

ІНТЕРАКТИВНИЙ ЖЕСТОВИЙ ІНТЕРФЕЙС ЛЮДИНА-КОМП'ЮТЕР

Ю.В. Крак^{1,2}, А.С. Тернов², Ю.В. Коваль¹, М.П. Лісняк²

¹Київський національний університет імені Тараса Шевченка

²Інститут кібернетики ім. В.М. Глушкова НАН України,

01601, Київ, вул. Володимирська, 64/13

Yuri.krak@gmail.com

Abstract

A common architecture of real-time sign language analysis system was discussed in the article. The main requirements to the system, main principles of modules and their structure were described. Integration of modules in main system utilizing high-speed TCP/IP network was proposed.

1. Вступ та постановка задачі

Сучасний науково-технічний прогрес характеризується принципово новими можливостями використання інформації, комп'ютерної техніки і комп'ютерних технологій в усіх сферах життєдіяльності. Це вимагає створення засобів взаємодії між користувачем і комп'ютерним середовищем у вигляді та формі, максимально зручній та природній для людини. Одним із засобів, звичних для людини, є передача інформації за допомогою зорової складової, а саме через рухи рук, тіла, міміки тощо [1]. Іншим важливим напрямом використання візуальної інформації є її передача і сприйняття людьми з вадами слуху засобами жестової мови спілкування [2].

Існування великої кількості методів і підходів до аналізу зорової інформації (математичних методів обробки, моделювання, виділення характеристичних ознак, класифікації, кластеризації, розпізнавання візуальної інформації) [3-5] викликає потребу в систематизації таких наробок та їх інтеграції в єдину систему. Це дозволить розглядати проблему створення технології аналізу інформації, що передається за допомогою рухів, в площині можливості застосування сучасних комп'ютерних засобів в розв'язанні проблеми комп'ютерного моделювання та аналізу жестової мови. В цьому контексті важливими напрямками досліджень слід вважати:

1. Розробка інформаційних моделей жестикуляції для вирішення задачі розпізнавання елементів жестової комунікації.

2. Формування критеріїв та вимог до технології аналізу інформації що передається за допомогою рухів, в контексті задачі розпізнавання чи ідентифікації структурних елементів жестової мови. Розроблення на їх основі загальної концепції такої інформаційної технології.

В даній роботі увага приділена технічному аспекту пов'язаному з питанням створення технології аналізу інформації, що передається за допомогою рухів. Проведений аналіз літературних джерел визначив напрямок досліджень і постановку задачі.

Необхідно розробити загальну архітектуру системи аналізу інформації основаної на існуючій інформаційній моделі жестової комунікації [6] і яка має задовольняти певним вимогам. Також запропонувати механізми реалізації складових такої системи і принципи їх дії. Вимоги до системи:

- функціонування відбувається в режимі реального часу;
- можлива реалізація паралельної обробки даних.
- використання модульного принципу розробки, свобода розширення.

2. Архітектура та деталі реалізації

Виходячи зі складності системи та можливого її розширення у майбутньому функціональними модулями, запропоновано загальну архітектуру системи, яка подана наступною схемою (рис. 1). Потоки інформації ідуть зверху вниз.

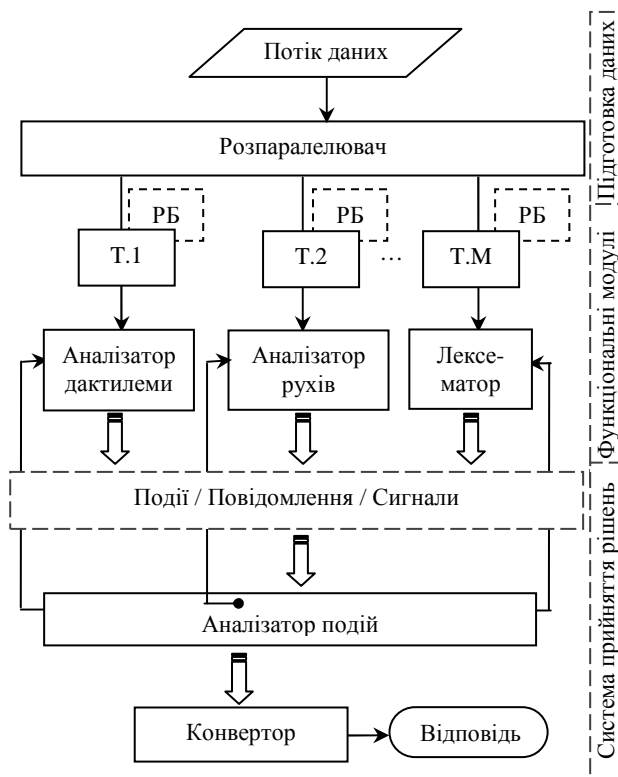


Рисунок 1. Загальна архітектура системи

Архітектура системи логічно розбита на три області:

- 1) частина системи, яка займається підготовкою даних для подальшої обробки і аналізу;
 - 2) функціональні та логічні модулі;
 - 3) система прийняття рішень.
- Окремо розглядається система керування і синхронізації модулів.

Функціонування системи, структура якої буде складатися з різних модулів кожен з яких має свою швидкодію, передбачає синхронізацію потоків для узгодження обчислень. Тому слід розглянути та визначити таке поняття як загальний час системи або системний час, який насправді в логіці системи буде прив'язаний до якогось типу події або даних, а не до певної сталої часової одиниці. В даному випадку за абстрактну одиницю системного часу береться найбільша неподільна одиниця інформації (наприклад, у випадку системи розпізнавання дактильної абетки це буде кадр відео).

Завдання, які покладені на Блок «Розпаралелювач», зводяться до одночасного і миттєвого (в ідеальному випадку) надання всім функціональним модулям системи даних для обробки. Буде реалізовано у вигляді системи керування потоками даних.

Додаткові перетворювачі даних Т.І, І=1,М призначені для підготовки інформації (її попередня обробка) для відповідного функціонального модуля системи: приведення її до стандартну вхідної інформації для модуля, якщо такий визначено. Кожен модуль має власний резервний буфер (РБ) даних.

Функціональні модулі системи – певні незалежні обробники інформації мета яких, інформувати систему про зміни в даних (аналіз інформаційних одиниць або «лексема») і при необхідності надання параметрів цих змін. Кількість модулів може варіюватись в залежності від набору алгоритмів обробки даних які є в наявності і обмежена лише технічними характеристиками мережі.

В єдину систему функціональні модулі поєднуються за допомогою швидкісної (одногігабітної) мережі TCP/IP. Це розв'язує задачу комунікації і робить можливим керування функціональними модулями для синхронізації їх роботи. Так само інші блоки системи поєднані в єдину мережу TCP/IP.

Кожний окремий модуль, згідно[7], є агентом. Для отримання потоку даних з метою подальшої обробки/аналізу, а також можливості відправки та отримання службових повідомлень він має пройти реєстрацію в системі, а саме в модулі розпаралелювання даних. Більшість модулів системи одночасно можуть виступати у вигляді або сервера, або клієнта в рамках свого незалежного каналу зв'язку (TCP-сесія), що говорить про наявність змішаної клієнт-серверної архітектури. Сесії в системі мають бути короткими для зменшення навантаження на мережу. Час їх життя пов'язано лише з часом передачі повідомлення і коли повідомлення передано, то сесія закривається.

2.1. Вхідні дані

За рішення системи відповідає аналізатор подій, який містить систему прийняття рішень. Основою такої

системи є таблиця прийняття рішень, яка будується з множини подій, яку, в свою чергу, породжують модулі системи.

Потік даних формується з поточкових даних, які отримані з пристроїв реалізації комп'ютерного зору. Наприклад, вхідними даними для системи розпізнавання дактильної абетки на початковому етапі досліджень буде потік відео (покадровий), який містить зміни конфігурації руки людини. У більш загальному випадку передбачено використовувати і інші джерела інформації (давачі\сенсори руху по типу Kinect [8] або LeapMotion [9] та інше). В цих випадках для кожного джерела інформації має бути присутній додатковий модуль розпаралелювання відповідних даних синхронізованих між собою у загальному часі системи.

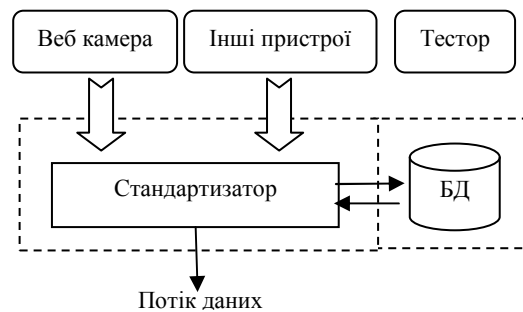


Рисунок 2. Система керування потоками даних

Стандартизатор потоку проводить лише загальну попередню обробку даних приводячи їх до єдиного стандарту (наприклад: за розміром, параметрами освітленості та інше).

Тестор – функціональний модуль, ініціатор початку роботи системи. Його завдання буде надавати системі данні зі сховища даних, емуляючи (імітуючи) поведінку людини.

2.2. Технічні вимоги і буфер даних

З огляду на вимоги загальної архітектури було сформульовано певний набір технічних вимог:

- 1) для кожного функціонального модуля системи має бути виділено свій процесор (реалізація вимоги паралельної обробки даних);
- 2) швидкісний обробник подій (час формування відправлення та отримання повідомлень має бути миттєвим у порівнянні з часом обробки одиниці інформації);
- 3) керування функціональними та логічними модулями за допомогою сигналів(подій) повідомлень;
- 4) кожен функціональний модуль роботи з даними має свій «буфер кадрів», розмір і призначення якого обумовлюється видом задачі яка вирішується цим модулем.

Буфер має властивість «вперед-назад»: дозволяє призупиняти чи прискорювати процес обчислень синхронно у відповідності до заданої у вигляді таймера часової вісі.

В системі передбачається використовувати кільцеві буфери даних на 3-4 лексеми: так перша лексема зчитується, зберігається у буфер і далі починається

зчитування наступної, не знаючи поки нічого про її початок чи кінець.

У випадках коли на момент ініціалізації лексематором обробки загальною системою потоку даних, поточний стан не є початком лексеми, потрібно відсікти кадри які віднесені до поточної лексеми – для кожного модуля приймається рішення в залежності від алгоритму обробки даних реалізованого для даного модуля.

Тому буфер даних є незалежним і власним для кожного логічного модуля системи. Для оцінки розміру буферу розглядається величина, яка дорівнює розміру чотирьох лексем максимальної довжини.

2.3. Режими функціонування системи

З точки зору практичної цінності система має декілька режимів різних за своїм призначенням і відповідно за логікою і послідовністю роботи структурних елементів: 1) Робочий режим; 2) режим навчання/тестовий режим; 3) режим калібрування. Логіку роботи системи в кожному режимі подано у вигляді діаграм переходів.

Режим навчання – отримання і збереження потоків даних з одночасним визначення параметрів цих даних в автоматизованому режимі.

Робочий режим – це основний режим функціонування системи за якого система реалізує своє призначення, а саме «аналізатор жестового мовлення» у реальному часі.

Тестовий – потоки даних у систему потрапляють не з оточуючого середовища, а зі сховища даних.

3. Оцінка потужності системи

Для визначення можливості використання програмної реалізації на окремому модулі системи потрібно визначити певний критерій, який засновано на єдиному принципі (на незалежних технічних характеристиках) та за яким можна враховувати цю оцінку ще на етапі розробки програмного забезпечення.

При незалежності обчислювальних ресурсів модуля від інших елементів системи важливою характеристикою є час виконання конкретної програмної реалізації на даному модулі. Тому для функціонування системи в режимі реального часу є критичним, щоб час виконання програмної реалізації не перевищував певної величини. Верхня оцінка часу виконання одного повного циклу обробки одиниці інформації для даної програмної реалізації буде мати такий вигляд: $t' = n \cdot t^{\max} / n'$, де n – продуктивність процесора незалежного модуля у MIPS (мільйон операцій за секунду), t^{\max} – максимально допустимий час виконання програмної реалізації на заданому модулі системи, n' – продуктивність тестового модуля на якому проводяться випробування.

4. Висновки

В результаті проведених досліджень була запропонована і розроблена загальна архітектура системи аналізу жестового мовлення в реальному часі, визначено вимоги

до такої системи реального часу, основні модулі такої системи та принципи їх взаємодії (підсистема комунікації модулів на основі високошвидкісної мережі). Дана система інтегрується до систем повноцінної підтримки інтерактивної взаємодії людини-носія жестової мови з комп'ютером у природній для людини спосіб [10].

Це дозволить розвинути напрямок досліджень пов'язаний з побудовою віртуального сурдоперекладача де однією з важливих проблем є перетворення висловлень жестової мови у зрозумілу для системи форму з подальшою трансформацією у вербальну мову. В цьому випадку віртуалізація досягається за рахунок використання клону (тривимірної моделі реальної людини сурдоперекладача) в якості інформаційної моделі та засобу комп'ютерного моделювання жестової мови.

Подальші дослідження направлені на реалізацію запропонованих функціональних модулів та створення загальної системи розпізнавання жестової мови.

5. Література

- [1] Кривонос Ю., Крак Ю., Бармак О., Тернов А., Троценко Б. Моделирование реалистических движений и мимики для задач визуализации жестовой информации. In book: Natural and Artificial Intelligence/ Ed. by K.Markov, V.Velychko, O.Voloshin. C/o Iusautor, Sofia. – 2010. – P. 137-143.
- [2] Ткачева Т. Жестовый язык глухих как особая знаковая коммуникативная система (на материале русского и французского жестовых языков) / Ткачева Т., Ахмедышева Н. // Материалы Международной научно-практической конференции «Приоритеты и интересы современного общества». – Астрахань: АГУ, 2010. – С. 361-365.
- [3] Rick Parent Computer Animation. Algorithms and techniques / Parent Rick. The Morgan Kaufmann Series in Computer Graphics, 2nd Edition. – San Francisco: Morgan Kaufmann Publishers Inc., 2007. – 624 p.
- [4] Panin G. "Model-based Visual Tracking: the OpenTL Framework" / Giorgio Panin – Peter Ratner; John Wiley & Sons Inc., New York, 2011 – 318 p.
- [5] Шапиро Л. Компьютерное зрение / Шапиро Л., Стокман Дж. – М.: БИНОМ, 2006 – 752 с.
- [6] Кривонос Ю.Г. Формалізація жестоутворення за допомогою трьохмірної моделі людини для української жестової мови / Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак, Г.М. Єфімов // Проблеми програмування. – 2012, №2-3. – С. 398-405.
- [7] Летичевський О.А. В.М. Глушков і сучасна інформатика (від теорії автоматів до когнітивних архітектур) / О.А. Летичевський // Вісн. НАН України. – 2013. – № 8. – С. 21-33.
- [8] Kinect for Windows Sensor Components and Specifications. Електронний ресурс. Режим доступу: <http://msdn.microsoft.com/en-us/library/jj131033.aspx>
- [9] Leap Motion Controller. Електронний ресурс. Режим доступу: <https://www.leapmotion.com/product>
- [10] Крак Ю.В., Тернов А.С., Лісняк М.П. Розробка архітектури та основних інструментів комп'ютерної анімації для побудови системи синтезу жестової мови Штучний інтелект. – 2013. – №3(61). – С.147-153.