УДК 004.93'12

Бабійчук А. А., **Сирота О. П.** *Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ*

АНАЛІЗ МЕТОДІВ ВІЗУАЛЬНОГО ВІДСТЕЖЕННЯ ОБ'ЄКТІВ ДЛЯ ВИКОРИСТАННЯ В СИСТЕМАХ РЕАЛЬНОГО ЧАСУ

Досліджуються алгоритми візуального відстеження об'єктів із застосуванням нерозрідженого кодування. Експериментальні дані показали, що використання нерозрідженого кодування дещо зменшує точність відстеження, однак, дає значний приріст в швидкодії, що має кращі перспективи для застосування в системах реального часу. Зроблено висновки щодо напрямку майбутніх досліджень та обрано модель, яка буде базою для подальшого вдосконалення системах відстеження реального часу.

***Ключові слова*:**машинне навчання, візуальне відстеження, система реального часу, розріджене кодування, нерозріджене кодування.

Babiichuk A. A., Syrota O. P. **Object visual tracking methods analysis for usage in real time systems**. . For several years the sparse coding model has been widely used for visual tracking. This model has good accuracy, but low speed. The focus on adaptation/development of visual tracking algorithms that can be applied in real-time systems will significantly expand the application area, it will allow the introduction of visual tracking in production, in recommendations systems that can immediately react to certain actions, autopilots, smart machines and other areas. In this paper several popular algorithms based on sparse coding were selected, and an experimental analysis of their speed and tracking quality was carried out. Also one of the algorithms, which showed the highest accuracy, was modified using of non-sparse coding. Experimental results showed that using of non-sparse coding slightly reduces the accuracy of tracking, however, it gives a significant improvement in speed. Based on this article, conclusions were drawn regarding the direction of future research, and model was selected that would be the basis for further improvement for real-time systems.

***Keywords*:** machine learning, visual tracking, real-time system, sparse coding, non-sparse coding.

1. Вступ. Розпізнавання та відстеження об'єктів — один з напрямів машинного навчання, задачею якого є визначення розташування рухомого об'єкта чи об'єктів на відео в певний проміжок часу. Алгоритм аналізує кадри відео і визначає положення рухомих цільових об'єктів відносно фону. Даний напрям розвивається вже понад 30 років, однак все ще залишається актуальною проблемою для дослідження, через низку невирішених проблем як точності розпізнавання, так і швидкості роботи алгоритмів.

Постановка проблеми. На сьогоднішній день візуальне відстеження об'єктів на відео має переважно наукове дослідницьке застосування та для різного роду спостережень. Застосування відстеження в системах реального часу дозволить значно розширити сферу застосування, зокрема дозволить впроваджувати його на виробництві, в рекомендаційних системах, які одразу можуть реагувати на певні дії, в автопілотах, розумних машинах та інших сферах.

Аналіз останніх досліджень і публікацій. Однією з найпопулярніших моделей для візуального відстеження об'єктів є модель розрідженого кодування. Після значного успіху у сфері розпізнавання обличь [3], Мей, Лінг та ін. перші пропонують розглядати візуальне відстеження як проблему розрідженого кодування [12]. На основі їх роботи було розроблено значна кількість алгоритмів L1 [12], T2CL1 [8], OT2CL1 [9], SCM [11], які будуть розглядатись в даній роботі. Ці алгоритми можуть показувати досить високу точність відстеження, але все ще мають високу складність обчислень та мають недостатню швидкодію для роботи в режимі реального часу.

Однак, в останні роки деякі дослідники в [6, 7]пропонують використання протилежної моделі, нерозрідженого кодування, для задачі розпізнавання обличь, яка показує значно кращу швидкість роботи. На основі їх дослідження є підстави вважати, що дану модель можна також застосувати і для задачі візуального відстеження. Тому в даній статті буде модифіковано алгоритм SCM, на основі робіт [7-6] та досліджень описаних в [10]. Назвемо модифікований алгоритм SCM2.

Крім цього в [10] та [15], надається порівняння точності роботи алгоритмів для візуального відстеження. На основі даних робіт було обрано для порівняння алгоритми, що показали кращі результати точності.

Невирішені проблеми. Популярна модель розрідженого кодування має досить гарні показники точності, однак не дуже високу швидкість роботи. Альтернативна модель нерозрідженого кодування, яка була застосована для задачі розпізнавання обличь показала досить хорошу швидкість роботи і непогану точність. Однак на сьогоднішній день ця модель не застосовується для вирішення задач візуального відстеження. Таким чином не можна судити про доцільність її застосування для візуального відстеження.

Також в розглянутих джерелах не приділяється достатньої уваги питанню швидкості алгоритмів, таким чином не можна зробити висновки щодо можливості використання алгоритмів в системах реального часу.

Мета та задачі дослідження. Проаналізувати продуктивність наявних методів відстеження об'єктів для роботи в режимі реального часу. Порівняти моделі розрідженого та нерозрідженого кодування для вирішення даних задач. Провести експериментальний аналіз швидкодії алгоритмів. На базі дослідження вибрати модель яка буде базою для подальшого вдосконалення, для роботи в системах реального часу.

Отже, в даній роботі буде перевірено експериментальним шляхом продуктивність алгоритмів L1 [12], T2CL1 [8], OT2CL1 [9], SCM [11], SCM2, а також перевірено точність їх роботи. На основі результатів, зроблено висновки, щодо застосування цих алгоритмів для систем реального часу.

2. Моделі візуального відстеження

2.1. Модель розрідженого кодування. Нехай $X \in R^D$– це вектор, отриманий в результаті перетворення пікселів матриці зображення у вектор, тобто всі рядки послідовно конкатинуються. D — це розмірність вектору вхідного зображення.

Розріджене кодування представляє вектор X як лінійну комбінацію набору базисних векторів $V = [v_1 ... v_k] \in R^{D \times K}$ та векторів коефіцієнтів $U = [u_1 ... u_k]^T \in R^K$. Набір базисних векторів V називають словником, а кожний базисний вектор — атомом. Варто звернути увагу, що словник описує як цільові зображення, так і фононові (нерухомі). D — це розширення базисного вектору, K — кількість базисних векторів. Нехай $n \in R^D$– це шум. Тоді з урахуванням шуму вхідне зображення можна описати за допомогою розрідженого представлення наступним чином:

$$\mathrm{X} = \sum_{k=1}^{K} u_k v_k + n. \quad (1)$$

За умови, що словник визначений (певними початковими значеннями), коефіцієнти можна обчислити за наступним рівнянням:

$$u = \arg\min(U) \, ||X - VU||_2^2 \, , \quad (2)$$

а рішення, відповідно, буде:

$$u = (V^T V)^{-1} V^T X \, . \quad (3)$$

Враховуючи, що базисні вектори розріджені, тобто більшість елементів нулі, вектор коефіцієнтів може бути отримано шляхом рішення задачі мінімізації першої норми наступним чином:

$$u = \arg\min(U) \frac{1}{2} ||X - VU||_2^2 + \lambda ||U||_1 \, , \quad (4)$$

де λ (множник Лагранжа) — регулюючий параметр, який знаходить баланс між помилкою перетворення та вагою коефіцієнтів.

Після того, як отримано коефіцієнти U, потрібно навчити словник V на основі набору природних зображень X так, щоб будь-яке зображення x_i може бути розріджено представлене отриманим словником. Нехай U матриця, що складається з коефіцієнтів векторів всіх тренувальних образів. Тоді словник V може бути вивчено шляхом вирішення наступного завдання оптимізації:

$$\min(U, V) \sum_{i=1}^{N} ||x_i - V u_i||_2^2 + \lambda ||u_i||_1 \, . \quad (5)$$

Таким чином процес розрідженого кодування відбувається у два кроки, що чергуються: 1) після призначення словнику початкових значень обчислюються коефіцієнти за (4); 2) потім використовуючи ці коефіцієнти навчається словник V за (5). І так повторюється поки не буде отримано оптимальний результат.

Розглянемо складність алгоритмів даної моделі. Основна вартість обчислень припадає на вирішення задачі мінімізації першої норми. При чому значний вплив на продуктивність має як обчислення кожної мінімізації окремо так велика кількість ітерацій мінімізації. Обчислювальна складність кожної мінімізації першої норми складає $O(D^2K^{3/2})$, де D розмірність кожного базисного вектору, а K – кількість базисних векторів.

Нехай N – кількість обчислень мінімізації першої норми. Таким чином, обчислювальна складність алгоритму буде рівна $O(ND^2K^{3/2})$. Для алгоритму L1 значення N буде рівне кількості цільових кандидатів, в даній статті це значення – 600. Для T2CL1 та OT2CL1 значення N, це число цільових показників, N=1. Для SCM N – це число частин зображень з цільового шаблону, тут 720.

2.2. Модель нерозрідженого кодування. В цьому методі ми працюємо з тими ж даними, що і в розрідженому кодуванні $X \in R^D$, $V = [v_1 \ldots v_k] \in R^{D\times K}$, $U = [u_1 \ldots u_k]^T \in R^{D\times K}$. Однак для зменшення обчислювальної вартості в словнику не використовуються шаблони фону, а лише цільові шаблони. Коефіцієнти так само обчислюються на основі (2). А рішення, відповідно можна обчислити за (3).

Однак, в даному випадку базисні вектори не розріджені, а отже може існувати лінійна залежність між двома або більше стовпців V, що приводить до значного зменшення точності МНК. Стовпці в цьому випадку називаються мульти-колінеарні. Отримана лінійна система, стає погано обумовлена і не дає нам отримати точний розв'язок системи. За таких умов рішення може бути отримано через мінімізацію другої норми.

$$u = \arg\min(U)\frac{1}{2}||X - VU||_2^2 + \lambda||U||_2 . \quad (6)$$

Процес навчання словника і визначення коефіцієнтів відбувається подібно як і в методі розрідженого кодування. Однак варто виділити такі основні відмінності між цими двома методами:

– використання меншої кількості шаблонів в словнику, оскільки в нерозрідженому методі не використовуються шаблони фону, яких, як правило значно більше;

– використання мінімізації другої норми замість першої.

3. Перевірка продуктивності

Алгоритми. Для перевірки продуктивності було використано наступні алгоритми: L1 [12], T2CL1 [8] та OT2CL1 [9], SCM [11] та модифікований алгоритм SCM2, з використанням мінімізації другої норми та нерозрідженого кодування. В даному експерименті було використано вихідні коди авторів даних алгоритмів, однак для справедливого порівняння було змінено деякі параметри, таким чином як в [12]: фіксований розмір частинок 32×32; кількість частинок 600; розміри словників: цільового – 50 та фонового – 200 (лише для алгоритмів з мінімізацією першої норми); λ дорівнює 0.01.

Слід зазначити, що, хоча було використано вихідні коди авторів, не вдалось точно відтворити результати отримані в їхніх роботах, зокрема через різні параметри та тестові вибірки.

Дані. Експерименти проводилися таких загальнодоступних відео-фрагментах, які включають: авто*, обличчя, птахи, мистецтво, улюбленці.* Дані відеофрагменти можна знайти за посиланнями:

http://www.cs.technion.ac.il/~amita/fragtrack/fragtrack.htm,http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/ ;

http://www.cvg.rdg.ac.uk/PETS2001/ .

Використовувались фрагменти різного розширення, з різних сцен та з різними умовами розпізнавання. Розширення та цільові розміри приведено в таблиці 1.

Розширення фрагментів Таблиця 1

Назва фрагменту	**Розширення**	**Розширення цільового об'єкту**
Обличчя	320×240	98×82
Авто	360×240	88×104
Птах	720×400	37×31
Мистецтво	320×240	79×61
Улюбленці	79×24	768×576

Оцінка точності. Для порівняння точності було використано один з основних критеріїв, а саме помилка розташування центру (ПРЦ). ПРЦ розраховувалася як відстань у пікселях між передбаченим положенням центру і реальним положенням центру. Дані наведено в Таблиці 2.

Порівняння точності алгоритмів на основі оцінки ПРЦ Таблиця 2

Фрагмент	L1	T2CL1	OT2CL1	SCM	SCM2
Обличчя	30.62	27.0	4.89	7.13	7.40
Авто	88.11	22.23	21.41	1.67	1.70
Птах	161.26	84.50	140.70	46.77	47.02
Мистецтво	251.38	27.50	1.19	1.42	5.99
Улюбленці	58.01	2.33	2.3286	2.31	6.15

Оцінка продуктивності. Для перевірки використовувався комп'ютер з наступними характеристиками: процесор: 2,6 GHz Intel Core i5 та оперативна пам'ять: 8 ГБ 1600 MHz DDR3. Було обрано параметри комп'ютера вказаної потужності, так як для застосування візуального відстеження на промисловому рівні, варто орієнтуватись на пристрої середньої потужності. Результати будуть порівнюватись в кадрах за секунду. Дані наведено в Таблиці 3.

Порівняння продуктивності алгоритмів (сек/кадр) Таблиця 3

Фрагмент	L1	T2CL1	OT2CL1	SCM	SCM2
Обличчя	5.16	0.90	1.00	5.11	0.22
Авто	4.1	0.84	0.88	3.87	0.08
Птах	7.02	1.14	1.21	6.18	0.31
Мистецтво	7.99	1.17	1.23	6.22	0.29
Улюбленці	5.12	0.89	0.97	4.59	0.10

Оцінка продуктивності в системах реального часу. Система реального часу (СРЧ) – це така система, яка має реагувати на зміни у зовнішньому по відношенню до системи середовищі та миттєво (для сприйняття людини) видавати результат. Якщо розглядати СРЧ в контексті вирішення задач обробки відео, то маємо наступне. Більшість систем які працюють з відео мають швидкість 60 кадрів в секунду, тому оптимальна швидкість роботи з відео має бути 0.017 сек/кадр. Однак, для людини достатньою є швидкість 24 кадри в секунду, тому можна допустити меншу швидкість обробки, при цьому не помітно для людського ока, а саме 30 кадрів в секунду. Отже, маємо 0.033 сек для обробки 1 кадру.

Порівняння результатів. Спершу розглянемо алгоритми на основі розрідженого кодування. Що до точності відстеження L1 має найгірші показники, так як це перший алгоритм на даній моделі, такі результати очікувані. Найкращу точність на більшості вибірках показує алгоритм SCM. Порівнюючи швидкість роботи, алгоритми T2CL1 та OT2CL1 мають значно кращі показники продуктивності ніж L1 та SCM. Однак навіть при цьому витрачають трохи менше секунди для обробки 1 кадру, тоді як для роботи в режимі реального часу необхідно 0.033 секунди. Застосування цих алгоритмів в системах реального часу потребує значного покращення продуктивності, однак на сьогоднішній день, результати далекі від бажаних.

При використанні мінімізації другої форми та нерозрідженого кодування можна отримати значне покращення продуктивності, до 0.08 сек/кадр на наших тестових даних. Такий результат є досить близьким до бажаного. Таким чином можна відзначити ефективність застосування нерозрідженого кодування та мінімізації другої норми при відстеженні об’єктів. Щодо точності відстеження,то *SCM2* показав дещо гірші показники на фрагментах улюбленці та мистецтво. Проте, на решті фрагментах, отримано точність близьку до тієї, яку дають методи розрідженого кодування. Що вказує на те, що мінімізація першої норми, що використовується в більшості сучасних алгоритмів може бути замінені мінімізацію другої норми, а розріджене кодування, нерозрідженим.

На основі цих результатів перспективним видається використання моделі нерозрідженого кодування з мінімізацією другої норми. Хоч точність дещо гірша, вона не значно поступається точності розрідженого кодування.

4. Висновки

На основі порівняння точності та продуктивності алгоритмів *L1*, *T2CL1*, *OT2CL1*,*SCM* та *SCM2* найбільш перспективним для нашої задачі є алгоритм *SCM2*, проте необхідно далі працювати над зменшенням похибки алгоритму та зменшенням його складності.

На основі результатів даної статті визначений наступний напрямок майбутніх досліджень – це модифікація алгоритмів на основі модель нерозрідженого кодування з мінімізацією другої норми.

Список використаної літератури. References

1. Arnold W M Smeulders, Senior Member, Dung M Chu, Student Member, Rita Cucchiara, and Simone Calderara. Visual Tracking : An Experimental Survey. 36(7): 1442–1468, 2014.

2. C. Bao, Y. Wu, H. Ling, H. Ji, Real time robust l1 tracker using accelerated proximal gradient approach, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1830–1837

3. John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell.,31(2):210–227, February 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.79. URL http://dx.doi.org/10.1109/TPAMI.2008.79.

4. K. Yu, Y. Lin, J. Lafferty, Learning image representations from the pixel level via hierarchical sparse coding, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition, 2011, pp. 1713–1720.

5. K. Zhang, L. Zhang, M. Yang, Real-time compressive tracking, European Conference on Computer Vision, in: Proceedings of the 12th European Conference on Computer Vision, 2012, pp. 864–877

6. Q. Shi, A. Eriksson, A. Hengel, C. Shen, Is face recognition really a compressive sensing problem, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition, 2011, pp. 553–560.

7. R. Rigamonti, M. Brown, V. Lepetit, Are sparse representations really relevant for image classification, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition, 2011, pp. 1545–1552.

8. S. Zhang, H. Yao, X. Sun, S. Liu, Robust object tracking based on sparse representation, in: Proceedings of the International Conference on Visual Communications and Image Processing, pp. 77441N-1-8, 2010.

9. S. Zhang, H. Yao, H. Zhou, X. Sun, S. Liu, Robust visual tracking based on online learning sparse representation, Neurocomputing 100 (2013) 31–40.

10. Shengping Zhang, Hongxun Yao n , Xin Sun, Xiusheng Lu, Sparse coding based visual tracking: Review and experimental comparison, Pattern Recognition 46 (2013) 1772–1788

11. W. Zhong, H. Lu, M. Yang, Robust object tracking via sparsity-based collaborative model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1838–1845.

12. X. Mei and H. Ling. Robust visual tracking using L1 minimization. Proceedings of the 12th International Conference on Computer Vision, pages 1436–1443, 2009.

13. X. Yuan, S. Yan, Visual classification with multi-task joint sparse representation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3493–3500

14. Yashar Deldjoo, Reza Ebrahimi Atani. A low-cost infrared-optical head tracking solution for virtual 3d audio environment using the nintendo wii-remote. Entertainment Computing, 12:9–27, 2016.

15. Yashar Deldjoo, Shengping Zhang, Bahman Zanj, Paolo Cremonesi, Matteo Matteucci, Sparse vs. Non-sparse: Which One Is Better for Practical Visual Tracking?, Computer Vision and Pattern Recognition, 2016

16. Zhangjian Ji and Weiqiang Wang. Object tracking based on local dynamic sparse model. Journal of Visual Communication and Image Representation, 2015.

Автори статті

Бабійчук Андрій Анатолійович – аспірант кафедри технічної кібернетики, Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ. Тел. +38 096 1017329. E-mail: babiychukandrey@gmail.com

Сирота Олена Петрівна – кандидат технічних наук, старший викладач кафедри технічної кібернетики, Національнйи технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ. Тел. +38 067 209 94 74. E-mail: sirotae@gmail.com

Authors of the article

Babiichuk Andrii Anatoliiovych – PhD student of technical cybernetics department, National Technical University of Ukraine ”Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv. Tel.: +380 (96) 101 73 29. E-mail: babiychukandrey@gmail.com

Syrota Olena Petrivna – candidate of science (technical), senior lecturer of technical cybernetics department, National Technical University of Ukraine ”Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv. Tel.: +380 (67) 209 94 74. E-mail: sirotae@gmail.com

Дата надходження в редакцію: 27.04.2017 р.

Рецензент:
доктор технічних наук, професор К. С. Козелкова
Державний університет телекомунікацій, Київ