CZY ZMNIEJSZANIE ILOŚCI DANYCH POCHODZĄCYCH Z SYSTEMU WIZYJNEGO HUMANOIDA MOŻE ZWIĘKSZYĆ JEGO MOŻLIWOŚCI?

Michał Podpora

Instytut Automatyki i Informatyki
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Opolska
ul. K. Sosnkowskiego 31, 45-272 Opole
ravyr@klub.chip.pl

Streszczenie: W artykule autor rozważa możliwość wykorzystania klastra komputerowego jako mocy obliczeniowej dla systemu wizyjnego humanoida. Biorąc pod uwagę sposób działania transmisji informacji między okiem a mózgiem u ludzi, rolę sakkad w tym procesie, jak również ideę funkcjonowania sieci HTM, zaproponowana zostaje transmisja spreparowanej struktury danych (zamiast kompletnej klatki obrazu) z dodatkowym sprzężeniem zwrotnym w postaci współrzędnych najbardziej interesującego fragmentu obrazu – odpowiednika „plamki żółtej". Dzięki temu w przyszłości stanie się możliwe wykorzystanie klastrów komputerowych, co z kolei zwiększy możliwości implementacji „inteligencji" humanoida.

1 Wstęp

Trudno precyzyjnie powiedzieć, kiedy robotyka pojawiła się w historii naszej cywilizacji. Według wielu źródeł nastąpiło to wraz z pierwszym użyciem słowa „robot", jednak niektórzy naukowcy są zdania, że o wiele wcześniej. Niezależnie od tego, początki robotyki są niezaprzeczalnie powiązane ze światem fantastyki i to właśnie literatura w znacznym stopniu zdefiniowała sposób, w jaki społeczeństwo rozumie pojęcie „robotyka". W latach trzydziestych XX wieku mianem robota określano autonomiczne urządzenie techniczne bardziej zbliżone do dzisiejszej definicji humanoida. Robotem potocznie nazywano maszynę mechaniczną o człekokształtnej budowie, zaprojektowaną, by wspomagać człowieka w jego codziennych czynnościach. W 1939 roku na targach światowych w Nowym Yorku, podczas prezentacji takiego właśnie „robota" powiedziano, że do końca wieku każdy będzie posiadał swojego własnego robota-pomocnika.

Po wojnie rozwój techniki znacznie przyspieszył, a roboty faktycznie zawitały pod strzechy. Nie były to jednak takie niezwykłe i inteligentne maszyny, o jakich mówiono wcześniej. Zmianie uległo podejście społeczeństwa do zagadnienia i definicji robota. Dziś pod tym pojęciem społeczeństwo rozumie jedynie mechaniczne urządzenie, które automatycznie wykonuje określone czynności. Mimo to ludzkość nie porzuciła nadziei na to, że „roboty" kiedyś staną się bardziej

uniwersalne, wszechstronne, a przede wszystkim bardziej inteligentne w naszym ludzkim rozumieniu.

Co kilka lat masmedia pokazują kolejnego robota, który prezentowany jest jako szczyt techniki, z komentarzem zupełnie podobnym do wspomnianego z targów w 1939 roku.

Wielu uważa, że najbardziej zaawansowanymi badaniami może się pochwalić firma Honda, od lat pracująca nad rozwojem humanoidów. Najnowszy model – Honda Asimo – potrafi naprawdę wiele. Wśród obserwatorów zdarzają się jednak sceptycy, którzy wbrew powszechnemu zachwytowi, wytykają niedociągnięcia. Przyglądają się, czy na podłodze są znaki kalibracyjne, czy robotowi przeszkadzają flesze aparatów, oceniają, na ile robot jest autonomiczny.

1.1 Wstęp: robot autonomiczny a istoty żywe

Humanoid jest dość specyficznym rodzajem robota. W jego przypadku zaprogramowane/wyuczone sekwencje ruchów są często zupełnie nieprzydatne. Co więcej, środowisko w jakim ma funkcjonować, ewolucja przez wiele lat optymalizowała dla ludzi, a nie dla robotów. Właśnie dlatego, aby humanoid mógł dobrze współistnieć u boku człowieka, potrzeba dołożyć wszelkich starań, by jego sposób postrzegania otoczenia i interakcji był maksymalnie zbliżony do biologicznych odpowiedników. Takie podejście często bardzo komplikuje budowę lub zasadę działania robota, czasami jednak jest źródłem bardzo cennych pomysłów. W mechanice podobieństwa są oczywiste – sposób poruszania się współczesnych humanoidów można określić mianem dużego sukcesu. Natomiast z punktu widzenia programisty trudno znaleźć podobieństwa pomiędzy współczesnymi humanoidami a ludźmi. Humanoid po prostu jest „zaprogramowany”, by działał w określony, bardziej lub mniej autonomiczny sposób.

2 System wizyjny humanoida

Podobnie jak „inteligencja” robota wynika bezpośrednio z umiejętności danego programisty lub zespołu programistów, tak samo funkcjonowanie systemu wizyjnego w dużej mierze zależy od zastosowanych algorytmów. Niestety w przypadku systemów wizyjnych programiści bardzo niechętnie korzystają z doświadczenia neurobiologii. Jedyne podobieństwa to powszechnie stosowana stereowizja oraz proste algorytmy wykrywania ruchu. Dalsza analiza i obróbka sygnału projektowana jest w oparciu o rozmaite filtry i przekształcenia. Niemal zawsze przetwarzany jest kompletny obraz, w jakości najlepszej możliwej do pozyskania z kamer danego robota. Podejście takie zapewnia najwyższą rozdzielczość, a więc i dużą dokładność przekształceń, z drugiej jednak strony w znacznym stopniu zwiększa koszty obliczeń (głównie czas i moc procesora).

Dysponując potężnym stacjonarnym komputerem można sobie pozwolić na masywne obliczenia i pewne opóźnienia czasowe między wejściem (obraz) a wyjściem (sterowanie). Tworząc aplikację, która na podstawie informacji z systemu wizyjnego miałaby sterować ruchem humanoida, nie tylko trzeba dopasować się do realiów zmniejszonego poboru prądu i mniejszej wagi sprzętu (nie zawsze będzie do dyspozycji taki komputer, jaki sugerowałby programista), ale co ważniejsze, robot

musi być sterowalny (przynajmniej w zakresie niektórych ruchów) w czasie rzeczywistym. Przy takich założeniach może być potrzebna weryfikacja, które algorytmy będą wydajne, a z których trzeba zrezygnować.

Tymczasem rozwiązanie stosowane od wielu lat (z powodzeniem) przez naturę, nie polega na analizie kompletnej informacji z całego pola widzenia.

2.1 System wizyjny oparty na kompletnym strumieniu wideo

System wizyjny humanoida zazwyczaj dostarcza informacji o jego otoczeniu w postaci pojedynczego strumienia wideo (lub dwóch – w przypadku niektórych implementacji stereowizji). Obraz przesyłany w ten sposób jest bardzo łatwy do przetwarzania przy pomocy rozmaitych filtrów i przekształceń. Aby możliwe było obserwowanie/rozpoznawanie określonych obiektów i/lub sytuacji, należy zapewnić określone parametry sprzętowe kamer, jak również uruchomić je w odpowiednim trybie. Niestety rozdzielczość obrazu, głębia koloru i stopień kompresji mają nie tylko wpływ na precyzję, ale również na czas i złożoność wszystkich przekształceń dokonywanych na obrazie.

Rys. 1. Obraz pochodzący ze standardowej kamery USB; rozdzielczość: 320x240 pikseli; kolory: 256 w skali szarości; kompresja: sprzętowa JPEG. Parametry obrazu są niewystarczające aby rozpatrywać szczegóły obrazu, jednocześnie zwiększenie rozdzielczości spowoduje m.in. wprowadzenie dodatkowego opóźnienia.

2.2 Możliwości rozbudowy systemu

System wizyjny zbudowany w oparciu o strumień video jest bardzo niewygodny, jeżeli chodzi o wszelkiego rodzaju modernizację czy rozbudowę. Każdy dopisany filtr czy procedura wnioskowania powoduje wydłużenie czasu oczekiwania między wejściem robota (obrazem z kamer) a jego sterowaniem. Dzieje się tak, ponieważ większość operacji jest wykonywana sekwencyjnie.

W świecie komputerów od dawna istnieje recepta na czasochłonne przekształcenia lub zwiększone zapotrzebowanie na moc obliczeniową. Rozwiązaniem tym mógłby

być klaster komputerowy połączony siecią bezprzewodową z humanoidem. Niestety klastrów nie można zastosować z dwóch powodów. Po pierwsze niektóre przekształcenia muszą być wykonywane w ścisłej kolejności (źle się je zrównolegla), a po drugie strumień video musi być przekazywany pomiędzy węzłami klastra przy użyciu stosunkowo wolnego medium transmisyjnego.

3 System wizyjny oparty na modelu biologicznym

Od niedawna w literaturze zaczęły się pojawiać wzmianki o sieciach HTM. Zbudowane w wyniku współpracy neurofizjologów i informatyków, dzięki doświadczeniom z sieciami neuronowymi (pamięcią autoasocjacyjną). Sieć HTM stanowią połączone ze sobą węzły (złożone z pamięci autoasocjacyjnej i odpowiedniego algorytmu) połączone według zdefiniowanego wcześniej sposobu (hierarchicznie). Na najniższym poziomie dane sensoryczne (w przypadku systemu wizyjnego – piksele obrazu) trafiają do sieci, gdzie na każdej kolejnej warstwie budowana jest nowa, abstrakcyjna reprezentacja, aż do warstwy najwyższej – kojarzeniowej.

W skrócie „HTM" środkowa litera pochodzi od angielskiego słowa „temporal", które oznacza, że dla poprawnego działania sieci HTM, dane wejściowe muszą być zmienne w czasie. Ten warunek ma swoje uzasadnienie w algorytmach, ale nie tylko. Naukowcy pracujący nad sieciami HTM wzorowali się na zasadzie funkcjonowania ludzkiego mózgu – a tutaj, jak nigdzie indziej – zmienność wzorców w czasie jest oczywista. Słysząc jedną częstotliwość nie rozpoznamy melodii, wpatrując się w jedno miejsce nie widzimy całego obiektu, nawet zmysł dotyku dostarcza zbyt mało informacji jeżeli nie możemy przesuwać palców po obiekcie.

3.1 System wizyjny oparty na niekompletnym strumieniu wideo

Lektura książki „On Intelligence" [2] wywołała pewne przemyślenia związane z naturą zjawiska widzenia u ludzi w kontekście zastosowania w humanoidach. Skoro zmysł wzroku działa w oparciu o sakkady – błyskawiczne przemieszczenia wzroku pomiędzy rozmaitymi szczegółami obrazu – oznacza to, że informacje docierają do mózgu w postaci występujących po sobie w czasie fragmentów obrazu. Siatkówka nie działa tak jak kamera. Pokryta jest nierównomiernie pręcikami i czopkami, które najbardziej zagęszczone są w miejscu tzw. plamki żółtej, dając właśnie tam najwyższą rozdzielczość obrazu. Może więc niepotrzebnie dzisiejsze humanoidy mają kamery o „stałej rozdzielczości", może – wzorem natury – warto przemyśleć stworzenie programowego odpowiednika plamki żółtej.

Gdyby okazało się, że taka redukcja ilości danych przesyłanych przez system wizyjny jest wystarczająca, mógłby być zastosowany klaster. Otworzyłoby to kolejne ogromne możliwości w dziedzinie sztucznej inteligencji humanoidów.

Struktura danych systemu wizyjnego, o której mowa, może być zrealizowana na kilka sposobów. Wszystkie te metody mają jednak jedną cechę wspólną – wymagają zaprojektowania mechanizmu, który określi współrzędne „plamki żółtej" (określające punkt fiksacji „wzroku" humanoida). Jeżeli taki mechanizm zostanie stworzony, wówczas można przystąpić do projektowania struktury danych systemu wizyjnego. Najprostsze rozwiązanie, przedstawione na rys. 2, polega na rezygnacji z części

danych poprzez pomijanie niektórych pikseli. Do pewnej odległości od „plamki żółtej" system przesyła wszystkie piksele, dalej – w pewnym zakresie – co drugi piksel i ostatecznie co trzeci piksel. Rys. 2 został spreparowany, by pokazać, które piksele są pomijane – w rzeczywistości węzeł odbierający taką niekompletną klatkę obrazu, piksele oznaczone kolorem białym (pominięte) zastąpi najbliższym znanym pikselem.

Rys. 2 przedstawia dwie koncepcje – z wykorzystaniem obszaru prostokątnego jako definiującego dany poziom dokładności oraz z wykorzystaniem odległości od punktu. Co prawda biologicznie bardziej uzasadniony jest ten drugi przypadek, jednak matematycznie podejścia te są równoprawne. Natomiast technicznie najszybszym wydaje się być pierwszy. Sprawdzanie przynależności danego piksela do jednego z podzbiorów jest tu zwykłym sprawdzeniem współrzędnych, podczas gdy bazując na odległości od punktu, trzeba dla każdego piksela osobno obliczać odległość.

Skuteczność obu rozwiązań w zakresie kształtu pola widzenia dla każdego z poziomów precyzji jako zagadnienie dyskusyjne nie była porównywana.

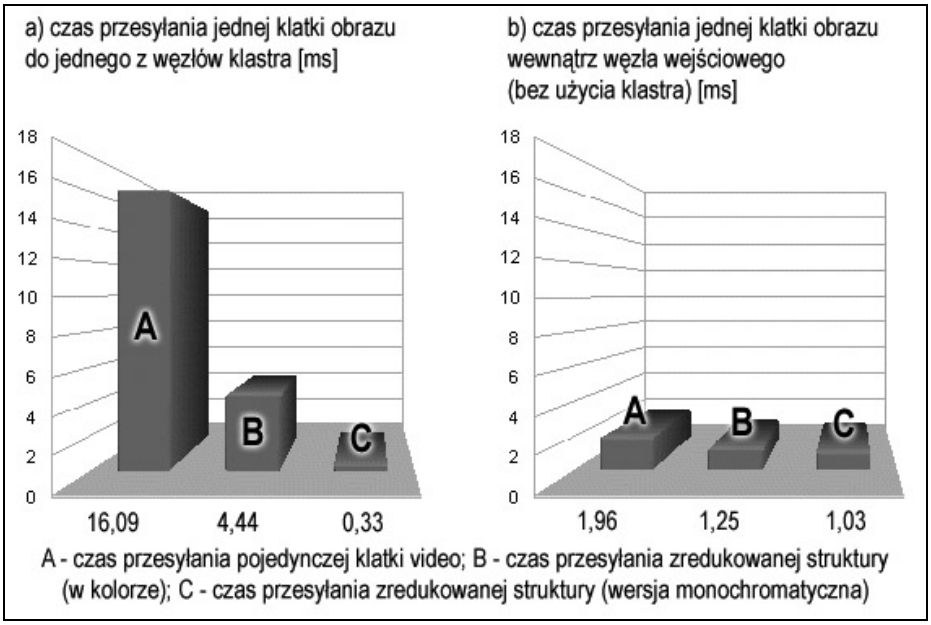
Rys. 2. Dwa przykładowe tryby redukcji danych strumienia video. Białym kolorem zamalowano piksele pomijane w transmisji. Rysunek został spreparowany w programie graficznym, aby jak najlepiej uwidocznić ideę utraty informacji.

Oczywiście istnieją inne metody redukcji danych. Można na przykład zaimplementować kompresję stratną o zmiennym (w dwuwymiarowej przestrzeni) współczynniku kompresji lub też zaproponować nowy format graficzny o „zmiennej wielkości piksela", gdzie przykładowo 24 bity oznaczałyby nie kolor jednego punktu, ale określonej (zależnej od odległości od „plamki żółtej") liczby punktów. Pomysłów jest wiele. Nie to jest jednak teraz najważniejsze, jak dokładnie zredukować ilość danych, lecz jakie taka redukcja przynosi korzyści.

4 System wizyjny a klaster komputerowy

Na rys. 3 przedstawione zostały wstępne wyniki badań czasu komunikacji – czasu przesyłania danych zawierających informacje z pojedynczej klatki obrazu. Pomiary zostały przeprowadzone na komputerze z procesorem 1,84GHz i 640MB pamięci operacyjnej oraz z kartą sieciową działającą w technologii 100Base-TX. Wyniki są

średnią arytmetyczną z 1000 pomiarów dla każdej kategorii. Parametry sprzętu, czy też zmierzone wartości nie są tu najbardziej istotne (wyniki pomiarów zależą od wielu czynników, w tym systemu operacyjnego, uruchomionych procesów, ruchu w sieci itp.), to raczej stosunek poszczególnych wartości (czy czas transmisji uległ wydłużeniu czy skróceniu w porównaniu z klasycznym rozwiązaniem i o ile).



Rys. 3. Wyniki pomiarów czasu transmisji pojedynczej klatki obrazu: a) do drugiego komputera (węzła klastra), b) wewnątrz węzła wejściowego (komputera z kamerą). Mierzone są czasy dla trzech przypadków: A- „kompletna" klatka obrazu wideo, B- klatka obrazu „zredukowana" według pomysłu opisanego w rozdziale 3, C- również klatka „zredukowana", ale w wersji monochromatycznej.

Wykres „b)", po prawej stronie rys. 3, pokazuje czasy transmisji, gdy aplikacja nie korzysta z klastra, natomiast wykres po lewej stronie – „a)" – gdy klaster jest w użyciu. Widać bardzo wyraźnie (kolumny „A" obu wykresów), że tradycyjne podejście do obrazu pochodzącego z systemu wizyjnego – czyli przesyłanie pojedynczych klatek w postaci „kompletnej" – całkowicie dyskwalifikuje użycie klastra. Czas transmisji „w obrębie jednego komputera" wynosi 1,96ms, natomiast przy użyciu zewnętrznego węzła do celów przetwarzania czas ten wzrósł do 16,09ms. Co prawda całość obliczeń przerzucana jest na ten zewnętrzny węzeł, ale sam czas transmisji jest tak ogromny, że aby otrzymać płynny ruch w obrazie systemu wizyjnego (28 klatek na sekundę), połowa z każdej sekundy czasu procesora musiałaby być poświęcona na transmisję obrazu. Taka sytuacja jest nie do przyjęcia w przypadku humanoida, którego procesor musi obsługiwać znacznie więcej zadań niż tylko wizję.

Kolumny „B" pokazują natomiast, że w przypadku obrazu, na którym przeprowadzono redukcję ilości danych opisaną w rozdziale 3, różnica tych czasów

nie jest aż taka duża. Można śmiało powiedzieć, że czas transmisji pełnej klatki wideo w obrębie jednego węzła (wykres „b)", kolumna „A", 1,96ms) jest zbliżony wartością do czasu transmisji zredukowanej klatki do klastra (wykres „a)", kolumna „B", 4,44ms). Taka różnica jest już akceptowalna i daje techniczną możliwość przerzucenia części operacji na inny komputer.

Warte uwagi są również ostatnie kolumny wykresów – „C" – pokazano tu czas transmisji obrazów monochromatycznych. Nie zawsze bowiem istotna jest informacja o kolorze. Jeżeli założymy wykorzystanie klastra komputerowego jako przetwarzającego dane systemu wizyjnego z rozdziałem funkcjonalnym zadań, to wówczas niektóre z zadań mogą wymagać uproszczonej postaci danych wejściowych. Dla przykładu jeden z węzłów klastra może analizować ruch obserwowanych obiektów – w tym przypadku wystarczy, że otrzyma różnicę dwóch kolejnych klatek obrazu (i to niekoniecznie w postaci kolorowej). Nawet analiza kolorów obiektów może przebiegać w oparciu o niepełne informacje dotyczące koloru – można na przykład na podstawie poziomów jaskrawości kolorów składowych zbudować obraz monochromatyczny zawierający informację o poziomie jaskrawości obserwowanych obiektów. Kolumna „C" na rys. 3 „b)" dowodzi temu, że w niektórych sytuacjach przesyłanie klatki do innego węzła może odciążyć węzeł wejściowy (ten z kamerą). Jest to sytuacja jak najbardziej pożądana, ponieważ przekształcenia dokonywane na obrazie, podobnie jak algorytmy wnioskowania, nie będą wówczas wykonywane na węźle wejściowym. Należy pamiętać, że humanoid zazwyczaj „nosi ze sobą" tylko jeden komputer, więc jego moc obliczeniowa jest niezwykle cenna. Okazuje się, że w przypadku przesyłania spreparowanych obrazów monochromatycznych, przesyłanie klatek poza węzeł wejściowy może trwać nawet krócej, niż przesyłanie do odpowiedniej funkcji algorytmu wewnątrz jednego programu.

5 Podsumowanie i dalsze prace

W wyniku przeprowadzonych badań udało się udowodnić, że wykorzystanie klastra komputerowego do zadań stawianych przez systemy wizyjne jest wykonalne. Komplikacje związane krytycznym parametrem funkcjonowania algorytmów aplikacji równoległych – czasem komunikacji – mogą zostać rozwiązane poprzez zastosowanie omówionej w rozdziale trzecim redukcji ilości danych. Czas procesora węzła wejściowego zyskany poprzez przerzucenie masywnych obliczeń systemu wizyjnego na inne węzły klastra można wykorzystać do umiejętnego preparowania obrazów źródłowych rozsyłanych do przynajmniej kilku węzłów klastra. Jeżeli zaistnieje taka sytuacja, w której użycie klastra nie spowoduje dodatkowego opóźnienia ani dodatkowego obciążenia procesora węzła wejściowego, będzie to ogromny zysk dla projektantów danego humanoida. Sam fakt „wyprowadzenia" części obliczeń do klastra, stworzy wówczas możliwość wykorzystania dodatkowej mocy, jaką dysponuje klaster, na przykład, aby zwiększyć poziom „inteligencji" humanoida.

Nie bez powodu na początku rozdziału trzeciego wspomniany został pomysł sieci HTM. Jest wysoce możliwe, że właśnie zestawienie omówionego pomysłu implementacji „plamki żółtej" i idei sieci HTM da interesujące efekty.

Nie uwzględniając faktu, że ludzkie oko w pewnych odstępach czasowych przesyła fragmenty obrazu, które dopiero w wyższych warstwach składają się na

oglądany obraz, stosowanie opisanej redukcji danych jest bezcelowe. To właśnie sieci HTM stanowią o dalszym rozwoju tego pomysłu.

Rys. 4. Praktyczna implementacja w projektowanym systemie wizyjnym. Na rysunku przedstawiony obraz w postaci przygotowanej do wysyłania (z usuniętymi pikselami) do węzłów klastra. Krzyżyk oznacza współrzędne „plamki żółtej" ustalonej w wyniku głosowania poszczególnych węzłów klastra zajmujących się rozpoznawaniem ruchu, koloru i innych atrybutów rejestrowanego obrazu.

Odpowiedź na pytanie postawione w temacie jest twierdząca. Umiejętna redukcja danych obrazu wejściowego może umożliwić użycie klastra komputerowego, co znacząco wpłynie na możliwości w zakresie implementacji nie tylko rozpoznawania, ale również pamiętania, a może i rozumienia widzianego przez humanoida świata ludzi.

Literatura

[1] Hawkins J. (2007), Learn Like a Human, *IEEE Spectrum*, **4**,
[2] Hawkins J., Blakeslee S. (2004), *On Intelligence*, Times Books,
[3] Podpora M. (2007), Computer vision in parallel computing, *Przegląd Elektrotechniczny*, **11**, 68-70,
[4] Podpora M. (2008), HTM – nowa era w historii sztucznych sieci neuronowych?, *Zeszyty Naukowe –Informatyka*, Oficyna Wydawnicza Politechniki Opolskiej, Opole 2008.