

АНАЛІЗ ВИСХІДНОЇ ІКН-МОДЕЛІ ЗОРОВОЇ УВАГИ

© Степанюк С., 2011

Розглядається моделювання зорової уваги на основі висхідних процесів, зокрема аналізується модель Ітті як одна із базових і широко використовуваних. Наведено переваги та обмеження цієї моделі, відкриті питання в цій області, також висунуто гіпотези щодо можливих шляхів вдосконалення моделі. Одними з таких є заміна нейронної мережі на ефективнішу, перехід від системи кольорів RGB до CMYK або HSV, врахування додаткових властивостей візуальної сцени.

Ключові слова: зорова увага, висхідна селективна увага, обчислювальна модель, карта властивостей, карта характерних особливостей, піраміда Гаусса, штучна WTA нейронна мережа, фокус уваги.

The consideration of modelling bottom-up visual attention is given in this paper. Particularly, saliency-based Itti's model as one of the basic and widely used is analysed here. Advantages and limitations, open questions, and some hypotheses of improvement of this model are shown. Especially, we propose to change WTA network of leakly integrate-and-fire neurons on most modern simple network. We assume that using CMYK or HSV color model will give possibility of allowance of additional visual features.

Keywords : visual attention, bottom-up process, computational model, feature map, saliency map, Gaussian pyramid, winner-take-all network, focus of attention.

Вступ

Доволі часто у системах комп'ютерного зору, зокрема в робототехніці, ставляться задачі як повної сегментації, так і виділення найінтенсивніших об'єктів (ділянок) візуальної сцени. Такі об'єкти, своєю чергою, характеризуються різноманітністю зовнішніх виглядів, контурів, накладанням один на одного, затіненням (їх форма може бути наперед і невідомою) та іншими властивостями, що ускладнюють сегментацію. А це є дуже складна задача комп'ютерного зору, для розв'язання якої в багатьох випадках традиційні алгоритми сегментації [1, 2], зокрема алгоритм нарощування регіонів [3] комбінуються із моделювання візуальної уваги людини.

Зорова, або візуальна увага (visual attention) – процес відбору людиною участків релевантної інформації із візуальної сцени для подальшої їх обробки [4]. З обчислювального погляду, візуальна увага – це процес селекції частин зображення для локалізації, ідентифікації та інтерпретації об'єктів довкілля. Відомі два основні підходи до моделювання уваги [4].

1. Підхід, що ґрунтується на висхідних процесах (bottom-up). Розподілення уваги визначається властивостями образу (наприклад, неочікуваний рух об'єкта, контраст кольору, напрямленість). При цьому рішення про зміщення уваги приймається без врахування свідомості людини. Зорова система функціонує за принципом висхідного процесу, коли створення загального образу є результатом об'єднання базових елементів візуальної сцени.

2. За другого підходу, оснований на нисхідних процесах (top-down), в основу покладено знання, попередньо отримані спостерігачем, його досвід, інтерпретація та очікування.

Механізм зорової уваги може складатися як з висхідних, так і з нисхідних процесів. У літературі із фізіології та психології уваги наголошується, що ключову роль відіграють саме

висхідні процеси [5]. Однією із найефективніших демонстрацій того, що людина керується значною мірою висхідною увагою, є експерименти з візуального пошуку, які проводили Трейсман і Геладе [6]. Існують задачі комп'ютерного зору, які можна розв'язувати за допомогою як bottom-up, так top-down моделей. Але є і значна частина таких, де наперед нічого не відомо про характер об'єктів для пошуку, що унеможливує застосування top-down моделей.

Методи і моделі візуальної уваги активно розвиваються, оскільки знаходять різноманітне застосування під час автоматизації задач комп'ютерного зору. Одними із найсучасніших застосувань є детекція об'єктів [7], сегментація зображень [8], компресія зображень [9], розпізнавання об'єктів [10], оцінювання якості зображень [9] та ін.

Сучасні методи роз'язання задач зорової уваги хоча й характеризуються значною ефективністю, але все одно мають ряд обмежень: вони не є універсальними відносно типу зображень чи об'єктів виділення, часто неробастні до наявності шумів, можуть бути складні в реалізації або з великою обчислювальною складністю, з недостатньою точністю чи швидкістю. Часові обмеження унеможливають ефективне використання моделей візуальної уваги в системах реального часу.

Широке застосування і ряд відкритих питань породжують актуальність дослідження та вдосконалення моделей зорової уваги. У зв'язку із універсальністю і широким застосуванням саме комп'ютерних моделей уваги, основаних на висхідних процесах, у цій статті ми проаналізуємо та дослідимо ці моделі.

Постановка проблеми

Через велику складність та практично 100-відсоткову нерозв'язність задачі моделювання зорової уваги кожна з моделей має ті чи інші обмеження, зокрема і одна з базових – модель Ітті [11]. У зв'язку з тим, що ця модель є однією із ключових висхідних моделей зорової уваги, яка має ряд модифікацій, інтеграцій з іншими моделями, прикладних застосувань, буде доцільним її подальше дослідження. Отже, завданням цієї роботи є детальний розгляд та аналіз моделі Ітті, якісні порівняння, наведення переваг та обмежень та формулювання відкритих питань. Також планується відшукання деяких можливих шляхів вдосконалення моделі.

Основні поняття та короткий огляд моделей висхідної уваги

Завданням кожної моделі зорової уваги є знаходження областей зображення, які в першу чергу притягували б увагу людини. В літературі відсутній однозначний термін для визначення таких областей (Salient Region). У роботах [5, 11] використовується поняття фокуса уваги (Focus of Attention, FOA), хоча чітке формулювання його відсутнє. Надалі розумітимемо під FOA область зображення (місце скупчення уваги), що містить його елементи, на яких концентрується увага людини в певний момент часу. З часом увага може зміщуватися з однієї частини зображення на іншу, тобто змінюються і фокуси. Зазвичай фокус на зображенні виділяється кругом заданого радіуса, як на рис. 1, взятий із [11].



Рис. 1. Ілюстрація фокусів уваги та послідовності їх знаходження, відповідно: А – перший Б – другий; В – третій фокус

Поряд із фокусом уваги в науковій літературі також зустрічається термін “протооб’єкт” (proto-object) [12]. Але на відміну від FOA, протооб’єкт визначається довільною формою, що є грубим наближенням деякого спостережуваного об’єкта.

Для того, щоб порівняти моделі, наведемо їх основні характеристики:

- 1) ступінь відповідності роботі зорової уваги людини;
- 2) характер задач (динамічні або статичні візуальні сцени, наявність шумів);
- 3) здатність до пошуку протооб’єкта;
- 4) швидкість пошуку фокусів уваги;
- 5) ресурсні (апаратні) затрати, необхідні для реалізації моделі;
- 6) можливість схемотехнічної реалізації.

Щодо відповідності людській зоровій увазі, то, з одного боку, можна говорити, що модель може реалізовувати реальні фізіологічні та психологічні процеси при моделюванні уваги, з іншого – лише результат, тобто виділення фокусів відбувається подібно як у людини, при цьому використовуючи довільний математичний апарат. Постає питання: як перевірити правдоподібність тієї чи іншої моделі уваги. Одним з методів є перевірка моделі на колекції картинок, де експерти позначили участки (фокуси) уваги. Іншим підходом є зіставлення траєкторії зміщення уваги із рухом зіниць людини під час огляду цього зображення [13].

Отже, з вищесказаного можна зробити висновок, що практично не може бути найкращої моделі в універсальному сенсі. Може бути та чи інша модель, краща за інші за одним чи кількома параметрами (в одному чи кількох розрізах). Тому на вибір моделі в реальних застосуваннях впливає ще й сам характер поставленої задачі та вимоги до результату.

Обчислювальні моделі зорової уваги переважно ґрунтуються на психологічних теоріях уваги. Однією з перших була висхідна модель, основана на так званих картах характерних особливостей або картах випуклості (saliency map, SM), які запропонували Коч і Ульман [14] і пізніше реалізована Ітті – ІКН-модель [11]. У цій моделі SM-карта виконує функцію комбінування випуклостей локальної інформації від кожного з нижчих чи попереднього рівня карт властивостей (feature maps). За допомогою штучної нейронної мережі знаходження максимальних сигналів [14] відбувається пошук найбільш інтенсивних (випуклих) участків результуючої карти. Загалом такого типу моделі складаються з трьох основних блоків: кілька паралельно розділених карт властивостей (feature maps) представляють візуальну сцену в таких вимірах, як колір, орієнтація, інтенсивність та ін.; одна чи більше карт характерних особливостей (SM) комбінують канали вхідних даних; штучна нейронна мережа знаходження максимальних сигналів (Winner-Take-All network, WTA) вибирає найбільш опуклі участки SM-карти. Загальну структуру висхідних моделей показано на рис. 2.



Рис. 2. Загальна схема трикомпонентної висхідної моделі зорової уваги.

На темному фоні зображено кожну з компонент, відповідно,

перша – видобування властивостей,

друга – безпосередньо модель уваги,

третья – механізм вибору фокусів

Ще однією спробою моделювати зорову увагу є НКР-модель [15], в якій використовується підхід, близький до ІКН-моделі. Відмінність полягає в тому, що в НКР-моделі для формування карт властивостей і створення карти характерних особливостей використовуються додаткові методи обробки зображення, що ґрунтуються на теорії графів. Оскільки НКР-модель – це по суті вдосконалення ІКН-моделі, то очевидно, що в якихось розрізах вона має бути більш результативною, що і показують дослідження в [15].

КВ-модель [16], запропонована *Kadir* і *Brady*, характерна тим, що в кожному пікселі зображення обчислюється ентропія наперед вибраної ознаки (наприклад, колір чи інтенсивність) в межах серії фреймів. Задається серія вікон різного масштабу S ($S_{\min} \leq S \leq S_{\max}$). Шляхом обчислення ентропії для кожного пікселя знаходиться масштаб, якому відповідає максимальна ентропія із заданим ваговим коефіцієнтом.

Варто згадати WK-модель (*Walther* і *Koch*) висхідної зорової уваги, запропоновану в [12]. Модель може працювати із сценами, на яких знаходяться кілька об'єктів і як знаходити фокус уваги, що збігається із найпомітнішим об'єктом, так і здійснювати їх перебір. У цій моделі із кольорового зображення при різних масштабах видобуваються карти властивостей, тобто цей етап відбувається аналогічно, як у моделі Ітті. Після того, як за допомогою WTA-нейронної мережі обрано координати елемента карти характерних особливостей з найбільшою випуклістю, відбувається процес обробки у зворотному порядку від SM-карти до карти властивостей. Знаходиться та матриця властивостей, яка зробила найбільший внесок у формування карти характерних особливостей і надалі використовується як маска для виділення із сцени протооб'єкта. Головна відмінність WK-моделі полягає в тому, що в ній визначається не лише фокус уваги, а й робиться спроба знайти протооб'єкт або хоча б його частину, на якій концентрується увага людини.

Як бачимо, ІКН-модель, крім того, що є однією із сучасних поширених висхідних моделей зорової уваги, ще є і базовою для багатьох інших моделей, тому її вдосконалення може поліпшити роботу багатьох існуючих систем комп'ютерного зору. У зв'язку із цим надалі будемо досліджувати саме модель Ітті.

Опис та аналіз ІКН-моделі

Як говорилося раніше, ІКН-модель ґрунтується на психологічних і фізіологічних принципах зорової уваги людини, тобто і механізми реалізації моделі, і знаходження фокусів тією чи іншою мірою наближають будову та поведінку зорової системи. Загальна архітектура моделі, представлена на рис. 3, складається із кількох взаємоузгоджених блоків компонент. Розглянемо детально кожен із них.

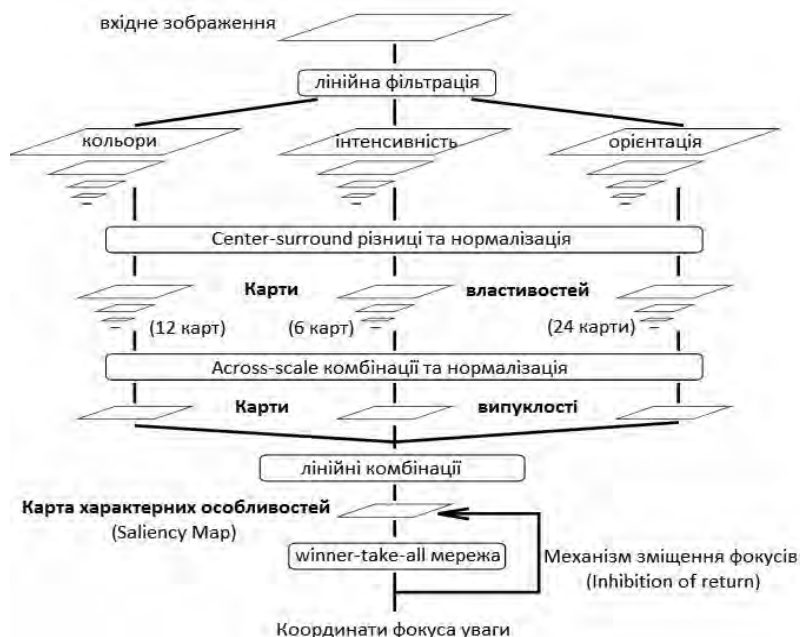


Рис. 3. Загальна архітектура ІКН-моделі

На вхід моделі подається двовимірне статичне кольорове зображення із розширенням 640×480 , яке представляється RGB системою кольорів. Початкове зображення кодується трьома каналами r, g, b для червоного, зеленого і голубого кольорів відповідно. Потім внаслідок лінійних перетворень формуються дев'ять каналів даних (матриць), що характеризують такі властивості: червоний, зелений, голубий і жовтий кольори, інтенсивність та орієнтацію в чотирьох напрямках. Інтенсивність I обчислюється за формулою $I = (r + g + b)/3$. Чотири матриці кольорів створюються з каналів r, g, b так: $R = r - (g + b)/2$, $G = g - (r + b)/2$, $B = b - (r + g)/2$, $Y = (r + g)/2 - |r - g|/2 - b$ для червоного, зеленого, голубого і жовтого кольорів відповідно.

Наступним кроком є створення пірамід властивостей, тобто на основі кожної матриці будують ще по вісім матриць в різних масштабах $\sigma \in [0..8]$. Для кольорів та інтенсивності формуються піраміди Гаусса $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, $Y(\sigma)$, $I(\sigma)$, для орієнтації – чотири піраміди Габора $O(\sigma, \theta)$, де $\sigma \in [0..8]$, а θ – кут нахилу, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ [17].

Потім формуються карти властивостей з використанням так званої „center-surround” різниці, які реалізуються шляхом віднімання від пікселів матриць із масштабами $c \in \{2, 3, 4\}$ відповідних пікселів матриць із масштабами $s = c + \delta, \delta \in \{3, 4\}$. Так утворюють шість карт властивостей для інтенсивності за співвідношенням:

$$I(c, s) = |I(c) \ominus I(s)| \quad (1)$$

Для каналів кольору будуються спарені карти властивостей (double-opponent color) $RG(c, s)$ та $BY(c, s)$:

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (2)$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (3)$$

Кarti властивостей для орієнтації обчислюються за виразом:

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (4)$$

де θ – кут нахилу і $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Загалом формується 42 карти властивостей: шість для представлення інтенсивності, дванадцять – кольору, двадцять чотири – орієнтації. Також на цьому етапі роботи моделі відбувається нормалізація утворених карт з використанням оператор нормалізації $N(\cdot)$, що функціонує за таким алгоритмом:

- 1) знаходять глобальне максимальне значення M серед усіх карт властивостей;
- 2) обчислюють середнє значення \bar{m} для кожної матриці властивостей;
- 3) максимальне в нормалізованих картах формують як $(M - \bar{m})^2$.

Приклад нормалізації карт інтенсивності та орієнтації показано на рис. 4.

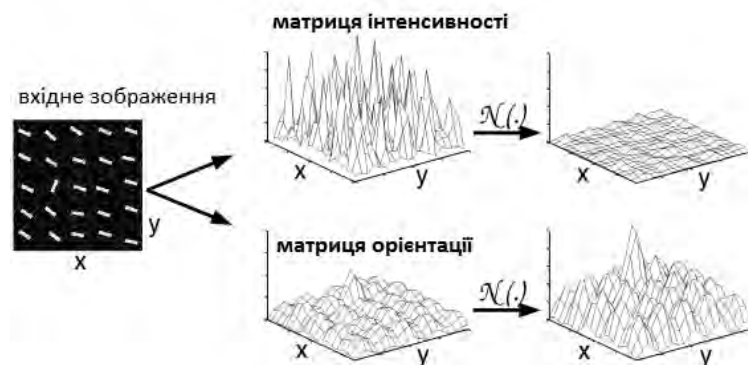


Рис. 4. Графічне представлення результатів нормалізації матриць інтенсивності та орієнтації

Як видно із рис. 2, наступним кроком функціонування моделі є так звані across-scale комбінації, позначені як “ \oplus ”, та повторна нормалізація, тобто тут створюються карти випуклості (conspicuity maps, CM) із масштабом $\sigma = 4$. Карти властивостей певним чином комбінуються і формують лише три CM-карти: \bar{I} для інтенсивності, \bar{C} для врахування кольору, \bar{O} для орієнтації. Карти випуклості отримуються через across-scale додавання шляхом зменшення кожної карти до масштабу $\sigma = 4$ і поелементним додаванням:

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c-4} N(I(c,s)) \quad (5)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c-4} [N(RG(c,s)) + N(BY(c,s))] \quad (6)$$

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N \left(\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c-4} N(O(c,s,\theta)) \right) \quad (7)$$

На етапі лінійних комбінацій створюється ключова карта ІКН-моделі – карта характерних особливостей (Saliency map). Це є двовимірний масив із дійсних елементів із масштабом $\sigma = 4$ відносно вхідного зображення (рис. 5). Вона формується на основі трьох карт випуклості за співвідношенням:

$$S = \frac{1}{3}(N(\bar{I}) + N(\bar{C}) + N(\bar{O})) \quad (8)$$

Інколи (8) обчислюють так:

$$S = w_I \cdot N(\bar{I}) + w_C \cdot N(\bar{C}) + w_O \cdot N(\bar{O}), \quad (9)$$

де коефіцієнт w – наперед задана вага для відповідної карти випуклості.

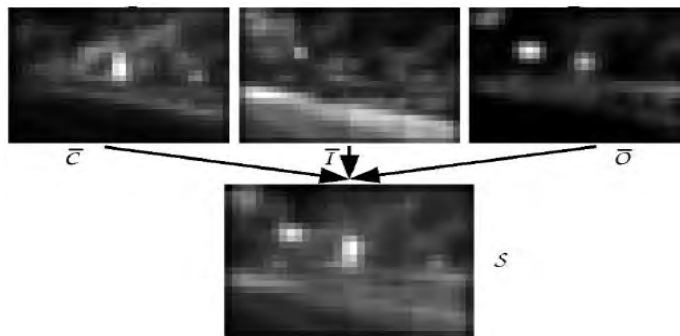


Рис. 5. S – карта характерних особливостей;
 $\bar{C}, \bar{I}, \bar{O}$ – карти випуклості кольору, інтенсивності та орієнтації відповідно

Завершальним елементом моделі є знаходження та зміщення фокусів уваги на SM-карті. На цьому кроці карта характерних особливостей зіставляється із двовимірним масивом штучних нейронів типу integrate-and-fire [14]. У довільний час максимальне значення цієї карти асоціюється із фокусом уваги. Схему механізму зміщення FOA показано на рис. 6.



Рис. 6. Графічне представлення результатів нормалізації матриць інтенсивності та орієнтації

На вхід нейронної WTA-мережі подаються значення із SM-карти. Мережа знаходить максимальний елемент на карті характерних особливостей, за яким можна визначити координати на вхідному зображенні, і цим встановлюють фокус уваги. Але для того, щоб вдруге не обрати попередньо знайденого FOA, використовують механізм зміщення уваги, який полягає у зменшенні („вибиванні”) значень SM-карти в деякому радіусі за співвідношенням:

$$IOR(t + \delta t, x, y) = IOR(t, x, y) - 5e^{-\frac{dx^2 + dy^2}{(rFOA)^2}} + e^{-\frac{dx^2 + dy^2}{rFOA}}, \quad (10)$$

де x та y – координати, dx і dy – відстань до центру FOA, $rFOA$ – радіус фокуса.

Динаміка нейронів мережі задається:

$$V(t + \delta t) = \left(1 - \frac{\delta t}{CR}\right)V(t) + \frac{\delta t}{C}I(t), \quad (11)$$

де C – ємність, R – опір, t – час, $V(t)$ і $V(t + \delta t)$ – попереднє та поточне значення на виході. Параметри налаштування мережі можна знайти в [11]. На рис. 7 показано приклад функціонування моделі Ітті. Параметри обрані так, що знаходження першого фокуса відбувається через 30–70 мс і обнулення області відповідної FOA дійсне на 500–900 мс, тобто так, як показують психологічні спостереження.

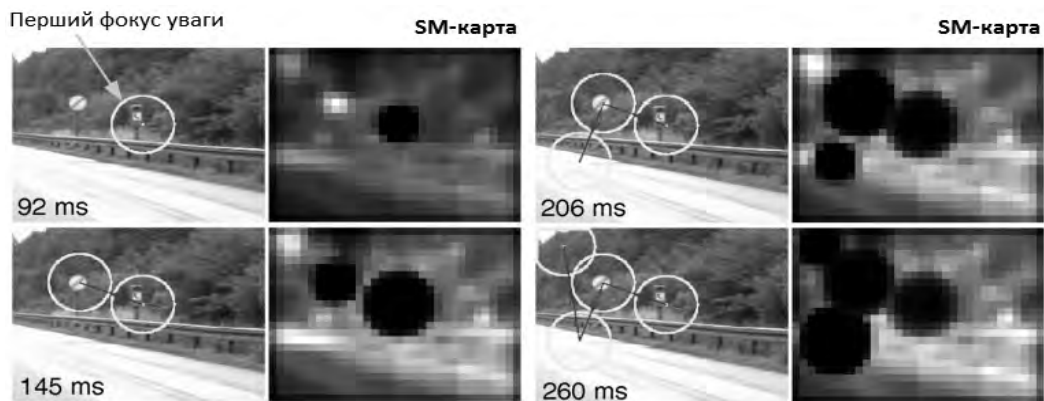


Рис. 7. Знаходження фокусів уваги на зображенні і відповідні їм частки на SM-картах

Детально розглянувши модель, можна навести основні переваги та обмеження, а також відкриті питання моделювання висхідної зорової уваги. Однією із переваг ІКН-моделі є можливість запаралелювання обчислень, що зменшує часові затрати, які особливо помітні у разі схематехнічної реалізації.

Оригінальна модель Ітті не передбачає можливості моделювання зорової уваги для відеопотоку, тобто не враховує руху об'єктів, вона розрахована тільки для статичних зображень.

Сучасні психологічні та фізіологічні дослідження зору людини свідчать про те, що модель представлення кольорів СМҮК більш відповідає фізіології ока, ніж RGB. У зв'язку з цим ми припускаємо, що перехід до системи кольорів СМҮК дасть змогу покращити якість результатів комп'ютерного моделювання зорової уваги людини. З іншого боку, в роботі [18] обґрунтовується перехід до моделей кольорів HSV або LAB у зв'язку із тим, що система RGB не розділяє інформації хроматичної від інтенсивності. Залишається відкритим питання вибору оптимальної моделі кольорів для представлення вхідного зображення.

Також залишається відкритим питанням щодо конкретно проаналізованої моделі: чому карти випуклості і карта характерних особливостей будуються при $\sigma = 4$, тобто за масштабом у чотири рази менші відносно початкового зображення. Немає досліджень функціонування моделі Ітті, використовуючи SM-карти інших розмірів.

Одним із ключових елементів моделі є WTA-мережа. Сьогодні існує велика кількість мереж для знаходження максимальних сигналів, кожна з яких має свої переваги та обмеження. У даному випадку використовується мережа з integrate-and-fire нейронів потребує режиму перезавантаження,

для неї немає доведення єдиності розв'язку. Режим повернення до початкового стану потребує додаткової схеми при апаратній реалізації і збільшення машинного часу, що зменшує швидкість роботи системи та ускладнює її реалізацію. Тому пропонується замінити цю мережу на сучаснішу звичайну WTA-мережу без integrate-and-fire нейронів.

Ще більш неоднозначними для вирішення є питання, які стосуються будь-якої моделі уваги, зокрема і вищерозглянутої. Одними з них є: “Що потрібно розглядати як збудники чи ознаки уваги?”, “Які властивості сцени оптимальні?”, “Як комбінувати ці властивості?” та ін. Незважаючи на інтенсивні дослідження міждисциплінарних наук, пов'язаних із зоровою увагою, такого роду питання ще не повністю вирішені.

Висновки

Як видно із вищерозглянутого матеріалу, моделювання зорової уваги людини відіграє важливе значення під час розроблення багатьох систем комп'ютерного зору. Зокрема модель Itti як одна із базових лідирує серед висхідних моделей, на основі якої розроблено ряд модифікацій та практичних застосувань. Тому вдосконалення цієї моделі приведе до можливості покращання роботи багатьох систем технічного зору. Детально проаналізувавши IKN-модель, пропонується замінити використовувану WTA-мережу на сучаснішу, із меншими затратами часу, яка гарантує єдиність розв'язку. Також варто дослідити та порівняти функціонування моделі, формуючи карту характерних особливостей при різних масштабах. Враховуючи той факт, що модель СМҮК представлення кольору більшою мірою відповідає фізіологічним принципам зорового сприйняття, ми припускаємо доцільність переходу від RGB до СМҮК.

1. Pedro F. Efficient graph-based image segmentation / F. Pedro, Felzenszwalb and Daniel P. Huttenlocher // *International Journal of Computer Vision*. – 2004. – V. 59(2)(5). – P. 167–181.
2. Malik J. Contour and texture analysis for image segmentation / J. Malik, S. Belongie, T. Leung, and J. Shi // *International Journal of Computer Vision*. – 2001. – V. 43(1). – P. 7–27.
3. Ouerhani N. Visual attention guided seed selection for color image segmentation / N. Ouerhani, N. Archip, H. Hgli, and Pierre-Jean Erard // *Computer Analysis of Images and Patterns*. – 2001. – V. 2124. – P. 630–637.
4. Chun M. Visual Attention / Marvin M. Chun, Jeremy M. Wolfe, Blackwell. – *Handbook of Perception*, 2000.
5. Шиффман Х. Ощущение и восприятие. 5-е изд. – СПб, 2003. – 928 с.
6. Treisman A. A Feature-Integration Theory of Attention / A. Treisman, G. Gelade // *Cognitive Psychology*. – 1980. – V. 12. – No. 1. – P. 97–136.
7. Moosmann F, Learning Saliency Maps for Object Categorization / F. Moosmann, D. Larlus and F. Jurie // *International Workshop on The Representation and Use of Prior Knowledge in Vision*. – 2006.
8. Achanta R. Salient Region Detection and Segmentation / R. Achanta, F. Estrada, P. Wils and S. Ssstrunk // in *Proceedings of the International Conference on Computer Vision Systems*. – 2008. – P. 66–75.
9. Guo C. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression / C. Guo and L. Zhang // *IEEE Trans. Image Processing*. – 2010. – V. 19(1). – P. 185–198.
10. Kanan C. Robust Classification of Objects, Faces, and Flowers Using Natural Image Statistics / C. Kanan and G.W. Cottrell // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2010.
11. Itti L. A model of saliency-based visual attention for rapid scene analysis / L. Itti, C. Koch, E. Niebur // *IEEE Trans. Pattern Anal. Machine Intel.* – 1998. – V. 20. – No. 11. – P. 1254–1259.
12. Walther D. Modeling attention to salient proto-objects / D. Walther, Ch. Koch // *Neural Networks*. – 2006. – V. 19. – P. 1395–1407.
13. Shic F. A Behavioral Analysis of Computational Models of Visual Attention / F. Shic, B. Scassellati // *International Journal of Computer Vision*. – 2007. – V. 73(2). – P. 159–177.
14. Koch C. Shifts in selective visual attention: towards the underlying neural circuitry / C. Koch, S. Ullman // *Human Neurobiology*. – 1985. – V. 4. – P. 219–227.
15. Harel J. Graph-Based Visual Saliency / J. Harel, C. Koch, P. Perona // *Proceedings of Neural Information Processing Systems (NIPS)*. – 2006.
16. Kadir T. An affine invariant salient region detector / T. Kadir, A. Zisserman, M. Brady // *Proceedings of the 8th European Conference on Computer Vision*. – 2004.
17. Greenspan H. Over complete Steerable Pyramid Filters and Rotation Invariance / H. Greenspan and other // *Proc. IEEE Computer Vision and Pattern Recognition*. – 1994. – P. 222–228.
18. Frintrop S., VOCUS: A visual attention system for object detection and goal-directed search / Simone Frintrop // *PhD thesis*. – 2006.