

Д.К. Марчук, ст. викладач

Державний університет «Житомирська політехніка»

Використання афінних перетворень для корекції зображень з подальшим використанням у системах розпізнавання

(Представлено: Граф М.С., PhD)

Нині нейронні мережі набувають все більшої популярності і все частіше використовуються в різних галузях життя. Їх застосовують для аналізу даних, кластеризації, класифікації, виявлення об'єктів або схованих закономірностей. Все частіше зустрічаються варіанти застосування нейронних мереж у завданнях бізнес-прогнозування. Системи комп'ютерного зору не є винятком. Важливою складовою цих процесів є попередня обробка зображень. Відомо, що зображення, отримані з відеокамер, можуть виявляти перспективні спотворення, що походять від неідеальних кутів розташування камери. Зазвичай техніка афінного перетворення використовується для корекції геометричних спотворень або деформацій, таких як масштабування, переміщення, зсув, обертання. В роботі були проведені дослідження особливостей афінних перетворень і їх застосування з нейронними мережами. Розглянуто методи розпізнавання геометричних параметрів під час трансформації та переміщення зображень. Увагу приділено елементарним перетворенням, які враховують перенесення, масштабування, зсув, обертання. Розроблено математичну модель перетворення зображень для систем комп'ютерного зору. На базі запропонованого методу побудовано алгоритм перспективних перетворень зображень, отриманих з відеокамер, які знаходяться на автостоянках або парковках міста, що значно полегшує в подальшому виявлення, сегментацію і класифікацію об'єктів. Для покращення роботи класичної Mask R-CNN було проведено дослідження, в якому було додано блок з афінними перетвореннями до згорткової нейронної мережі. Афінні трансформації застосовано для корекції перспективного сходження ліній у кадрі, паралельних у реальності.

Ключові слова: відеозображення; геометричні параметри; комп'ютеризована система; афінні перетворення; нейронні мережі.

Актуальність теми. Завдяки стрімкому розвитку обчислювальної техніки виникають нові моделі та методи, які працюють з фото- і відеозображенням і дозволяють вирішувати складні завдання комп'ютерного зору. Для розпізнавання застосовують різноманітні методи, серед яких можуть бути кореляційні, синтаксичні або методи нормалізації. Нормалізація в цьому випадку передбачає автоматичний розрахунок деяких геометричних параметрів.

Зазвичай відеозображення, отримані з відеокамер, страждають від перспективних спотворень геометричних параметрів, що виникають через неідеальні кути камер. Техніка афінного перетворення зазвичай використовується для виправлення геометричних спотворень або деформацій, таких як масштабування, переміщення, зсув, обертання тощо.

Аналіз останніх досліджень та публікацій, на які спирається автор. В аналізі відеозображень і комп'ютерному зорі інтенсивно використовуються геометричні характеристики, такі як орієнтації, лінії та геометричні перетворення, починаючи від простих (орієнтації, лінії, тверде тіло або афінні перетворення тощо) до складних, таких як криві або поверхні. В сучасних програмних розробках це може створювати певні проблеми. Для їх вирішення потрібно використовувати методи та способи для зменшення невизначеності, до яких належать також афінні перетворення при побудові нейронної мережі.

У [1] розглядаються інваріанти, що є інваріантними відносно загального афінного перетворення і можуть бути використані для розпізнавання афінно-деформованих об'єктів. Описаний підхід базується на теорії алгебраїчних інваріантів. Також описано кілька числових експериментів, що стосуються розпізнавання образів за допомогою інваріантів афінного моменту як ознак образу. Оцінка мільйонів параметрів глибокої CNN вимагає великої кількості анованих зразків, що дозволяє запобігати застосуванню багатьох згорткових глибоких нейронних мереж CNN до проблем із обмеженими навчальними даними. Наприклад, таких як AlexNet, VGG, ResNet. Для вирішення описаної проблеми пропонується новий двофазний метод, що поєднує навчання передачі CNN і розширення вебданих.

Незважаючи на те, що за останні роки згорткові нейронні мережі (англ. convolutional deep neural networks, CNN) досягли великого успіху в багатьох застосуваннях комп'ютерного зору, інваріантність обертання все ще є складною проблемою для таких мереж. Особливо це стосується деяких зображень, вміст яких може відображатися на зображенні під будь-яким кутом. Наприклад, це можуть бути медичні зображення, мікроскопічні зображення або зображення дистанційного зондування та астрономічні зображення. У [2] авторами запропоновано нову операцію згортки, яка називається

Gradient-Aligned Convolution (GAConv) і яка може допомогти CNN досягти інваріантності обертання шляхом заміни ванільних згорток у CNN.

За допомогою, методу представленого в [3], презентація корисних функцій попередньо навченої мережі може бути ефективно передана цільовому завданню, а вихідний набір даних може бути доповнений найціннішими інтернет-зображеннями для класифікації. Описаний метод не тільки значно зменшує потребу у великих навчальних даних, але й ефективно розширює набір навчальних даних. Розширення даних походить від традиційних методів розширення даних, таких як обертання, трансляція, масштабування, повороти, зсув, дзеркальне відображення та збурення кольору, що вирішують проблему обмежених навчальних даних шляхом збагачення навчального набору трансформованими оригінальними зразками. Для зображень із невеликою можливістю змін всередині класу (таких як зміна точки огляду, кадрування, зміна кольорів або застосування афінних перетворень).

У [4] описано систему розпізнавання об'єктів, що використовує новий клас локальних ознак зображення. Функції інваріантні до масштабування зображення, трансляції та обертання, а також частково інваріантні до змін освітлення та афінної або 3D-проекції. Ці особливості мають схожі властивості з нейронами нижньої скроневої кори, які використовуються для розпізнавання об'єктів у зорі приматів. Функції ефективно виявляються за допомогою підходу поетапної фільтрації, яка визначає стабільні точки в просторі. Створюються ключі зображення, які визначають локальні геометричні деформації, представляючи розмиті градієнтом зображення в кількох площинах та в кількох масштабах.

У [5] пропонується алгоритм для розпізнавання афінно-спотворених плоских форм. Сутність алгоритму полягає в такому: кожна фігура ділиться на послідовність упорядкованих афінно-інваріантних сегментів на основі властивостей дескриптора форми простору шкали кривизни (CSS). Потім застосовується алгоритм Сміта – Вотермана для зіставлення пари афінно-інваріантних послідовностей сегментів. Коли отримані відповідності точок вздовж двох форм, оцінюється афінне перетворення. Експерименти показують, що алгоритм є ефективним для розпізнавання форм, які спотворюються афінними перетвореннями, враховуючи обертання, масштабування та зсув. Крім того, він здатний мати справу з випадком, коли одна з форм або обидві форми частково закриті.

У [6] автори розширюють детектор інваріантності масштабу до афінної інваріантності шляхом оцінки афінної форми оточення точки. Ітераційний алгоритм змінює розташування, масштаб і оточення кожної точки та збігається до афінних інваріантних точок. Метод може працювати зі значними афінними перетвореннями, враховуючи великомасштабні зміни. Характерний масштаб і афінна форма оточення визначають афінну інваріантну область для кожної точки.

Інваріанти афінних перетворень є важливими дескрипторами форми в розпізнаванні образів у комп'ютерному зорі. Існуючі методи афінних інваріантів базуються на геометричних і комплексних моментах. У [7] дослідники пропонують набір афінних інваріантів, отриманих з перетворень Лежандра. Ці інваріанти виводяться на основі співвідношення між перетворенням Лежандра до афінно-перетвореного зображення. Продуктивність запропонованого дескриптора оцінюється за допомогою набору бінарних і сірих зображень. Експериментальні результати показують, що запропонований метод поводиться краще, ніж існуючі методи з точки зору точності розпізнавання образів.

Метою статті є дослідження особливостей афінних перетворень для відеозображень у системах паркування, опис розробки математичної моделі перетворення відеозображення для систем комп'ютерного зору та застосування афінних трансформацій для вдосконалення роботи згорткової нейронної мережі.

Викладення основного матеріалу. Для вирішення завдання знаходження вільного місця для паркування під час отримання відеоданих у реальному часі необхідно створити комп'ютеризовану систему, що буде надавати швидкий доступ до інформації про доступність та розташування вільних місць для паркування на обраній парковці міста. Як вхідні дані для системи використовуються кадри з відеопотоку, що отримані з камер відеоспостереження, що встановлені на парковках [8]. Для реалізації розпізнавання використовується архітектура згорткової нейронної мережі Mask R-CNN [9], яка є двоетапною. На першому етапі сканується зображення і генеруються області, які можуть містити деякі об'єкти. На другому – виконується класифікація пропозицій і створення обмежувальних рамок та масок. Однак під час використання Mask R-CNN іноді виникають труднощі. Обчислювальна складність алгоритму розпізнавання відеозображень за геометричними параметрами для визначення вільних місць для паркування автомобілів побудована на основі Mask R-CNN також стикається з труднощами під час роботи мережі в реальних умовах. Модель має багато рівнів і параметрів, що вимагають високоякісного обладнання та ресурсів для навчання й запуску [10]. Крім того, така модель може повільно обробляти зображення для зображень з високою роздільною здатністю або якщо на зображеннях з відеопотоку розпізнано декілька однотипних об'єктів. Це накладає обмеження на застосування моделі для досягнення потрібної продуктивності в реальному часі. Для оптимального функціонування мережі рекомендується використовувати GPU з об'ємом пам'яті від 10 до 12 гігабайт. Окрім того, у Mask R-CNN є обмеження, коли алгоритм працює з нічним (рис. 1), ранковим або вечірнім зображенням (рис. 2).



Рис. 1. Автостоянка, майдан Соборний, місто Житомир, кадр нічної відеозйомки



Рис. 2. Автостоянка, майдан Соборний, місто Житомир, кадр з ранкової відеозйомки

На рисунку 3 зображено автостоянку, що розташована на майдані Перемоги в місті Житомир, зі стертою розміткою, що негативно впливає на навчання нейронної мережі та на розпізнавання вільних місць для паркування на відеозображенні.



Рис. 3. Автостоянка майдан Перемоги, місто Житомир, кадр нічної відеозйомки

На рисунку 4 зображено автостоянку, що розташована по вулиці Героїв Небесної сотні в місті Житомир. Автостоянка має чітку розмітку, але на рисунку видно скошування перспективи, що також впливає на розпізнавання за геометричними параметрами.



Рис. 4. Автостоянка, вулиця Героїв Небесної сотні, місто Житомир

Описана модель показує гарні результати навчання і роботи зі стандартними наборами, але можуть виникати деякі помилки і неточності, особливо якщо дані надходять з відеопотоку в режимі реального часу. Спотворення, викривлення, зміна освітлення, фонові перешкоди можуть вплинути на ефективність і точність. Для вирішення цієї проблеми використовуються методи, що описані нижче.

Для підвищення ефективності визначення вільних і зайнятих місць для паркування з врахуванням геометричних параметрів у системах відеоспостереження пропонується використовувати афінні перетворення, а також метод перетворення вхідного зображення з різного роду геометричних викривлень та можливим фоновим забрудненням (рис. 1–4).

Проведемо розрахунок афінного перетворення за трьома точками. Три точки на площині переходять в інші три точки. Знайдемо це афінне перетворення. Система рівнянь для афінного перетворення виглядає таким чином:

$$\begin{aligned} x' &= a_0 + a_1x + a_2y \\ y' &= b_0 + b_1x + b_2y \end{aligned} \quad (1)$$

де x, y – вхідні координати, x', y' – вихідні координати, $a_0, a_1, a_2, b_0, b_1, b_2$ – відповідно коефіцієнти.

Для розрахунку трансформації відомо початкові координати точок і кінцеві координати (рис. 5).



Рис. 5. Зображення автостоянки з обраними початковими і кінцевими координатами

$$\begin{aligned} X1s &= -3.432001 \\ X2s &= -2.42 \\ X3s &= -3.894 \\ X1e &= -2.04051 \\ X2e &= 2.040509 \\ X3e &= -0.3362256 \\ Y1s &= -2.167 \\ Y2s &= -1.683 \\ Y3s &= -0.9570001 \\ Y1e &= -2.475 \\ Y2e &= -2.475 \\ Y3e &= 2.475 \end{aligned} \quad (2)$$

Задача трансформації зводиться до рішень системи рівнянь і знаходження коефіцієнтів.

$$\begin{aligned} x'_1 &= a_0 + a_1 \cdot (-3.432001) + a_2 \cdot (-2.167) \\ x'_2 &= a_0 + a_1 \cdot (-2.42) + a_2 \cdot (-1.683) \\ x'_3 &= a_0 + a_1 \cdot (-3.894) + a_2 \cdot (-0.9570001) \\ y'_1 &= b_0 + b_1 \cdot (-3.432001) + b_2 \cdot (-2.167) \\ y'_2 &= b_0 + b_1 \cdot (-2.42) + b_2 \cdot (-1.683) \\ y'_3 &= b_0 + b_1 \cdot (-3.894) + b_2 \cdot (-0.9570001) \end{aligned} \quad (3)$$

Для знаходження коефіцієнтів a і b перейдемо до матричної форми:

$$\begin{bmatrix} -2.04051 \\ 2.040509 \\ -0.3362256 \end{bmatrix} = \begin{bmatrix} 1 & -3.432001 & -2.167 \\ 1 & -2.42 & -1.683 \\ 1 & -3.894 & -0.9570001 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \quad (4)$$

$$\begin{bmatrix} -2.475 \\ -2.475 \\ 2.475 \end{bmatrix} = \begin{bmatrix} 1 & 3.432001 & -2.167 \\ 1 & -2.42 & -1.683 \\ 1 & -3.894 & -0.9570001 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}. \quad (5)$$

Знаходимо коефіцієнти $a_0, a_1, a_2, b_0, b_1, b_2$:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 3.432001 & -2.167 \\ 1 & -2.42 & -1.683 \\ 1 & -3.894 & -0.9570001 \end{bmatrix}^{-1} \begin{bmatrix} -2.04051 \\ 2.040509 \\ -0.3362256 \end{bmatrix}, \quad (6)$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 3.432001 & -2.167 \\ 1 & -2.42 & -1.683 \\ 1 & -3.894 & -0.9570001 \end{bmatrix}^{-1} \begin{bmatrix} -2.475 \\ -2.475 \\ 2.475 \end{bmatrix}. \quad (7)$$

Для розв'язання задачі потрібно виконувати операції оберненої матриці та множення матриць. Після розрахунків значення коефіцієнтів будуть дорівнювати:

$$\begin{aligned} a_0 &= 13.1098 \\ a_1 &= 2.840328 \\ a_2 &= 2.492988 \\ b_0 &= -0.6567969 \\ b_1 &= -1.654411 \\ b_2 &= 3.459227 \end{aligned} \quad (8)$$

Далі коефіцієнти використовуються для отримання нових координат для кожного пікселя вихідного растра. На рисунку 6 зображено автостоянку після застосування афінних перетворень.



Рис. 6. Результат афінного перетворення

Афінні перетворення є комбінацією лінійних перетворень, що супроводжуються трансформацією зображень і формують зручну підсистему білінійних перетворень, часто використовуються в комп'ютерній графіці і значно спрощують масштабування, поворот і зсув зображень. Запропонований алгоритм одночасно адаптує розташування, масштаб і форму оточення відносно трьох точок.

Для вдосконалення роботи Mask R-CNN було розглянуто архітектуру згорткової нейронної мережі, визначено оптимальну архітектуру CNN для використання. Для покращення навчання згорткової нейронної мережі додано блок з інваріантними афінними перетвореннями.

Висновки та перспективи подальших досліджень. У дослідженні описано використання афінних перетворень для проведення корекції зображень, що отримуються з відеокамер у режимі реального часу з врахуванням геометричних параметрів об'єктів на зображенні задля можливості подальшого використання в системах розпізнавання для визначення вільних місць для паркування. Відомо, що зображення, отримані з відеокамер, можуть виявляти перспективні спотворення, що походять від неідеальних кутів розташування камери. Зазвичай техніка афінного перетворення використовується для корекції геометричних спотворень або деформацій, таких як масштабування, переміщення, зсув, обернення.

Розроблено математичну модель перетворення зображення для систем комп'ютерного зору. Наведено приклад розрахунку параметрів такої моделі. Запропоновано метод побудови алгоритму перспективних перетворень зображень, що дозволяє спростити виявлення, сегментацію та розпізнавання об'єктів у подальшому під час використання такої мережі.

References:

1. Flusser, J. and Suk, T. (1993), «Pattern recognition by affine moment invariants», *Pattern recognition*, No 26 (1), pp.167–174.
2. Hao, Y., Hu, P., Li, S., Udupa, J.K., Tong, Y. and Li, H. (2022), «Gradient-Aligned convolution neural network», *Pattern Recognition*, No 122, p. 108354.
3. Han, D., Liu, Q. and Fan, W. (2018), «A new image classification method using CNN transfer learning and web data augmentation», *Expert Systems with Applications*, No 95, pp. 43–56.
4. Lowe, D.G. (1999), «September. Object recognition from local scale-invariant features», *In Proceedings of the seventh IEEE international conference on computer vision*, Vol. 2, pp. 1150–1157. IEEE.
5. Mai, F., Chang, C.Q. and Hung, Y.S. (2010), «September. Affine-invariant shape matching and recognition under partial occlusion». *In 2010 IEEE international conference on image processing*, pp. 4605–4608. IEEE.
6. Mikolajczyk, K. and Schmid, C. (2004), «Scale & affine invariant interest point detectors», *International journal of computer vision*, No 60, pp.63–86.
7. Zhang, H. and Wu, Q.J (2011), «September. Pattern recognition by affine legendre moment invariants», *In 2011 18th IEEE International Conference on Image Processing*, pp. 797–800. IEEE.
8. Levkivskiy, V. et al. (2022), «Available parking places recognition system», *CEUR Workshop Proceedings 4th Workshop for Young Scientists in Computer Science & Software Engineering*, Virtual Event, Kryvyi Rih, 18 December 2021, P. 123–134, [Online], available at: <https://ceur-ws.org/Vol-3077/paper07.pdf>.
9. He, K., Gkioxari, G., Dollár, P. & Girshick R. (2017), *IEEE International Conference on Computer Vision, ICCV*, Mask r-cnn, pp. 2980–2988, [Online], available at: <https://ieeexplore.ieee.org/document/8237584>
10. Marchuk, D.K. (2023), «Analiz suchasnykh alhorytmiv vvyavlennya i rozpoznavannya ob'ektiv z video potoku dlya system upravlinnya parkuvannyam v real'nomu chasi», *Visnyk Khmel'nyts'koho natsional'noho universytetu Seriya: Tekhnichni nauky*, No 3 (321), pp. 17–23.

Марчук Дмитро Костянтинович – старший викладач кафедри КН, Державний університет «Житомирська політехніка».

Наукові інтереси:

– інтелектуальний аналіз даних.

Marchuk D.K.**The use of affine transformations for image correction with further use in recognition systems**

Today, neural networks are gaining popularity and are increasingly used in various fields of life. They are used for data analysis, clustering, classification, object detection, or hidden patterns. Neural networks are increasingly used in business forecasting tasks. Computer vision systems are no exception. An important component of these processes is image preprocessing. It is known that images obtained from video cameras can exhibit perspective distortions originating from imperfect camera angles. Usually, the affine transform technique is used to correct geometric distortions or deformations, such as scaling, moving, shifting, rotating. In this paper, we study the features of affine transforms and their application with neural networks. Methods for recognizing geometric parameters in image transformation and moving are considered. Attention is paid to elementary transformations, which include transfer, scaling, shift, rotation. A mathematical model of image transformation for computer vision systems is developed. On the basis of the proposed method, an algorithm for perspective transformations of images obtained from video cameras located in parking lots or city parking lots is built, which greatly facilitates the further detection, segmentation and classification of objects. To improve the performance of the classical Mask R-CNN, a study was conducted in which a block with affine transformations was added to the convolutional neural network. Affine transformations are used to correct the perspective convergence of lines in the frame that are parallel in reality.

Keywords: video images; geometric parameters; computerized system; affine transformations; neural networks.

Стаття надійшла до редакції 02.09.2023.