

Warszawa, dn. 9 kwietnia 2021

dr hab.inż. Marcin Iwanowski, prof.uczelni
Politechnika Warszawska, Wydział Elektryczny
Instytut Sterowania i Elektroniki Przemysłowej

Recenzja rozprawy doktorskiej mgr inż. Anety Bery pt. „Szybka ekstrakcja cech obrotowo-niezmiennej za pomocą obrazów całkowych w zadaniach detekcji”

1 Zawartość rozprawy

Recenzowana rozprawa liczy 110 stron, w tym 97 stron tekstu zasadniczego. Została napisana w języku polskim i składa się z nienumerowanego wprowadzenia oraz ośmiu rozdziałów.

Wprowadzenie zawiera ogólne omówienie obszaru tematycznego pracy, w tym uczenia maszynowego, ze szczególnym uwzględnieniem kluczowego – z punktu widzenia tematyki pracy – jego etapu przygotowawczego, czyli doboru cech obrazów. W ostatnim akapicie omówiona jest struktura dalszej części dysertacji. W osobnym, nienumerowanym, punkcie podane zostały cel, motywacja i teza pracy.

Pierwszy rozdział pracy zawiera wstęp teoretyczny, który składa się z czterech punktów. W pierwszym omówiono zadanie detekcji i typowe sposoby jej przeprowadzania na obrazach cyfrowych. Kolejny punkt jest poświęcony obrazowi całkowemu, następne dwa – technikom klasyfikacji, które zostały wykorzystane w eksperymentach opisanych w pracy: metodom łączenia słabych klasyfikatorów (boosting) oraz samym tego typu klasyfikatorom.

Rozdział drugi jest omówieniem kluczowych zagadnień i pojęć związanych z momentami Zernika i jako taki stanowi wprowadzenie teoretyczne do zasadniczych rozważań zawartych w pracy. W pierwszym punkcie rozdziału przedstawiono niezbędne definicje poczynając od wielomianów Zernika. Kolejny punkt jest poświęcony kluczowemu z punktu widzenia pracy zagadnieniu niezmienniczości względem obrotu i skalowania. W kolejnych dwóch punktach omówiono związki momentów Zernika z momentami Fouriera-Mellina. Ostatni punkt zawiera przegląd zastosowań obu rodzajów momentów w literaturze.

Rozdział trzeci zawiera najważniejsze osiągnięcie Autorki - algorytm wyznaczania momentów Zernika o złożoności niezależnej od wielkości okna detekcji. W pierwszym punkcie opisano podstawowe przekształcenia niezbędne do wykazania poprawności zaproponowanego rozwiązania. W drugim punkcie pokazano związek obrazów całkowych ze stałym czasem obliczenia momentów bez względu na wielkość okna. W trzecim podpunkcie przedstawiono podstawowe własności rozwiązania, zaś w kolejnym – autorską koncepcję zwiększenia liczby cech poprzez podział okna detekcji na współśrodkowe pierścienie. W ostatnim, piątym punkcie omówiono sposoby przyspieszenia obliczeń z wykorzystaniem tablicy korekcji.

Wyniki testów, zaproponowanej przez Autorkę, metody zostały przedstawione w rozdziale czwartym. Rozdział ten prezentuje rezultaty wstępnych eksperymentów walidacyjnych metody opisanej

w rozdziale poprzedzającym. Eksperymenty zostały przeprowadzone na dwóch zbiorach danych. Ich pierwsza część polegała na detekcji znaków litery 'A' losowo rozmieszczonych na sztucznie wygenerowanych obrazach ze zróżnicowanym tłem. Druga część eksperymentów dotyczyła detekcji sylwetek samolotów na zdjęciach płyt lotnisk. Wyniki przedstawione w rozdziale czwartym są wynikami cząstkowymi, które stanowiły motywację do podjęcia prac badawczych opisanych w dalszej części pracy.

W rozdziale 5 poruszane jest zagadnienie redukcji błędów numerycznych, występujących w metodzie podstawowej, opisanej w rozdziale 3, a które zostały zaobserwowane podczas testów opisanych w rozdziale 4. Rodział rozpoczyna się od analizy przyczyn powstawiania tych błędów. Następnie została zaproponowana technika redukcji błędów numerycznych poprzez podział obrazów całkowych, zaś na końcu rozdziału – omówienie dodatkowych obliczeń wymaganych do poprawnego stosowania zaproponowanej techniki.

W rozdziale szóstym omówiono rozszerzoną przestrzeń niezmienników Zernika, w tym – w podpunktach – schemat indeksowania przez grupowanie iloczynów oraz procedurę ekstrakcji cech.

Rodział siódmy zawiera opis zrealizowanych eksperymentów zasadniczych metody podstawowej wraz z zaproponowanym mechanizmem redukcji błędów. W eksperymentach zasadniczych wykorzystano te same zbiory danych, które zostały użyte w eksperymentach wstępnych. Ich wyniki opisano w dwóch pierwszych punktach rozdziału. W ostatnim podpunkcie zaprezentowano wyniki eksperymentów wydajnościowych.

Rodział ósmy jest podsumowaniem pracy i zawiera konkluzje wraz z krytyczną analizą osiągniętych wyników oraz wskazaniem dalszych kierunków badań.

Na końcu pracy zamieszczono spisy rysunków, tablic, algorytmów oraz bibliografię zawierającą 77 pozycji literatury, w tym dwie pozycje autorskie.

2 Ocena pracy

Metody i techniki analizy treści obrazowych, na przestrzeni ostatnich kilkadziesiąt lat, okazały się jednymi z najbardziej intensywnie rozwijających się obszarów w naukach technicznych. Przyczyną tego jest gwałtowny rozwój narzędzi do pozyskiwania, gromadzenia, przesyłu i przetwarzania obrazów pozwalający na tworzenie coraz bardziej efektywnych modeli obliczeniowych. Modeli, które z uwagi na bardzo dużą różnorodność treści obrazowych i mnogość możliwych do realizacji zadań, stają się coraz bardziej złożone i zasobożerne. Dlatego, niezależnie od rozwoju po stronie sprzętowej, niezwykle ważna jest optymalizacja algorytmów obróbki obrazu cyfrowego na wszystkich poziomach, od przetwarzania wstępnego, filtracji, segmentacji, wyznaczania cech aż do etapu stosowania narzędzi uczenia maszynowego.

W tym kontekście tematyka recenzowanej pracy jest aktualna, zaś jej wyniki mogą przyczynić się do poprawy efektywności algorytmów wizyjnych. Celem pracy jest bowiem opracowanie algorytmu umożliwiającego szybkie wyznaczanie momentów Zernika należących do grupy klasycznych cech momentowych obrazów cyfrowych. Tak zdefiniowany cel pozwolił na precyzyjne sformułowanie tezy badawczej, którą jest stwierdzenie, że możliwa jest realizacja algorytmu, realizującego zdanie ekstrakcji momentów Zernika w czasie niezależnym od wielkości okna detekcji.

Metoda podstawowa została przedstawiona w artykule A.Bera, P.Kłęsk, D.Sychel "Constant-Time Calculation of Zernike Moments for Detection with Rotational Invariance", opublikowanym w 2019 roku, w prestiżowym czasopiśmie IEEE Pattern Analysis and Machine Intelligence, którego

Doktorantka jest pierwszą autorką. Metoda ta została opisana szczegółowo w rozdziale trzecim pracy. Metoda podstawowa jest wartościowym wkładem w teorię i praktykę metod momentowych. Dzięki zastosowaniu zbioru zespolonych obrazów całkowych (wzór 3.14), które kumulują wartości pośrednie wymagane do wyznaczenia momentów Zernika z wykorzystaniem operatora przyrostu (wzór 3.18), wyznaczanie wartości momentów nie zależy od wielkości okna. Zaproponowane podejście jest przykładem rozwiązania w którym poprawę paramterów wydajnościowych uzyskuje się kosztem zwiększenia wymagań pamięciowych, w tym przypadku niezbędnych do przechowywania zbioru obrazów całkowych. Jest ono ciekawym i zręcznym wykorzystaniem koncepcji obrazu całkowego do optymalizacji czasowej algorytmu wizyjnego.

Drugie, także udokumentowane stosowną publikacją, osiągnięcie jest związane ze zwiększeniem dokładności metody podstawowej poprzez podział obrazów całkowych. Ścisłej rzecz ujmując, obrazy te są wyznaczane nie dla całego obrazu jak w koncepcji oryginalnej, lecz dla mniejszych podobrazów, na które jest dzielony obraz wejściowy. Dzięki takiemu podejściu negatywne efekty numeryczne wynikające z wielkości iloczynów kumulowanych w obrazie całkowym są znacząco zredukowane. Koncepcja została opisana w artykule konferencyjnym P.Kłęsk, A. Bera, D.Sychel "Reduction of Numerical Errors in Zernike Invariants Computed via Complex-Valued Integral Images" z roku 2020. Także i to osiągnięcie należy ocenić zdecydowanie pozytywnie.

Weryfikacja obu metod została przeprowadzona dwuetapowo. Celem pierwszego etapu weryfikacji było uzyskanie potwierdzenia skuteczności metody podstawowej oraz określenie warunków brzegowych stosowalności podejścia. Jak wykazały eksperymenty, wraz ze zwiększaniem się rozmiaru obrazu, rosły błędy numeryczne. Efektem wniosków z pierwszego etapu testów było opracowanie metody redukcji tych błędów. Efekty tej redukcji są pozytywne, co pokazują wyniki drugiego etapu testów.

Oprócz dwóch zasadniczych osiągnięć opisanych powyżej Autorka opisała w pracy także i inne, mniejszego kalibru. Jest nim zwiększenie liczby cech przez wyznaczanie momentów we fragmentach okna, tzw. pierścieniach. Takie rozwiązanie, polegające na wyznaczaniu cech w określonych obszarach obrazu, prowadzi do zwiększenia liczby cech i istotnie może przyczynić się do poprawy skuteczności identyfikacji obiektu przez zwiększenie ilości informacji w procesie uczenia.

Warto jednak zaznaczyć, co zresztą Autorka stwierdza w podsumowaniu pracy, iż momenty Zernika nie należą do najbardziej efektywnych wskaźników opisujących treść obrazu. Lepsze efekty detekcji można uzyskać prostszymi i szybszymi algorytmami, takimi jak np. cechy Haara. Ich bezapelacyjną zaletą jest jednak niezmienniczość względem skali i obrotów. Powyższa konkluzja nie neguje w żaden sposób wartości recenzowanej pracy. Jest ona bowiem efektem solidnej i dobrze udokumentowanej pracy badawczej wykonanej przez Autorkę.

3 Uwagi krytyczne

Lektura pracy pozwoliła na sformułowanie następujących uwag krytycznych do dalszej dyskusji:

1. **Złożoność obliczeniowa, a dane wejściowe.** Cel jaki został postawiony przez Autorkę i teza pracy jest całkowicie zrozumiała, i została w pracy skutecznie udowodniona. Jednak użyte w tezie pracy pojęcia "w czasie stałym" i oznaczenie " $O(1)$ " nie oddaje w pełni specyfiki zagadnienia. Złożoność obliczeniową definiuje się jako funkcję rozmiaru danych wejściowych. Złożoność $O(1)$ odnosi się do sytuacji, w której czas realizacji algorytmu nie zależy od rozmiaru tych danych. W odniesieniu do wielkości okna detekcji tak jest w istocie. Jednak

w przypadku metod obrazowych najczęściej za rozmiar danych wejściowych przyjmuje się rozmiar obrazu wejściowego, liczbę jego pikseli. Ostateczną miarą efektywności algorytmu jest bowiem możliwość przetworzenia kompletnego obrazu. Określenie użyte przez Autorkę odnosi się do efektywności wyznaczania cech momentowych w pojedynczym oknie detekcji. Liczba takich okien jest jednak na tyle duża, że wpływa znacząco na efektywność algorytmu rozpatrywaną w odniesieniu do obrazu wejściowego. Na tę ostatnią wpływa także rozmiar samego obrazu, im bowiem jest on większy, tym więcej trwa wyznaczanie obrazu całkowego (złożoność wynosi w tym przypadku $O(n)$, przy czym n jest liczbą pikseli obrazu wejściowego). Jak należałoby oszacować złożoność algorytmu w funkcji danych wejściowych w powyższym rozumieniu ?

2. **Zakres testów.** Teza i cel pracy wprost odnoszą się do efektywności i czasu działania algorytmu. Z tej perspektywy zakres eksperymentów mógłby być szerszy i w większym zakresie dotyczyć zagadnień efektywnościowych. Także w kontekście złożoności algorytmu, ale z punktu widzenia danych wejściowych tj. obrazu na którym przeprowadzana jest detekcja oraz liczby okien detekcji (punkt 1 powyżej). Tego typu eksperymenty miałyby istotne znaczenie dla możliwości praktycznego zastosowania metody.
3. **Analiza statystyczna cech.** Wątkiem, który przewija się w pracy, jest dobór odpowiedniej liczby cech. Mowa jest m.in. o mechanizmie zwiększania ich liczby, czemu służy koncepcja pierścieni (punkt 3.4, którego kończący fragment brzmi "Warto jednak przypomnieć, że główną motywacją jest generowanie większej liczby cech i zapewnienie algorytmowi uczącemu więcej informacji"). W pracy nie przeprowadzono jednak głębszej analizy statystycznej cech, np. w formie analizy korelacyjnej. Analiza taka przeprowadzona, jeśli nie teoretycznie, to przynajmniej na konkretnym zbiorze danych, pozwoliłaby na stwierdzenie jaka jest najwłaściwsza liczba cech, możliwe przecież, że część spośród cech jest silnie ze sobą skorelowana, co mogłoby spowodować nadmierną ich redundancję. Zrealizowana w pracy weryfikacja cech poprzez klasyfikację jest rozwiązaniem słusznym, jednak powinna zostać poprzedzona bardziej szczegółową analizą cech jako takich.
4. **Przykłady praktyczne.** Ciekawe i wartościowe wyniki zaprezentowane w pracy zostały zweryfikowane na przykładach dość prostych. Z uwagi na koncentrację zakresu pracy na wydajności obliczeniowej, takie przykłady można uznać (z zastrzeżeniem uwag z punktu 2) za wystarczające. Szkoda jednak, że w przeprowadzonych badaniach nie przyjęto szerszego ujęcia tematu, które mogłoby polegać na wykazaniu, że proponowane podejście ma zastosowanie do wybranych zagadnień związanych z realnymi wyzwaniem dla systemów wizyjnych, przyczyniając się do istotnego polepszenia wydajności obliczeniowej. Kilka dobrych przykładów takich wyzwań zostało wymienionych w punkcie 2.6 "Momenty Zernika w literaturze", choćby zastosowania do detekcji symboli na schematach elektrycznych czy detekcja poruszających się obiektów.
5. **Uzasadnienie przyjętych założeń.** Podczas testowania algorytmów przyjęto szereg założeń. Założenia te nie zostały jednak dostatecznie wyraźnie umotywowane. Jednym z podstawowych była liczba wyznaczanych cech momentowych: 375, 450, 540,.. Nie zostało jednak wyjaśnione z czego wynikają takie liczby cech. Innym parametrem jest zakres zmienności okna detekcji, pośrednio wynika on (w przypadku Zbioru 1) z wielkości obrazów zbioru uczącego oraz z podanej sumarycznej liczby okien w zbiorze testowym, nie jest to jednak

informacja kompletna. Nie podano także np. na jakie mniejsze obrazy zostały podzielone obrazy lotnisk na zdjęciach ze Zbioru 2. Jeśli bowiem konkurują z pierwszymi testów na zbiorze 2 jest iż kumulacja wartości liczbowych w obrazie całkowym jest przyczyną błędów numerycznych to nasuwa się pytanie jaka jest maksymalna wielkość obrazu dla której błędy takie okazały się znaczące ?

4 Uwagi redakcyjne

Tekst pracy, w jej warstwie informacyjnej, dobrze relacjonuje uzyskane wyniki. Jednak struktura pracy oraz układ rozdziałów i punktów nie zawsze jest czytelna. Pewne rozwiązania kompozycyjne należy ocenić pozytywnie. Takim jest choćby umieszczenie rozdziału "Rezultaty wstępnych eksperymentów" po pierwszej części osiągnięć własnych Autorki opisanych w rozdziale 3, a drugą częścią związaną z poprawą dokładności (rozdział 5). Takie, dość nietypowe przemieszanie opisu osiągnięć własnych i testów, ma tutaj swoje uzasadnienie – wyniki wstępnych eksperymentów były motywacją do podjęcia prac nad poprawą dokładności.

Można jednak znaleźć w pracy także i mniej fortunate rozwiązania kompozycyjne. I tak, rozdział 1 "Wstęp teoretyczny" nie jest w istocie tym czy być powinien. Wprowadzenie teoretyczne do meritum jest bowiem zawarte w rozdziale drugim, zaś pierwszy zawiera omówienie kwestii w isotcie pomocniczych. Dużo lepszym rozwiązaniem byłoby włączenie punktów 1.1 i 1.2 do podstaw teoretycznych zawartych w rozdziale drugim, wraz z przeniesieniem punktów 1.3 i 1.4 dotyczących klasyfikacji do załącznika – w kontekście celu pracy, rodzaj zastosowanych klasyfikatorów ma znaczenie drugorzędne. Drugim niezbyt fortunnym rozwiązaniem jest rozbięcie przeglądu literatury na dwie części – ogólny znajduje się we wprowadzeniu, zajmując większą jego część, zaś drugi dotyczący momentów Zernika - na końcu rozdziału drugiego. Tutaj lepszym rozwiązaniem byłoby umieszczenie na początku pracy pracy odrębnego rozdziału z szerokim przeglądem literatury. Do nieciągłości kompozycyjnych zaliczyć można także układ i położenie rozdziału 6 pracy. Trudno jest na bazie lektury tego rozdziału zorientować się odnośnie kontekstu i celu stosowania przestrzeni rozszerzonej. Dopiero szersza analiza treści pracy pozwala na pozyskanie wystarczającego zasobu wiedzy w tym zakresie.

W pracy można także znaleźć szereg, mniejszej wagi, niedociągnięć językowych i redakcyjnych, takich jak np.:

- Na str.16 znajduje się następujące zdanie "Od tej pory w tekście pracy f będzie używane jako oznaczenie funkcji obrazu,,. Tymczasem, na str.23 pojawia się inne "w całej pracy używany będzie symbol f dla oznaczenia pewnej funkcji matematycznej, która w zależności od kontekstu może być funkcją dwóch zmiennych ciągłych lub też funkcją obrazu z dwiema zmiennymi dyskretnymi,, , które rozszerza tę pierwotną definicję. Jest to pewna niekonsekwencja.
- Str.37 niepoprawne określenie "zapis na momenty", chodzi przecież o stosowny wzór.
- Na str.40 użyto wdzięcznych, choć raczej nie używanych w takim kontekście, określeń procesu przeglądania obrazu piksel po pikselu: "obraz jest przemierzany (...) oknem", "obraz jest przechodzony oknem".
- Strona 41 określenie "współrzędne zawierają połówki" jest zrozumiałe, lecz dość kolokwialne.

- Na stronach 45 i 46 znajduje się szereg wzorów w których po lewej stronie znajduje się podwójna suma jedynek. Lepszym rozwiązaniem byłoby wpisane w tym miejscu stosownej wielkości liczbowej, wynikającej zakresów sumowania. Użyty zapis, odnosi się do obrazów całkowych, lecz w zupełności wystarczałoby użycie takiego zapisu jednokrotnie wraz z odpowiednim wyjaśnieniem.
- Str.56 oraz w wielu innych miejscach tekstu użyto znaku kropki ”.” jako separatora części ułamkowej. O ile jest to typowe i poprawne dla języka angielskiego, to w przypadku tekstu napisanego w języku polskim – już nie. Separatorem części ułamkowej w języku polskim jest przecież znak przecinka ”,”.

5 Podsumowanie

Podsumowując, niezależnie od powyższych uwag krytycznych, które w większości są elementem dyskusji nad pracą oraz sugestią ewentualnych dalszych prac badawczych, uważam, że Autorka pracy wykazała swoje wysokie kompetencje w wybranym przez siebie obszarze badawczym. Recenzowana rozprawa zawiera oryginalne rozwiązania problemów naukowych oraz pokazuje umiejętności Autorki w zakresie samodzielnego prowadzenia pracy naukowej. **Całościowa ocena rozprawy jest pozytywna.** Uważam, że rozprawa spełnia wymagania stawiane rozprawom doktorskim przez stosowne, aktualnie obowiązujące, akty prawne. Dlatego **wnioskuję o dopuszczenie rozprawy doktorskiej mgr inż. Anety Bery do publicznej obrony.**

