

ОПЕРАЦІЙНА СИСТЕМА ОБРАЗНОГО КОМП'ЮТЕРА¹

Тарас Вінцюк

Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій та систем
40 проспект Академіка Глушкова, Київ 03680, Україна
Тел.: +380 44 526-4356 Факс: +380 44 526-4356
vintsiuk@uasoiro.org.ua

ABSTRACT

Taras K. Vintsiuk. Operative System for Pattern Computer. Pattern computer (PatCom) is such intellectual cybernetic system that provides a functional simulation of intelligent, mainly subconscious activity of all living and a human being particularly. In PatCom this activity is related to image, sound and other patterns perception, scene analysis, action and movement planning, generalisation of observations, discovering of regularities, prediction, decision making etc. Pattern computer operates with patterns and other complex notions. It actualises both pattern and logical reasoning.

Pattern computer is a parallel system. It has several information perception channels (acoustic, visual, scential) that is multimodal perception, pattern operation system, improved human-machine interface. Unlike usual computer, which is based on a rapid arithmetic-logical processor and a large RAM space, PatCom is grounded on external world models including physical, geometry, acustical, language, linguistic, semantical ones.

Here a so-called pattern operation system for a pattern computer is debated. This system synchronizes an intellectual processing of multimodal information, which come by different perception channels, and operating by external world models fulfils a complex semantic interpretation for whole received information.

ВСТУП

З часу започаткування Державної науково-технічної програми “Образний комп’ютер” пройшло п’ять років. Є певні здобутки, є й розчарування. Настав час оглянутися, підбити підсумки, визначити подальші кроки.

Ініціатори програми вважали, що 10-річний цикл робіт повинен завершитись розробленням технічного завдання на створення образного комп’ютера. Тоді як їх опоненти (а це, в основному, спеціалісти з обчислювальної техніки) заперечували: дайте, мовляв, нам технічне завдання і ми вам швидко зробимо який завгодно комп’ютер.

Досвід п’яти років тільки підтвердив всю складність проблем, що виникають, і виявив всі труднощі написання технічного завдання. Водночас цей досвід показав, що уже на часі створення початкової платформи образного комп’ютера, а, значить, і технічного завдання на неї.

Тому важливо ще раз повернутись до основних визначень образного комп’ютера та ролі образної операційної системи в ньому.

ЩО Ж ТАКЕ ОБРАЗНИЙ КОМП'ЮТЕР

В роботах [1, 2] змістовно визначені два базових поняття: образний комп’ютер (ОК) та образна операційна система (ООС).

Образний комп’ютер — це така кібернетична система, в якій виконується функційне моделювання інтелектуальної, головно підсвідомої, діяльності людини та всього живого, що пов’язана зі сприйняттям зорових, слухових та інших образів, аналізом сцен та складних ситуацій, плануванням дій та рухів, узагальненням спостережень, встановленням закономірностей, прогнозуванням, прийняттям рішень тощо. ОК оперує образами та іншими складними поняттями, реалізує як **образне**, так і логічне мислення.

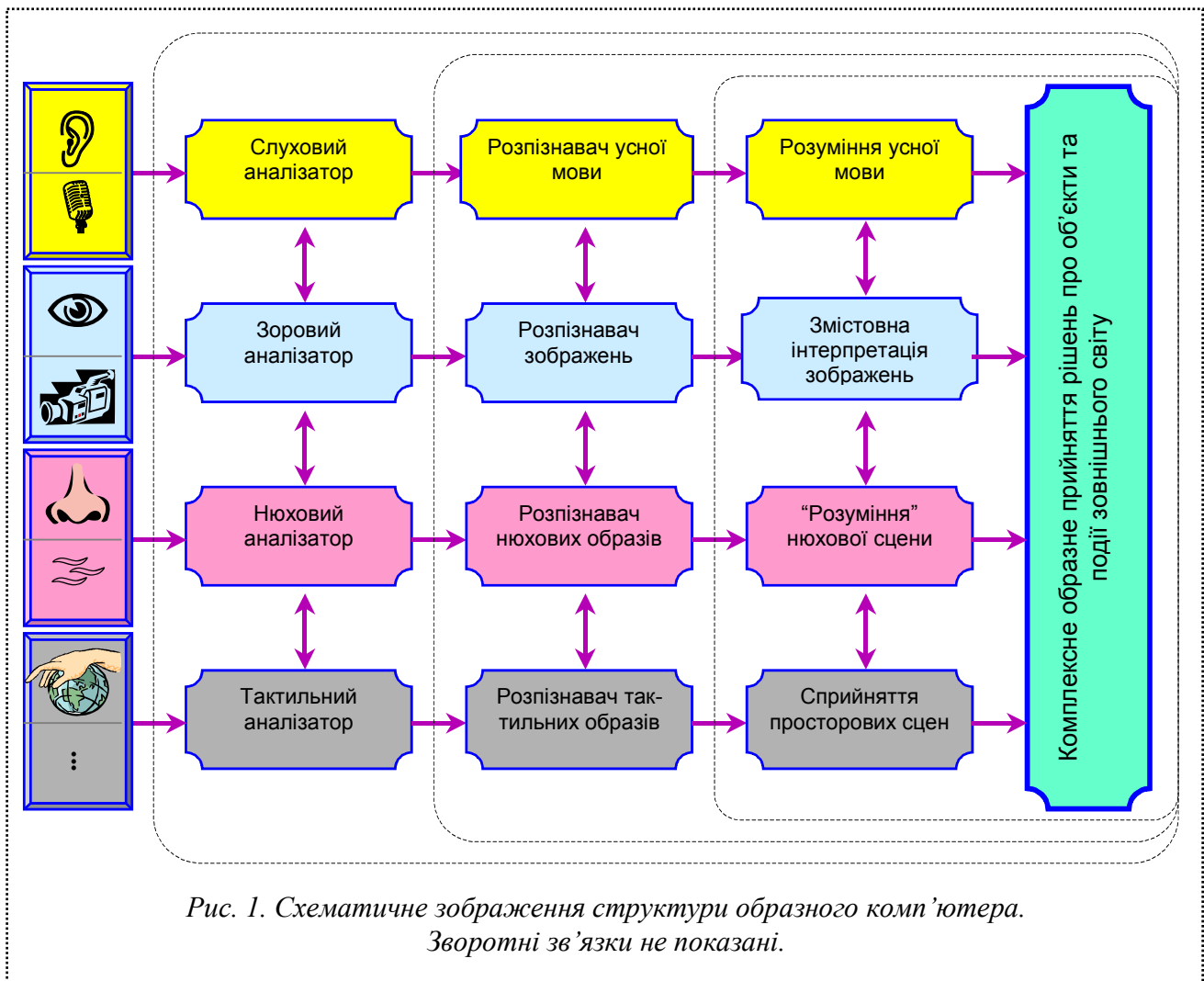
Образний комп’ютер є паралельною мультимодальною системою, яка має у своєму складі декілька каналів сприйняття інформації (слухової, зорової, текстової, смакової, нюхової тощо), **образну операційну систему**, моделі зовнішнього світу (в тому числі акустичну, оптичну, геометричну, лінгвістичну, семантичну, канонічних форм тощо), розвинений інтерфейс з людиною, засоби взаємодії з існуючими комп’ютерними та телекомунікаційними мережами.

Образна операційна система “синхронізує” оброблення інформації, що надходить різними каналами її сприйняття, та, оперуючи моделями зовнішнього світу, виконує комплексну семантичну інтерпретацію всієї отриманої інформації, організує остаточне прийняття рішень..

Отже, поняття ОК та ООС суттєво відрізняються від загально прийнятих понять “комп’ютер” та “операційна система”.

Так, у центрі звичайного комп’ютера знаходяться центральний процесор та пам’ять (відносно примітивні пристрої), а від периферійних пристроїв вимагаються (чи “справедливо” ?!) такі, наприклад, високоінтелектуальні функції як введення-виведення (сприйняття-генерація) рукописної або мовленнєвої інформації. Операційна система звичайного комп’ютера забезпечує (чи

¹ Робота виконана в рамках Контракту № ОК_2005_2 за ДНТП “Образний комп’ютер”



успішно?!) взаємодію центрального процесора та пам'яті (примітивних пристроїв) з периферійними пристроями та, опосередковано, з людиною-користувачем.

Пропонується поглянути на проблему підвищення інтелектуальності комп'ютерів дещо інакше.

На відміну від звичайного комп'ютера у центрі образного комп'ютера поставимо моделі зовнішнього світу з усіма його об'єктами, явищами та їх проявами, а від образної операційної системи вимагатимемо як інтелектуального прийняття рішень, так й інтелектуальної взаємодії з людьми.

СТРУКТУРА ОБРАЗНОГО КОМП'ЮТЕРА

На рис. 1 подано схематичне зображення ОК. Наведені чотири канали сприйняття інформації: слуховий, зоровий, нюховий та тактильний. В кожному каналі виділено перцептор, аналізатор, розпізнавач та інтерпретатор.

Пунктирними лініями зображається ієрархія рівнів оброблення інформації. Найнижчий рівень

стосується первинного аналізу, найвищий — змістовної інтерпретації та розуміння.

Образна операційна система виконує комплексну, узгоджену за всіма каналами, інтерпретацію та образне прийняття рішень про об'єкти та події зовнішнього світу. Ієрархії в ООС відображено вкладеними пунктирними фігурами.

Взаємодії між каналами показані вертикальними стрілками. Зворотні ж зв'язки в каналах та між каналами не відмічені.

МУЛЬТИМОДАЛЬНІСТЬ В ОК

В [1, 2] запропоновані два підходи (дві моделі) до створення образного комп'ютера: генеративна модель та багаторівнева багатозначна модель.

Ці моделі ґрунтуються на конструктивній реалізації загально визначених принципів оброблення інформації в живій природі, техніці та суспільстві, таких як: мультимодальність, аналіз через синтез, індуктивне та дедуктивне виведення, генерація та направлений перебір варіантів, зворотній зв'язок, використання апріорної інформації, адаптація, навчання та самонавчання, розпізнавання образів,

побудова моделей зовнішнього світу та розумової діяльності.

Тут термін “мультиmodalність” має більш глибокий смисл, ніж просто використання двох чи більше modalностей (зір, слух, дотик, запах) сприйняття сенсорної інформації. Мульти-modalність у розпізнаванні образів взагалі означає сприйняття та оброблення не тільки різномірної, а й однорідної інформації, що надходить багатьма каналами.

Розрізнятимемо modalності двох типів: синхронного та асинхронного.

Оскільки в образному комп’ютері на образну операційну систему покладається комплексне, узгоджене за всіма каналами, прийняття рішень, то синхронні чи синхронізовані modalності передбачають одночасне, синхронізоване на всіх рівнях ієрархії, оброблення взаємопов’язаних потоків інформації. Наприклад, при розпізнаванні мовлення від особи, яка говорить, потрібно синхронізовано сприймати як фонемні (акустичне явище), так і віземи (візажами) – зображення обличчя, яке говорить.

При асинхронному сприйнятті мульти-modalної інформації узгоджуються тільки кінцеві рішення. Наприклад, розпізнавання особи за обличчям (зоровий канал) та за голосом (акустичний канал). Якщо одночасно не аналізувати віземи то виходить, що лише на кінцевій стадії приймається узгоджене (“синхронізоване”) рішення про особу.

При прийнятті рішень на підставі інформації, що надходить мульти-modalними каналами, виходять з того, що сигнали у каналах за умов класів є статистично незалежними [3]:

$$P(x/j) = \prod_{k=1}^K P(x_k/j),$$

де $x = (x_1, x_2, \dots, x_k, \dots, x_K)$ – вхідна інформація у вигляді об’єднання інформації від K modalностей.

Вирішувальне правило засновується на формулі Байеса для $|J|$ класів:

$$\begin{aligned} j^*(x) &= \arg \max_{j \in J} \frac{P(j)P(x/j)}{P(x)} = \\ &= \arg \max_{j \in J} \frac{P(j) \prod_{k=1}^K P(x_k/j)}{P(x)} = \\ &= \arg \max_{j \in J} \frac{P(j)}{P(x)} \prod_{k=1}^K \frac{P(x_k)P(j/x_k)}{P(j)} = \\ &= \arg \max_{j \in J} \frac{P(j)}{P(x)} \prod_{k=1}^K \frac{P(j/x_k)}{P(j)}, \end{aligned}$$

де J – множина класів при мульти-modalному розпізнаванні.

Як впливає з попередньої формули, “мульти-modalне” рішення про клас приймається на підставі середньогометричної, за всіма modalностями, апостеріорної ймовірності класу.

При прийнятті рішень загальний ефект, згідно з теорією розпізнавання образів, такий: при додаванні нових modalностей помилка розпізнавання не збільшується, а головню зменшується, а при викиданні однієї якої-небудь modalності помилка розпізнавання не зменшується (головню збільшується). Причому ці властивості мають місце не тільки у випадку незалежних спостережень modalностей, але і для залежних (корельованих).

ГЕНЕРАТИВНА МОДЕЛЬ ДЛЯ БІМОДАЛЬНОГО РОЗПІЗНАВАННЯ МОВЛЕННЯ

Те, що образна операційна система грає роль супервізора, покажемо на прикладі біmodalної моделі розпізнавання мовленнєвих сигналів. Отже, нехай модель не тільки сприймає сигнал мікрофона, але й додатково слідкує за мімікою обличчя людини, яка говорить.

В трирівневій генеративній моделі автоматичного розпізнавання та розуміння мовленнєвого сигналу за канонічною формою передаваного смислу генератор семантично еквівалентних речень (ГСЕР) породжує всі можливі речення, що передають один й той самий смисл, визначений вхідною канонічною формою. На цьому рівні ієрархії зображення не породжуються.

На другому рівні генератор модельних сигналів зв’язного мовлення (ГМСЗМ) ставить синхронно у відповідність прийнятому орфографічному текстові (віртуально породжує, генерує або синтезує) як всі можливі модельні сигнали злитого мовлення, так і всі відповідні можливі модельні динамічні зображення міміки обличчя та жестикуляції, які відображають розмаїті мовні сигнали та динамічні зображення, що відрізняються фонемним складом, темпом, інтенсивністю, інтонацією, моделюють індивідуальні фізичні та інші особливості мовця, його стану, його мовлення, міміки та жестикуляції.

В компараторі (К) модельні сигнали порівнюються з мовним сигналом та динамічним зображенням, що синхронно пред’являються для аналізу. Результат поточного порівняння використовується у синхронізованому зворотньому зв’язку для направленої відбору та пошуку найкращих модельних звукових сигналів та зображень. Останні аналізуються, тобто вказується, якому реченню (орфографічному текстові) він відповідає (текст можна надрукувати — автоматична машинка, що друкує під диктування) і/або отримати відповідну цьому текстові канонічну форму, що передається (автоматичне розуміння мовного сигналу аналізуються аналізуються — цей результат можна

використати для виконання дії, щоб задовольнити інтереси людини, яка говорить).

Водночас, при необхідності, може бути видана інформація про емоційний стапн людиниЮ що говорить. 3.

Отже, щоб синхронізувати генерацію та аналіз мовленнєвого сигналу та зображень міміки обличчя і таким чином підвищити достовірність розпізнавання та інтерпретації, треба побудувати адекватну бімодальну модель породження мовлення, яка враховує індивідуальні особливості мови, мовлення та поведінки людини (фізичні, фонетичні, лінгвістичні, семантичні тощо).

За цю синхронізацію відповідатиме образна операційна система, яка заснована на моделях зовнішнього світу та маніпуляціях ними.

ЩО ТАКЕ БАЗОВА ПЛАТФОРМА ОБРАЗНОГО КОМП'ЮТЕРА

На сьогодні розроблені окремі підсистеми образного комп'ютера, які є одноmodalними і стосуються окремо оброблення візуальної, мовленнєвої, звукової, текстової, біомедичної чи якоїсь іншої інформації. Всі вони тою чи іншою мірою оперують моделями зовнішнього світу, вивчаючи різні його аспекти.

Пора інтегрувати розробки, а саме:
1). Узагальнити моделі зовнішнього світу;
2). Розробити мову канонічних форм для опису понять та подій у зовнішньому світі;
3). Розробити образну операційну систему, що інтегруватиме поточні результати.

Сьогодні ясно, що не вдасться створити образний комп'ютер на базі персонального комп'ютера. ОК має іншу структуру та принципи функціонування.

Базова платформа ОК – це така апаратно-програмна модель, яка виконує мультимодальне інтелектуальне оброблення інформації в темпі її надходження. В основі базової платформи лежить образна операційна система.

До базової платформи ОК та її ООС ставляться такі вимоги.

ПРИКІНЦЕВІ ПОЛОЖЕННЯ. ВИМОГИ ДО ООС

1. **Одночасне введення-виведення та базове мультимодальне оброблення звукової та зорової інформації:** стереозвук, стереозір, символна інформація, сканування зображень, відтворення введених зображень (в тому числі об'ємних та стерео- зображень) на екрані, генерація звуків і зображень.
2. Ця вимога є першою і чи не найголовнішою. Так, більшість розробок одноmodalних систем працює за принципом: спершу сигнали вводяться,

потім аналізуються та розпізнаються, або навпаки – спершу генеруються, потім – виводяться.Тоді як процеси введення-виведення та оброблення інформації повинні виконуватись одночасно.

Реальний час та синхронізація – це принципово для базової платформи.

2. Сегментація та анотування потоків аудіо-, відео- та аудіовідео- інформації на різномодальні ділянки (мовлення, музика, шум, різні диктори, різні мови, друковані тексти, рукописні тексти, рисунки, малюнки, кінофільми, реклама тощо).

3. Пошук інформації за заданим змістом у мультимодальному потоці (за ключовими словами, характерними значками, виділеними об'єктами, фотокадрами тощо).

Образний комп'ютер можна порівняти з інтелектуальним роботом, система управління якого може бути аналогом образної операційної системи. Подібно до того, як окремі складові робота мають свої локальні системи управління і інформація про стан підсистем надходить у центральну систему управління для прийняття скоординованих рішень в роботі в цілому та його частинах, так і в образному комп'ютері його ООС “синхронізує” узгоджене сприйняття різних модальностей та каналів надходження інформації.

Тому можемо сказати, що, наприклад, в основі бімодальної системи сприйняття та синтезу мовленнєвого сигналу й міміки-емоцій обличчя та жестикуляції повинна бути механо-семантична модель мовотворення та сприйняття людиною мовленнєвої інформації.

ЛІТЕРАТУРА

- [1] Тарас Вінцюк, “Образний комп'ютер: концепції, методологія, підходи”, *Праці П'ятої Всеукраїнської міжнародної конференції «УкрОбраз'2000»*. Видання УАсОІРО, Київ, 2000, с. 9-16.
- [2] Тарас Вінцюк, “Генеративна модель образного комп'ютера”, *Праці Шостої Всеукраїнської міжнародної конференції «УкрОбраз'2002»*. Видання УАсОІРО, Київ, 2002, с. 7-14.
- [3] J.Kittler@eim.surrey.ac.uk