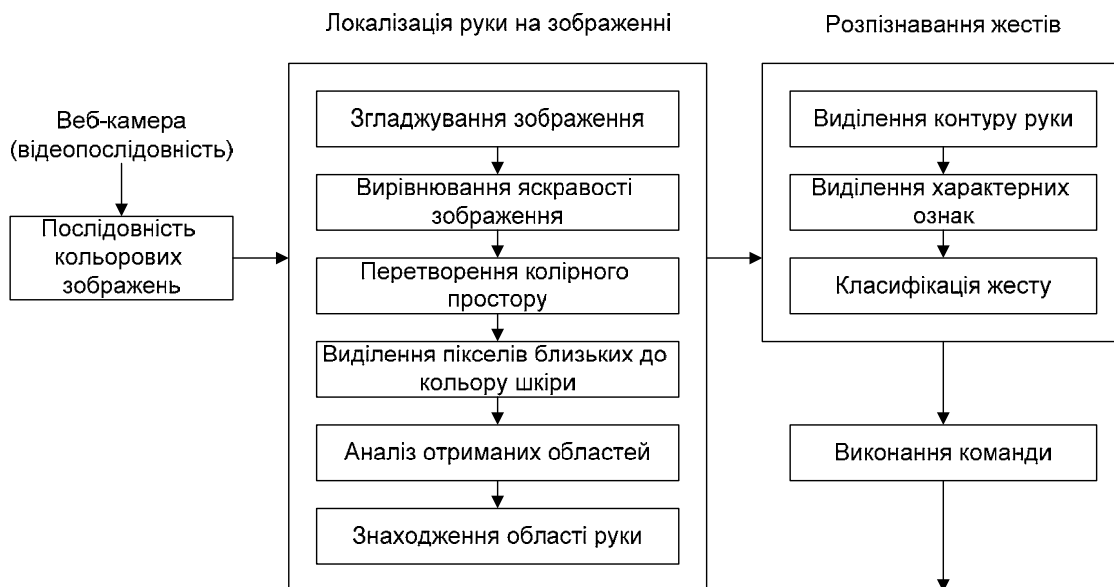


Рис. 1. Основні етапи процесу розпізнавання жестів

Таблиця 1. Набір жестів, які розпізнаються системою

N	Жест	Позначення	N	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»

просторів: RGB простір перетворюється на колірний простір HSV з виділенням каналу *S* та проводиться нормалізація RGB простору з виділенням каналу *R*. Для виділення руки на зображенні проводиться сегментація по кольору в каналах *RS*. Після закінчення сегментації аналізуються отримані області, з метою позбутися невірної сегментованих часток зображення і знайти область на зображенні, що відповідає області руки;

- розпізнавання жестів - на отриманому зображенні визначається контур руки і його характеристики, що дозволяє класифікувати жест;
- розпізнаний жест подає відповідну команду програмі.

Локалізація руки на зображенні

Для виділення потрібної області кисті руки було обрано простий в реалізації і виконуваний з високою швидкістю метод сегментації за кольором [2]. З метою вибору найкращої реалізації сегментації були вибрані RGB, HSV і нормалізований RGB колірні простори.

Розглянемо колірні простори, які використовуються при моделюванні:

а) RGB колірний простір - адитивна колірна модель (кольори виходять шляхом додавання (англ. addition) до чорного), що описує спосіб синтезу кольору для кольоровідтворення. Сегментація шкіри в колірному просторі RGB вважається самою нестійкою і сильно залежить від зовнішніх умов наприклад, освітлення;

б) HSV колірний простір - колірна модель, в якій координатами кольору являються: 1) hue -

колірний тон, який варіюється в межах 0 - 360°, проте іноді приводиться до діапазону 0 - 100 (чи 0-1); 2) saturation – насиченість, що варіюється в межах 0 - 100 (чи 0 - 1); 3) value (значення кольору) - яскравість, що задається в межах 0-100 (чи 0-1).

Не зважаючи на те, що ця колірна модель дозволяє зменшити вплив освітленості і виділити ділянки шкіри краще, ніж при використанні колірної моделі RGB, вона недостатньо стійка до перепадів освітленості і при роботі з веб-камерою з низьким розрізненням показує не кращі результати.

в) нормалізований RGB - це представлення, яке легко отримати процедурою нормалізації RGB величин.

Нормалізований RGB простір має властивість:

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

Третя компонента *b* не містить ніякої важливої інформації і може бути відкинута, тим самим зменшуючи розмірність простору. Компоненти, що залишилися, часто називають "чистими кольорами", оскільки залежність *r* і *g* компонент від освітленості зменшена нормалізацією. Особливістю цього представлення є те, що окрім незалежності від навколишнього освітлення, нормалізований RGB незмінний (при певних допущеннях, оскільки неможливо повністю уникнути впливу освітлення) до зміни орієнтації поверхні до джерела світла.

Теоретично, пікселі, які належать до області руки повинні бути близькими по кольору до кольору шкіри і належати певному діапазону на гістограмі [3].

Для перевірки цього на практиці були захоплені п'ять зображень з камери і на цих зображеннях область руки була виділена вручну. Потім були побудовані гістограми цих зображень. В результаті було встановлено, що пікселі, які належать шкірі, знаходяться в діапазоні яскравості від 80 до 130 в R-каналі, а в G-каналі від 60 до 95. Виявилось, що B-канал не дуже інформативний для сегментації шкіри, тому далі інформація цього каналу не використовувалась. Це допомогло зменшити розмірність і кількість обчислень.

Тестування розроблених функцій проводилось на 3-х вибірках при різних умовах освіт-

лення: - штучне освітлення з розміщенням джерела світла перед об'єктом; - штучне освітлення з розміщенням джерела світла за об'єктом; - в умовах недостатньої освітленості (наприклад, екран монітора).

Об'єднавши результати по всім трьом тестам (табл. 2), та проаналізувавши їх, можна зробити висновок, що сегментація, яка проводиться в R (R канал у нормалізованому RGB просторі) і S (S канал HSV колірного простору) каналах є найбільш стійкою до змін освітлення, а також має найменший відсоток помилок 2-го роду.

Таблиця 2. Середній результат сегментації по всім тестам

Колірний простір	Точність розпізнавання, %	Помилки 1-го роду, %	Помилки 2-го роду, %
RGB	64,03	35,97	23,74
HS	44,87	55,13	20,95
Нормалізований RGB	76,1	23,9	13,87
RS	91,81	8,19	4,03

В результаті етапу моделювання колірний простір RGB виявився самим нестійким до змін освітлення, оскільки в цій моделі представлення компонента яскравості не відокремлена від колірної компоненти. Нормалізований RGB простір частково допомагає уникнути залежності від освітлення. У просторі HSV компонента V відповідає за яскравість зображення і, якщо проводити аналіз тільки H і S каналу, то результати менше залежатимуть від освітлення, але повністю позбавитися від нього, особливо при роботі з камерою, неможливо.

Також, якщо фон містить області співпадаючі з кольором шкіри, вони можуть бути невірно класифіковані. З такими помилками можна впоратися, перевіряючи площу області або її форму.

Розпізнавання жестів

Для розпізнавання жестів використовується метод розпізнавання з урахуванням характеристик і властивостей контуру [4]. Долоня людини і пальці мають специфічну форму, тому можна знайти властивості контуру, які будуть однозначно описувати жест.

В результаті аналізу існуючих методів і алгоритмів був розроблений алгоритм, який використовує характеристики контуру (області опуклості та їх дефекти) і співвідношення осей еліпса прямокутника, в який вписана долонь.

Для реалізації цього алгоритму необхідно знайти контур долоні, а потім точки опуклостей контуру, які відповідають кінчикам пальців (рис.

3 точки A1, B1, C1, D1, E1). Потім точки опуклостей контуру з'єднуються і утворюють область, яку займає долонь.

Далі необхідно обчислити дефекти опуклості (на рис. 4 області A, B, C, D, E, F, G, H). Вони обчислюються як відстань від прямої, що з'єднує дві сусідні точки опуклості, до найглибшої западини (рис. 3 точки A, B, C, D) контуру руки.

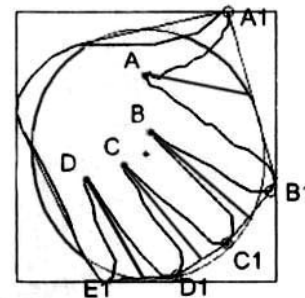


Рис. 3. Фрагмент результату моделювання: контур опуклості (contour convexity), точки опуклості (A1, B1, C1, D1, E1) і точки дефектів опуклості (A, B, C, D)

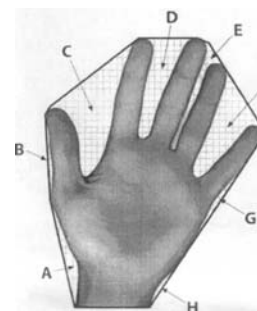


Рис. 4. Контур опуклості долоні (contour convexity) і області дефектів опуклості (A, B, C, D, E, F, G, H)

За допомогою властивостей опуклостей контуру можна легко вичислити кількість пальців, однак це є недостатньою ознакою для розпізнавання усіх десяти жестів. Тому використовується додаткова ознака: співвідношення осей (більшої осі до меншої) еліпса (або прямокутника), в який вписана долоня.

Таблиця 3. Набір жестів і їх властивості

№	Позначення жесту	Кількість пальців
1	«Start»	0
2	«Stop»	0
J	«Pause»	1
4	«Previous»	1
5	«Next»	2
6	«Volume Up»	1
7	«Volume Down»	2
8	«Close»	4
9	«Forward»	J
10	«Full Screen»	5

У програмній реалізації підсистеми був створений клас *GestureRecognition*, який дозволяє провести сегментацію зображення у потрібних каналах.

Реалізований алгоритм був протестований на 75 зображеннях рук. Результати тестування представлені в табл. 4.

Таблиця 4. Результати розпізнавання жестів

Позначення	Вірно розпізнані, %	Помилки 1-го роду, %	Помилки 2-го 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

Висновки

Наведено структуру та досліджено роботу системи управління жестами як перспективного інтерфейсу людина-машина, що містить наступні етапи: 1) захоплення зображення; 2) локалізація руки; 3) розпізнавання жестів; 4) формування команд. Така система дозволяє

реалізувати дистанційне управління автоматизованими пристроями не тільки звичайними користувачами, а й людьми з вадами слуху і мови.

Обраний для локалізації руки на зображенні метод сегментації за кольором, забезпечує достатню швидкість для того, щоб його програмна реалізація могла працювати в реальному часі. Згідно аналізу результатів тестування за різних умов освітлення, яке виконувалося над послідовністю зображень з камери, було визначено, що сегментація, яка проводиться в R (r канал у нормалізованому RGB просторі) і S (S канал HSV колірного простору) каналах є найбільш стійкою до змін освітлення, а також має найменший відсоток помилок 2-го роду. Це свідчить про доцільність застосування саме цього методу сегментації в R і S каналах різних колірних просторів.

Для розпізнавання жестів пропонується алгоритм, який використовує характеристики контуру (області опуклості та їх дефекти) і співвідношення осей еліпса прямокутника, в який вписана долоня. Відповідно до результатів тестування запропонованого алгоритму на 75 зображеннях рук помилки розпізнавання 2-го роду складають не більше 6,5%, що свідчить про надійність та завадостійкість даного методу.

Література

1. В.Г. Абакумов, Е.Ю. Ломакина. Автоматическое распознавание жестов в интеллектуальных системах // Материалы конференции "Искусственный интеллект. Интеллектуальные системы 2010". – 2010. – С. 269-272 с.
2. Skin detection in video under changing illumination conditions / Maricor Soriano, Birgitta Martinkauppi, Sami Huovinen // Proc. of International Conference on Pattern Recognition. – 2000. –vol.1.– p. 839-842.
3. Skin Detection using Neighborhood Information / Ruiz-del-Solar J., Verschae R. // IEEE International Conf. on Image Processing.– 2004.– vol.1. –p.207-210.
4. Fast and Accurate Hand Pose Detection for Human-Robot Interaction / Luis Antón-Canalís, Elena Sánchez-Nielsen, Modesto Castrillón Santana // –2005. –p.553-560.

¹ Національний технічний університет України «Київський політехнічний інститут»

³ Одеський національний політехнічний університет