УДК 681.142.2; 622.02.658.284; 621.325

Д.Д. Пелешко, Н.Лотошинська, Н.Кустра
Національний університет "Львівська політехніка"

ФІЛЬТРАЦІЯ НАБОРІВ ОДНОТИПНИХ ЗОБРАЖЕНЬ НА ОСНОВІ КВАНТУВАННЯ МАТЕМАТИЧНИХ СТАТИСТИК

На основі квантування математичних статистик запропоновано інтервальні оцінки для цих параметрів з метою організації фільтрацію наборів однотипних зображень. На основі практичних експериментів визначено особливості використання кожної математичної статистики.

There is proposed the filtration of the sets of similar images. The filtration is based on the build of interval estimate and used quantification of parameters of model presentation images. Comparative analysis of the results of quantification filtration and parameter frequency (for the all statistical parameters).

Вступ

Розробка методів видобутку інформації із наборів цифрових зображень висуває певні вимоги до формування і попередньої обробки не лише до окремих зображень, а й до параметрів та характеристик наборів в цілому. До останніх, зокрема відносять однотипність й однакова розрядність зображень, суміщення в межах піксела тощо. Це спричинює потребу у розвитку і постійному вдосконаленню методів попередньої обробки наборів. Однією із найбільш актуальних задач в царині попередньої обробки є фільтрації з точки зору покращення характеристик множини параметрів обраної моделі представлення зображень набору [1].

Постановка задачі

Нехай існує набір **P** однотипних зображень. Шумом набору будемо називати стохастичні варіації параметрів вибраної моделі представлення зображень набору [1]. При цьому параметрами виберемо математичні статистики набору однотипних зображень, а саме: математичне сподівання, дисперсію та девіацію [3, 4].

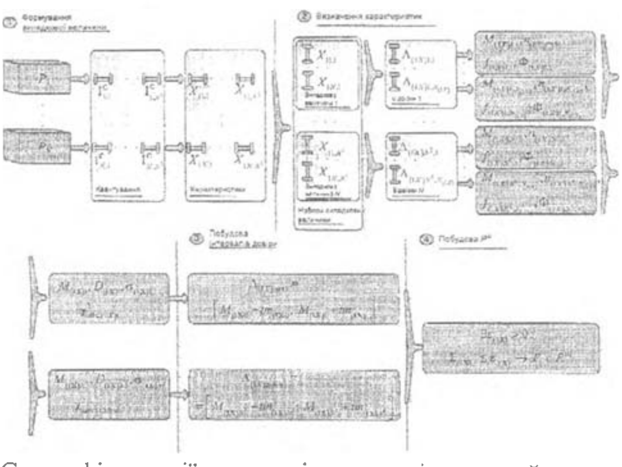

Рис.1. Схема фільтрації на основі параметрів моделей представлення

Якщо через $\mathbf{P}^{\Phi}$ позначити відфільтровану підмножину зображень набору **Р**, а через $\mathbf{P}^{\text{Ш}}$ – множину зображень з **Р**, які складають "шум", то очевидно повинна задовольнятись операція об'єднання $\mathbf{P}^{\text{Ш}}$ і $\mathbf{P}^{\Phi}$. Тоді задача фільтрації полягає в побудові множини $\mathbf{P}^{\text{Ш}}$ такої, щоб мала місце операція

$$P = P^{\Phi} \cup P^{\text{Ш}}; \qquad P^{\Phi} \cap P^{\text{Ш}} = \varnothing . \tag{1}$$

Фільтрація на основі квантування параметрів моделей представлення зображень застосовується у випадку малих розмірностей наборів. Тобто у тих випадках, коли розмірність вибірки із генеральної сукупності є недостатньою для статистичної перевірки гіпотез.

Схема фільтрації наведена на рис.1, на якому параметр X, треба замінити на будь-який статистичний параметр моделі представлення.

Основною ідеєю фільтрації на основі квантування є розбивання кожного зображення на проміжки (процес квантування) і побудова множини випадкових величин як середніх значень параметрів зображення на кожному проміжку квантування. Тоді при великих кількостях рівнів квантування N^{I} можна отримати більшу розмірність вибірки в порівняння із заданою розмірністю набору. Більше того, в результаті квантування інтервали довіри будуються для усіх N^{I} наборів випадкових величин. А тому, на відміну від фільтрації за параметром, рішення про шум приймається на основі аналізу потраплянь у декілька інтервалів довіри, а не в один.

Обов'язковість розгляду N^{I} наборів випадкових величин є деяким обмежувальним фактором до необмеженого зростання значення N^{I}, що призводить до пропорційного зростання обчислювальних затрат .

Фільтрація набору зображень на основі квантування математичних статистик

Формування наборів випадкових величин. Квантуємо кожне P_z набору Р, в результаті чого отримуємо

$$P_z \to \left\{ \mathrm{I}^{\mathrm{X}}_{z,m} \to \mathrm{I}^{\mathrm{C}}_{z,m} \right\}, \qquad m = \overline{1, N^{\mathrm{I}}}, \tag{2}$$

де $\mathrm{I}^{\mathrm{X}}_{z,m}$, $\mathrm{I}^{\mathrm{C}}_{z,m}$ - відрізки квантування відповідно в координатній та колірній областях; N^{I} - розмірність квантування. Приймаємо, що для усіх z вона є однаковою.

На кожному $\mathrm{I}^{\mathrm{X}}_{z,m}$ обчислюємо значення параметрів моделі представлення. У нашому випадку це математичне сподівання кольору (інтенсивності) $M_{1z,m}$ усіх пікселів, які потрапили до $\mathrm{I}^{\mathrm{C}}_{z,m}$. В результаті для кожного z отримаємо однакові за розмірністю набори характеристик

$$\forall m \in \left[1; N^{\mathrm{I}}\right] \; \exists \left\{ M_{1z,m} \right\}_{z=1..N} : M_{1z,m} = \frac{1}{N_{z,m}} \sum_{c_{z,m} \in \mathrm{I}^{\mathrm{C}}_{z,m}} c_{z,m} \, ; \tag{3}$$

де $c_{z,m}$ - значення кольору (інтенсивності), яке потрапило в інтервал $\mathrm{I}^{\mathrm{C}}_{z,m}$; $N_{z,m}$ - розмірність $\mathrm{I}^{\mathrm{C}}_{z,m}$. І при будь-якому z для кожного P_z отримуємо співвідношення з набором розмірністю N^{I}

$$\forall z \in [1;N]: \quad P_z \to \begin{pmatrix} \mathrm{I}^{\mathrm{X}}_{z,m} \\ \mathrm{I}^{\mathrm{C}}_{z,m} \end{pmatrix} \to \{M_{1z,m}\}, \qquad m=\overline{1,N^1}, \tag{4}$$

За (3) і (4) по кожному m набору **Р** можна сформувати N^1 наборів випадкових величин розмірністю N.

$$\mathbf{P} \to \left\{ \{M_{1z,m}\}_{z=1..N} \right\}_{m=1..N^1} \tag{5}$$

Квантування наборів випадкових величин. Для кожного набору $\left\{\{M_{1z,m}\}_z\right\}_m$ після сортування характеристик сформуємо набір відрізків. В результаті цього для кожного m отримаємо

$$\begin{gathered} \forall m \in [1..N^1]: \\ \forall z \in [1..N],\ \exists k \in \left[1..N_{(1M)}\right]: \quad M_{1z,m} = M_{1z,m,k} \in \Lambda_{(1M)m,k} = \left[\mathrm{I}_{(1M)m,k-1}, \mathrm{I}_{(1M)m,k}\right[, \end{gathered} \tag{6}$$

де $\Lambda_{(1M)m,k}$ - k-й інтервал m-ї випадкової величини $\left\{\{M_{1z,m}\}_z\right\}_m$; $N_{(1M)}$ - кількість інтервалів m-ї випадкової величини. В загальному випадку для різних наборів може бути різна кількість інтервалів. Для спрощення вважаємо, що для усіх наборів $\left\{\{M_{1z,m}\}_z\right\}_m$ значення $N_{(1M)}$ є однаковим (таке спрощення не є принциповим). Значення $\mathrm{I}_{(1M)m,0}$ та $\mathrm{I}_{(1M)m,N_{(1M)}}$ для усіх m визначається так

$$\forall m \in [1..N^1]: \quad \mathrm{I}_{(1M)m,0} = \min_{z\in[1..N]}\left(M_{1z,m}\right); \qquad \mathrm{I}_{(1M)m,N_{(1M)}} = \max_{z\in[1..N]}\left(M_{1z,m}\right). \tag{7}$$

В результаті (6) отримаємо N^1 наборів відрізків $\left\{\Lambda_{(1M)m,k}\right\}_k$. Розмірність кожного набору рівна $N_{(1M)}$

$$\Lambda_{(1M)m} = \left\{\Lambda_{(1M)m,k}\right\}_{k=1..N_{(1M)}}, \quad m=1..N^1; \tag{8}$$

$$\Lambda_{(1M)} = \left\{\Lambda_{(1M)m}\right\}_{m=1..N^1}; \tag{9}$$

У кожному наборі (9) для будь-якого k визначимо середні значення випадкової величини $\left\{\{M_{1z,m}\}_z\right\}_m$

$$\begin{gathered} M_{(1M)m,k} = \mathrm{M}\, M_{1z,m,k} = \frac{1}{n_{(1M)m,k}} \sum_{M_{1z,m,k}\in\Lambda_{(1M)m,k}} M_{1z,m,k}; \\ \Lambda_{(1M)m,k} \to M_{(1M)m,k}, \quad \mathbf{M}_{(1M)} = \left\{M_{(1M)m,k}\right\}, \quad k=\overline{1,N_{(1M)}}, \end{gathered} \tag{10}$$

де $n_{(1M)m,k}$ - кількість m-ї випадкової величини на $\Lambda_{(1M)m,k}$. Експериментальна частота величини $M_{1z,m,k}$ буде визначатись так

$$f_{(1M)m,k} = \frac{n_{(1M)m,k}}{N}. \qquad (11)$$

Зауважимо, що кількість та експериментальна частота є залежними від m та k. Це означає, що при однакових m та k значення $f_{(1M)m,k}$ є різними. При цьому повинно виконуватись співвідношення

$$N = \sum_{k=1}^{N_{(1M)}} n_{(1M)1,k} = \ldots = \sum_{k=1}^{N_{(1M)}} n_{(1M)N^1,k}. \qquad (12)$$

Перевірка гіпотез про нормальність розподілу наборів випадкових величин. Для перевірки гіпотези для кожного набору за визначимо статистики [3, 4]

$$M_{(1M)m} = \mathrm{M}\, M_{1z,m} = \frac{1}{N}\sum_{z=1}^{N} M_{1z,m}; \qquad (13)$$

$$D_{(1M)m} = \mathrm{D}\, M_{1z,m} = \frac{1}{N-1}\sum_{z=1}^{N}\left(M_{1z,m} - M_{(1M)m}\right)^2; \qquad (14)$$

$$\sigma_{(1M)m} = \sqrt{D_{(1M)m}}. \qquad (15)$$

Теоретична частота величини $\left\{\left\{M_{1z,m}\right\}_z\right\}_m$, яка відповідає k-у інтервалу m-ї величини обчислюється за формулою Муавра-Лапласа [3]

$$\Phi_{(1M)m,k} = \Phi\left(M_{(1M)m,k}\right) = \frac{\mathrm{I}_{(1M)m,k} - \mathrm{I}_{(1M)m,k-1}}{K}\frac{1}{\sigma_{(1M)m}\sqrt{2\pi}}e^{-\left(\frac{M_{(1M)m,k} - M_{(1M)m}}{\sqrt{2}\sigma_{(1M)m}}\right)^2}. \qquad (16)$$

Тоді з врахуванням (11) обчислюване значення критерію Пірсона визначається за формулою

$$\chi^2_{\text{обч}\,(1M)m} = N\sum_{k=1}^{N_{(1M)}}\frac{\left(f_{(1M)m,k} - \Phi_{(1M)m,k}\right)^2}{\Phi_{(1M)m,k}}. \qquad (17)$$

Порівнюючи ці значення із табличними приймаються рішення про правильність гіпотези для кожного m. Очевидно, що не для усіх наборів випадкових величин $\left\{\left\{M_{1z,m}\right\}_z\right\}_m$ гіпотеза про нормальний розподіл справдиться. Більше того, для різних m вона може справдитись із різними рівнями значимості. Тоді такі набори випадкових величин видаляються із (5). В результаті цього отримаємо набір (5) розмірністю $N^{1\prime}$.

Побудова інтервалів довіри. Для наборів випадкових величин $\left\{\left\{M_{1z,m}\right\}_z\right\}_m$, серед яких приймається гіпотеза про нормальний розподіл, будуються інтервали довіри [3, 4] із довірчою імовірністю $p = 2\Phi(t_M)$

$$\Lambda_{(1M)\text{дов}m} = \left[M_{(1M)m} - tm_{(1M)m};\ M_{(1M)m} + tm_{(1M)m}\right];\quad m = 1..N^{1'}. \qquad (18)$$

де $m_{(1M)m}$ - середньоквадратичне відхилення середнього арифметичного m-ї випадкової величини

$$m_{(1M)m} = \frac{\sigma_{(1M)m}}{\sqrt{N}}. \tag{19}$$

Звернімо увагу на те, що у (18) вказано, що m набуває максимального значення $N^{I'}$, а не N^{I}. Це зумовлено тим, що не для усіх наборів $\left\{\left\{M_{1z\,m}\right\}_z\right\}_m$ може справдитись гіпотеза про нормальний розподіл. Якщо таких наборів буде небагато, то ними можна знехтувати при подальшому формуванні $P^{Ш}$. Це підтверджується результатами практичних експериментів. Видалення, до 5 випадковим чином вибраних наборів, не спотворювали суттєво (помилка в межах 1%) остаточні результати. Це означає, що не має потреби негайно змінювати функцію розподілу.

Формування Р^{Ш}. У результаті (18) для кожного P_z отримаємо вектор розмірністю $N^{I'}$, елементами якого будуть 0 та 1. При цьому значення 0, наприклад в позиції m, буде означати не потрапляння в m-й інтервал довіри математичного сподівання (як характеристики) отриманого на $I^{C}_{z\,m}$ даного зображення. Значення 1 буде свідчити протилежне – потрапляння m-й інтервал довіри.

Тоді кожному P_z набору **Р** поставимо у відповідність значення $\Sigma_{(1M)z}$, яке є сумою усіх значень координат описаного вектора.

$$P_z \to \Sigma_{(1M)z}. \tag{20}$$

Тоді формування $P^{Ш}$ буде здійснюватись за правилом

$$\exists \varepsilon_{(1M)} > 0: \quad \Sigma_{(1M)z} \le \varepsilon_{(1M)} \to P_z \in P^{Ш} \tag{21}$$

Висновки і результати практичних експериментів

Практична реалізація теоретичних викладок наведених вище здійснювалась для фільтрації набору однотипних зображень (рис.2), який має такі характеристики: розмірність набору – N = 27 зображень; зображення в градаціях сірого; розмірність кожного зображення – l = 34× h = 54 пікселів. При цьому базовою характеристикою вибрано ентропію H^{δ}_z. Параметр квантування N^{δ}_z при обчисленні H^{δ}_z рівний 15.

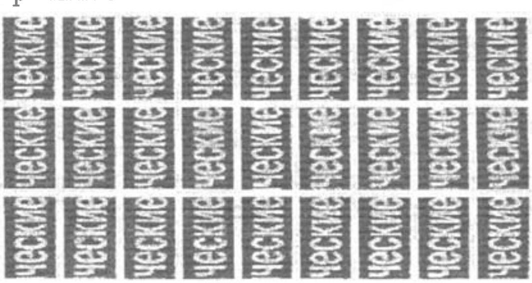

Рис.2. Набір однотипних зображень

На рис.3 наведено результати вирішення задачі (1) для НРОЗ при різних значеннях рівня значимості α. (рис.2), який має такі характеристики: розмірність набору – $N = 27$ зображень; зображення в градаціях сірого; розмірність кожного зображення – $l = 34\times h = 54$ пікселів. При цьому базовою характеристикою вибрано ентропію M_2.

Параметр квантування N^i рівний 37. Отже розмірність вибірки для побудови інтервальних оцінок вдалось збільшити в 1.3 рази.

На основі результатів практичних експериментів наведених на рис.3 можна стверджувати, що:

- інтервальні оцінки для кожного параметра є різними, що призводить до залежності побудови $\mathbf{P}^{Ш}$ від вибраного параметра. А це дає можливість розрізняти фільтрацію за параметром;
- при падінні рівня значимості інтервал довіри розширюється, що призводить до зменшення розмірності $\mathbf{P}^{Ш}$;
- характер залежності зміни $\mathbf{P}^{Ш}$ від рівня значимості не є однаковим для різних параметрів;

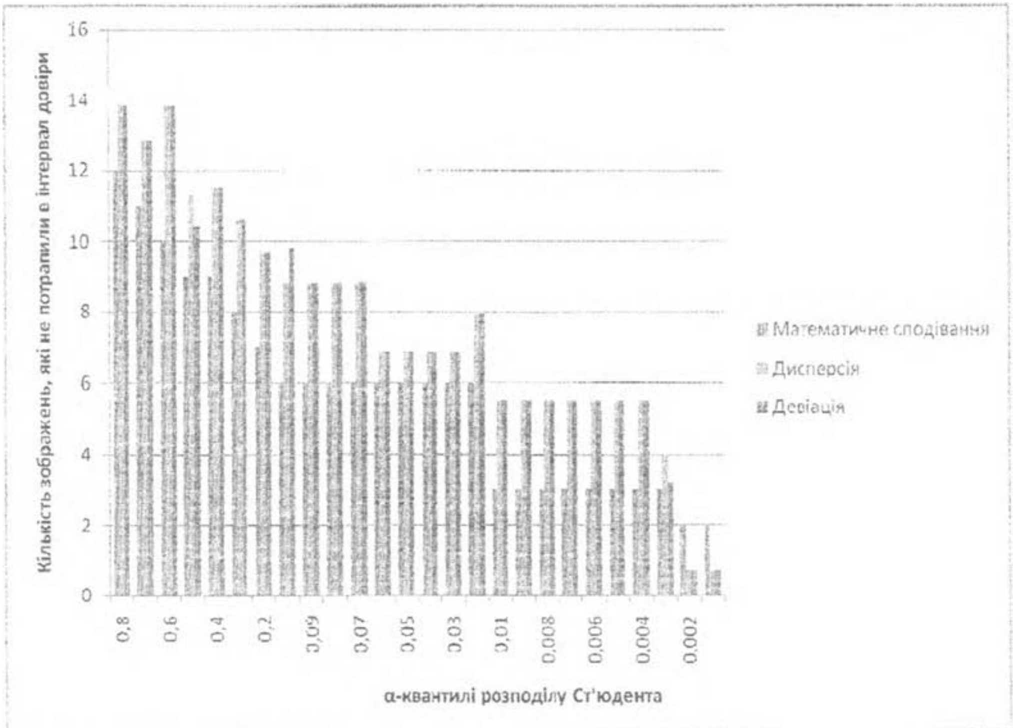

Рис.3. Результати фільтрації НРОЗ на основі квантування параметрів при різних значеннях α-квантилю розподілу Ст'юдента

На рис. 4 наведено результати кореляційних залежності результатів побудови $P^{Ш}$ за алгоритмами фільтрації на основі параметрів та на основі квантування відповідних параметрів. Отримані результати свідчать про дуже високий ступінь кореляції і близькість її значень - усі значення вклались в діапазон [0.95; 0.96].

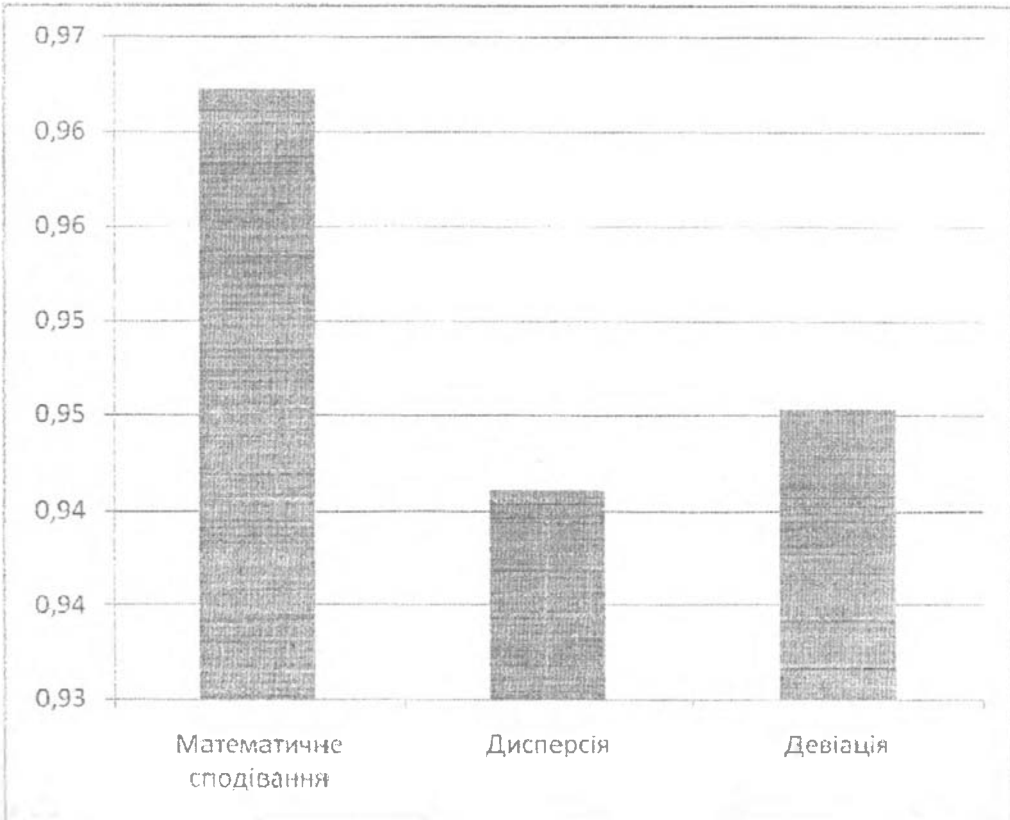

Рис.4. Кореляційні залежності результатів фільтрації за параметром і на основі квантування параметрів

Оскільки кореляція виражає взаємозв'язок між наборами даних, то вона не може повністю підтверджувати правильність отриманих результатів. А тому на рис.5 приведено відхилення отриманих результатів для різних параметрів при кожному значенні α–квантилю. З наведених результатів випливає, що:

- в усіх випадках похибка відхилення не перевищувала двох зображень, що становить 7%. Ц е означає, що задовільними є результати фільтрації на основі квантування для усіх параметрів.
- у випадках окремих параметрів завжди існують ділянки значень α–квантилю, при яких повністю збігаються результати фільтрації;

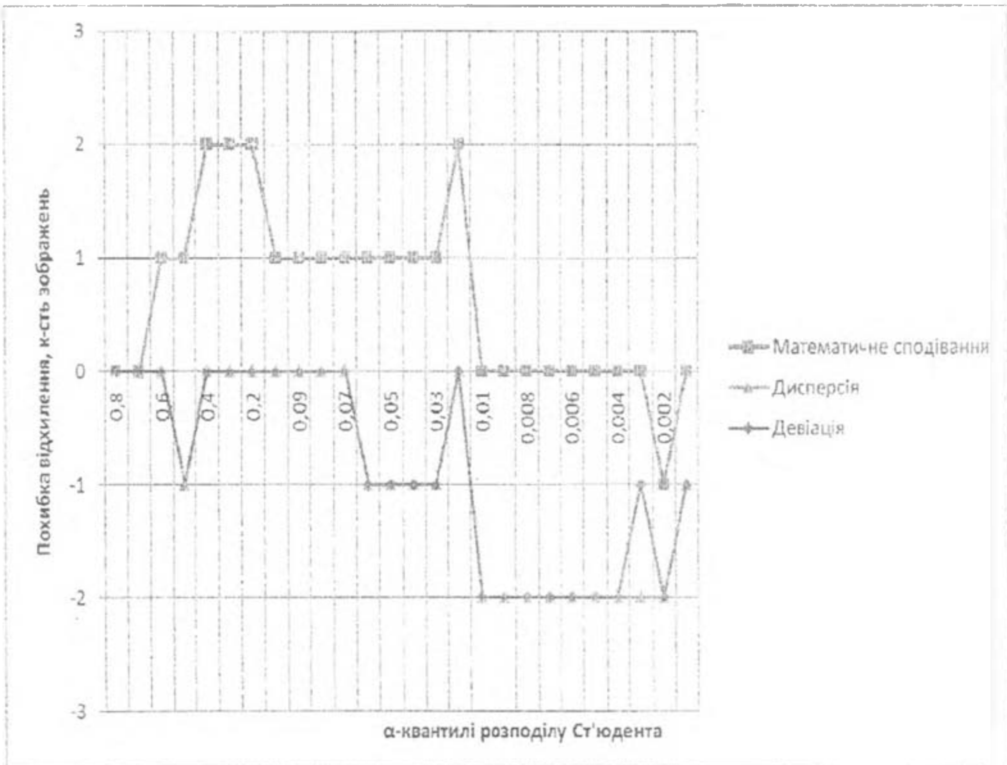

Рис.5. Відхилення результатів фільтрації НРОЗ за параметром та на основі квантування параметрів при різних значеннях α-квантилю розподілу Ст'юдента

1. Класифікація моделей представлення зображень та наборів зображень як стохастичних зображень та полів: Матеріали науково-практичної конференції ["Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту ISDMCI'2009"], (Євпаторія, 18-22 травня 2009) / Херсонський морський інститут.- Херсон: Видавництво Херсонського морського інституту, 2009.– т.2. – С. 401-405.

2. Пелешко Д.Д. Частотна фільтрація наборів зображень / Д. Пелешко // Науковий вісник НЛТУ України: Збірник науково-технічних праць. - 2008. - Вип. 18,6. – С. 291-303.

3. Колмогоров А.Н. Основные понятия теории вероятностей: учеб. пособие / А.Н. Колмогоров - 2-е изд. - М.: Наука, 1974. – 544с

4. Соловьев А.А. Лекции по теории вероятностей и математической статистике: учеб. пособие / А.А.Соловьев. – Челябинск: Челябинский гос. ун-т, 2003. – 118с.