

Jan KLIMASZEWSKI, Barbara PUTZ

Instytut Automatyki i Robotyki Politechniki Warszawskiej, Warszawa

BŁOKOWY ALGORYTM DOPASOWANIA OBRAZÓW MULTIMODALNYCH W SYSTEMACH MONITOROWANIA

Słowa kluczowe

Dopasowanie obrazów, fuzja obrazów, głowica obserwacyjna, korelacja fazowa, FFT.

Streszczenie

Dopasowanie obrazów jest jednym z etapów fuzji obrazów. W artykule zaprezentowano blokowy algorytm dopasowania obrazów multimodalnych, bazujący na korelacji fazowej. Algorytm wykorzystuje podział obrazu na prostokątne bloki w celu lepszego dopasowania kilku planów obserwowanej sceny. Rozwiązanie zostało opracowane dla systemu monitorowania złożonego z kamer IR oraz TV, przy wynikających stąd założeniach upraszczających.

Wprowadzenie

Fuzja obrazów w systemach monitorowania jest zadaniem prowadzącym do połączenia ze sobą obrazów pochodzących z różnych sensorów (tzw. obrazów multimodalnych) celem uzyskania jednego obrazu zawierającego najbardziej pożądane informacje występujące na obrazach składowych. Dzięki temu otrzymany obraz dostarcza więcej informacji o obserwowanej scenie niż każdy z obrazów składowych osobno. Fuzja obrazów jest stosowana w monitoringu

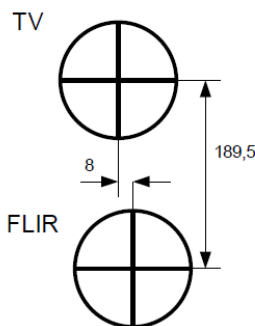
wizyjnym i systemach nadzoru, w systemach diagnostyki stanu maszyn czy budynków, w diagnostyce medycznej.

Fuzja obrazów może zostać przeprowadzona jedynie w przypadku, gdy obrazy składowe są przestrzennie do siebie dopasowane, tzn. odpowiadają sobie położeniem, rotacją, kształtem i skalą. Dopasowanie obrazów jest jednym z istotnych zadań koniecznych do zrealizowania w celu dokonania fuzji.

Artykuł prezentuje algorytm rozwiązania zagadnienia dopasowania z wykorzystaniem korelacji fazowej i podziału obrazów na bloki, na przykładzie konkretnego systemu monitorowania i wynikających stąd założeń upraszczających, co opisano w rozdziale 1. Rozdział 2 zawiera krótki przegląd literatury dotyczącej zadania dopasowania obrazów. W rozdziale 3 opisana została opracowana metoda dopasowania. Rozdział 4 przedstawia wyniki testowania opracowanego algorytmu.

1. Opis problemu

Głowica obserwacyjna, z której pochodzą obrazy składowe przeznaczone do fuzji, składa się z dwóch kamer (rys. 1). Jedna kamera pracuje w paśmie podczerwonym, a druga w paśmie światła widzialnego. Rozdzielczość kamery IR wynosi 384x256, a kamery TV 756x581 pikseli.



Rys. 1. Rozmieszczenie kamer IR (FLIR) i TV w głowicy obserwacyjnej; wymiary na rysunku podane w milimetrach

Zadanie dopasowania obrazów pochodzących z takiej głowicy musi rozwiązywać wiele problemów. Pierwszym problemem jest znalezienie metody porównania ze sobą obrazów IR i TV tak, aby możliwe było automatyczne określenie, czy dwa obrazy są do siebie dopasowane poprawnie. Z uwagi na przestrzenne rozmieszczenie kamer (rys. 1), na obrazach pochodzących z systemu, podczas dopasowywania do siebie obiektów widocznych na obu obrazach bę-

dzie występował efekt stereowizyjny [7]. Przesunięcie pionowe obiektów na obrazach będzie zależne od odległości obiektu od kamer (w standardowych zadaniach związanych ze stereowizją przesunięcie występuje w poziomie).

W pracy przyjęto dodatkowo założenie, że optymalne dopasowanie obrazów poszukiwane jest dla stałego współczynnika rotacji, skalowania i zmiennej współczynniki translacji. Oznacza to przyjęcie założenia o pomijalnej nierównoległości względem siebie osi optycznych kamer (skomplikowana rektyfikacja obrazu nie jest konieczna). Głównym poszukiwanym parametrem dopasowania jest przesunięcie obrazów w osi pionowej Y obrazu. Dodatkowo zostało wprowadzone przeszukiwanie w niewielkim zakresie dopasowania w osi poziomej X w celu zniwelowania efektów drgań głowicy i braku synchronizacji czasowej par obrazów.

W celu znalezienia poprawnego dopasowania konieczne jest opracowanie takiej metody, która będzie skuteczna dla obrazów zawierających kilka różnych, bliższych i dalszych planów, czyli uwzględni różne wartości dopasowania dla fragmentów obrazu zawierających obiekty znajdujące się w różnej odległości od obserwatora. Dodatkowym wymaganiem, bardzo istotnym w przypadku systemu monitorowania, jest możliwość realizacji algorytmu w czasie rzeczywistym.

2. Przegląd rozwiązań

Algorytmy dopasowania obrazów najczęściej dzielone są na dwie grupy [9]: algorytmy obszarowe, analizujące obraz na poziomie pikseli (zwane często algorytmami bazującymi na intensywności [8]) oraz algorytmy bazujące na detekcji cech charakterystycznych w obrazie.

Algorytmy obszarowe wykorzystują metody, w których obrazy lub ich fragmenty są dopasowywane do siebie w wyniku maksymalizacji różnorodnych miar podobieństwa różnymi metodami optymalizacji. Tradycyjne miary wymagają wyznaczania sumy kwadratów błędów (SSD, MSE) bądź sumy błędów bezwzględnych (SAD, MAE) [4, 7]. Stosowane są też lokalne miary podobieństwa, z użyciem przesuwne okna [4]. Specyficzny dla algorytmów dopasowania rodzaj miary stanowią współczynniki korelacji wzajemnej [3, 4] oraz miara informacji wzajemnej (MI) [4, 9], uważana za szczególnie korzystną dla obrazów multimodalnych. Do algorytmów obszarowych o największej efektywności należą algorytmy poszukujące maksymalnej miary dopasowania w dziedzinie częstotliwości obrazu, np. algorytm korelacji fazowej. Bazują one na przekształceniach całkowitych Fouriera i Mellina i wydajnym algorytmie FFT, [6, 9], toteż pozwalają na implementację w czasie rzeczywistym [1].

Druga grupa algorytmów bazuje na wyznaczaniu w dopasowywanych obrazach cech charakterystycznych (takich jak wydzielone segmenty obrazu, krawędzie, narożniki, punkty przecięcia linii) oraz tzw. punktów kontrolnych reprezentujących te cechy – są to np. środki ciężkości wyznaczonych segmentów.

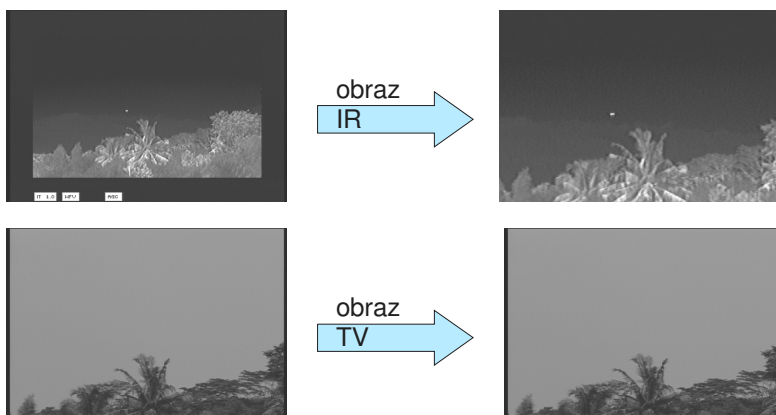
Stosowane w tej grupie metody bazują na wykrywaniu przestrzennych zależności pomiędzy punktami kontrolnymi, dopasowaniu punktów kontrolnych pod kątem ich własności, podejściu relaksacyjnym z iteracyjnym powtarzaniem aż do momentu uzyskania stabilnego dopasowania, reprezentacji piramidowej czy zastosowaniu falek [9]. Ze względu na konieczność ekstrakcji cech charakterystycznych mają one wciąż zbyt dużą złożoność obliczeniową wobec wymagania ich realizacji w czasie rzeczywistym.

3. Algorytm blokowego dopasowania obrazów

Artykuł przedstawia metodę dopasowania do siebie synchronicznych czasowo par obrazów pochodzących z kamery wizyjnej i kamery termowizyjnej. Metoda dopasowania bazuje na klasycznym algorytmie korelacji fazowej [6] i wykorzystuje podział obrazu na bloki. Dla każdej pary obrazów wykonywane są cztery etapy algorytmu: dopasowanie wstępne, zgrubne, dokładne i integracja danych. Każdy z etapów został szczegółowo opisany w kolejnych podrozdziałach.

3.1. Wstępne dopasowanie obrazów – etap I

Pary obrazów pochodzą z kamer o różnych parametrach optycznych. Kamery posiadają różne kąty widzenia oraz dostarczają obrazy o różnych rozdzielczościach. Pierwszy etap dopasowania niweluje różnice w rozdzielczości i kącie widzenia głównie za pomocą skalowania i interpolacji.



Rys. 2. Wstępne dopasowanie obrazu IR do TV za pomocą skalowania, interpolacji i przycinania obrazu IR

Obrazy IR i TV przed i po operacjach wstępnego przygotowania ilustruje rys. 2. Fizyczne rozmieszczenie kamer jest znane (rys. 1) i niezmiennie w czasie. Ponadto znane są wartości kątów widzenia oraz rozdzielczość kamer. Z uwagi na to przyjęte zostały stałe współczynniki skalowania obrazu IR względem TV. Jako obraz referencyjny przyjęty został obraz TV z uwagi na jego większą rozdzielczość. Dzięki takiemu założeniu nie następuje utrata informacji o scenie w wyniku skalowania obrazu IR. Fragmenty obrazu IR przedstawiające elementy sceny niewidoczne na obrazie TV nie są brane pod uwagę na dalszych etapach analizy. Dodatkowo przyjęte zostało założenie o wzajemnej równoległości osi optycznych kamer IR i TV. Eksperymenty potwierdziły, że przyjęcie takiego założenia jest zasadne z uwagi na znikome zniekształcenia wynikające z transformacji perspektywicznej obrazów względem siebie. W efekcie tego dopasowanie obrazów może ograniczyć się do operacji rotacji, skalowania i translacji, z pominięciem przekształceń rzutowych. W znacznym stopniu zmniejsza to koszt obliczeniowy algorytmu.

3.2. Wyznaczenie zgrubnego dopasowania – etap II

Na podstawie obrazów IR i TV, dopasowanych wstępnie do siebie pod względem rozdzielczości, skali i rotacji, następuje wyznaczenie dopasowania zgrubnego jako wartości translacji w kierunkach x i y metodą korelacji fazowej [1, 6]. Funkcję korelacji wzajemnej wyznacza się jako odwrotną transformatę Fouriera względem funkcji R będącej wzajemną gęstością widmową:

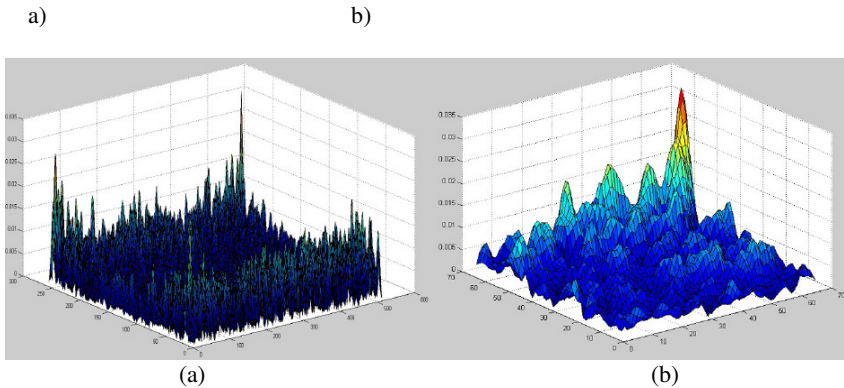
$$R = \frac{F_{IR} \cdot F_{TV}'}{|F_{IR} \cdot F_{TV}'|} \quad (1)$$

We wzorze tym występują transformaty Fouriera (F_{IR} , F_{TV}) obrazów IR i TV i funkcje względem nich sprzężone. Ważnym krokiem w metodzie korelacji fazowej jest filtracja otrzymanej funkcji R przed wyznaczeniem transformaty odwrotnej w celu uzyskania stabilnych i jednoznacznych wyników. Wartość zgrubnego dopasowania obrazów IR i TV, czyli wartości (x_{\max}, y_{\max}) przesunięcia ich względem siebie w osi x i osi y uzyskuje się poprzez wyznaczenie współrzędnych maksimum modułu funkcji korelacji wzajemnej:

$$x_{\max}, y_{\max} = \arg \max_{x,y} F^{-1}(R) \quad (2)$$

Przykładowe wartości przeszukiwanej macierzy przedstawia rys. 3.

Do wyznaczonych współczynników dopasowania dla obrazów w sekwencji wideo stosowany jest filtr uśredniający (np. medianowy) o szerokości 16 ramek obrazu. Dzięki temu uzyskuje się płynność tych parametrów w sekwencji wideo.



Rys. 3. Wykres modułu odwrotnej transformaty Fouriera: a) w całym zakresie $x = 0-511$, $y = 0-255$; b) w okolicy maksimum $x = 2$, $y = 8$

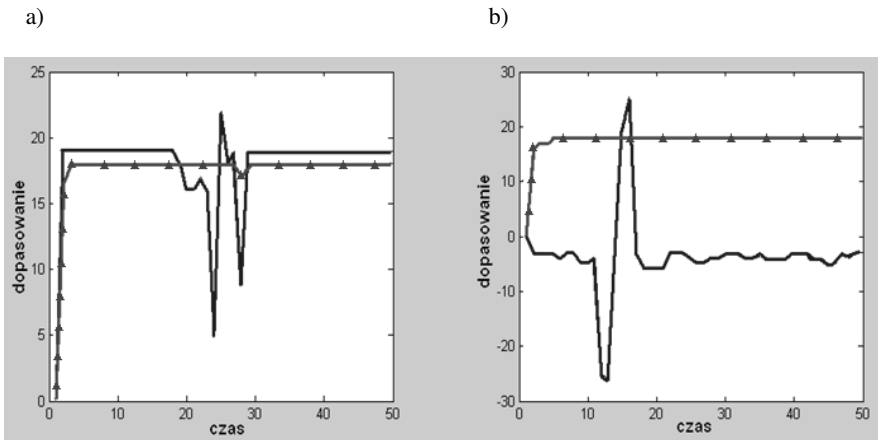
3.3. Wyznaczenie dopasowania dokładnego – etap III

Ten etap algorytmu rozpoczyna się od podziału obrazu IR na jednakowe prostokątne bloki w liczbie od 4 do 64. Podział na bloki następuje w taki sposób, aby ich rozmiary pozwalały na łatwe zastosowanie transformacji Fouriera, mając na uwadze wymagania implementacji sprzętowej. Dla każdego z nich zostają wyznaczone wartości dopasowania względem odpowiadającego mu fragmentu obrazu TV (opisane na etapie dopasowania zgrubnego). Z uwagi na bardzo niewielkie rozsuniecie w poziomie osi optycznych kamer, dopasowanie poziome dla każdego bloku przyjmuje się jako stałe i równe dopasowaniu zgrubnemu w osi X wyznaczonemu na etapie II. Na tym etapie algorytmu pojawiają się bloki obrazów, dla których wyznaczone współczynniki dopasowania są nieprawidłowe, albowiem niektóre fragmenty obrazu przedstawiają obiekty, które nie są wizualizowane w jednakowy sposób w paśmie termowizyjnym i wizyjnym (na którymś z obrazów mogą być np. całkiem niewidoczne). Dodatkowo w wyniku zaszumienia obrazów metoda wyznaczenia dopasowania zwraca czasem niepoprawne wyniki, nie znajdując właściwego maksimum. Przykładowe wartości dopasowania dla dwóch różnych bloków obrazu w przestrzeni czasu (dla sekwencji wideo IR i TV) przedstawiają wykresy na rys. 4 (zaznaczone linią ciągłą).

Z uwagi na występujące błędy w wyznaczeniu dopasowania konieczne stało się wykrywanie niepoprawnie dopasowanych bloków w celu uniknięcia propagacji błędów do ostatniego etapu algorytmu.

Opracowana została metoda wykrywania i korekcji niepoprawnych wartości dopasowania bazująca na statystycznych cechach zmienności wartości dopasowania dla bloków w przestrzeni obrazu oraz w przestrzeni czasu. Wyniki działa-

nia metody przedstawiono na rys. 4 (zaznaczone linią z trójkątami). Metoda niweluje chwilowe błędy w dopasowaniu, zachowując płynnie zmienną wartość dopasowania (rys. 4a). Może się jednak zdarzyć, że odrzuca stabilne w czasie wartości dopasowania, jeśli są położone zbyt daleko od wartości dopasowania zgrubnego. Uzyskane po korekcji wartości (rys. 4b) okazują się być lepszym rozwiązaniem problemu dopasowania niż te, które wynikałyby z uśrednienia dopasowania dokładnego.



Rys. 4. Przykładowy efekt wykrycia i korekcji niepoprawnego dopasowania w pionie wybranego bloku dla dwóch różnych przypadków; linia ciągła – wyznaczone pierwotnie dopasowanie dokładne, linia z trójkątami – dopasowanie po korekcji

Wykrycie niepoprawnego dopasowania następuje poprzez porównanie wyznaczonej wartości dopasowania dla danego bloku z wartością dopasowania zgrubnego wyznaczoną dla całego obrazu na etapie II. Blok zostaje uznany za poprawnie dopasowany, jeśli odległość wartości dopasowania dla tego bloku od wartości dopasowania zgrubnego nie przekracza pewnej wartości progowej ustalonej dla każdego obrazu i bieżącej chwili czasowej. Zależność tę przedstawia wzór (3):

$$|y_z - y_d(j)| < a \cdot \frac{\sum_{t=1}^{t=T} \sum_{i=1}^{i=n} |y_z - y_d(i,t)|}{nT} \quad (3)$$

gdzie:

y_z – zgrubna wartość dopasowania,

$y_d(j)$ – sprawdzana wartość dopasowania dokładnego dla j -tego bloku,

$y_d(i,t)$ – wartość dopasowania dokładnego dla bloku i -tego w obrazie pochodzącym z konkretnej ramki obrazu t liczonej wstecz od ramki bieżącej,

n – liczba bloków obrazu,

T – liczba ramek wstecz, które brane są pod uwagę podczas analizy,

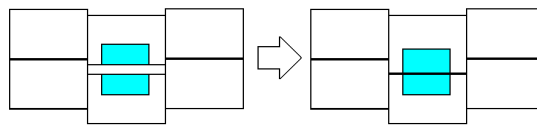
a – współczynnik dobrany eksperymentalnie.

Końcowa korekcja niepoprawnego dopasowania bloku polega na wyznaczeniu poprawnej wartości dopasowania na podstawie wartości dopasowania dokładnego bloków sąsiadujących oraz wartości dopasowania zgrubnego dla całego obrazu.

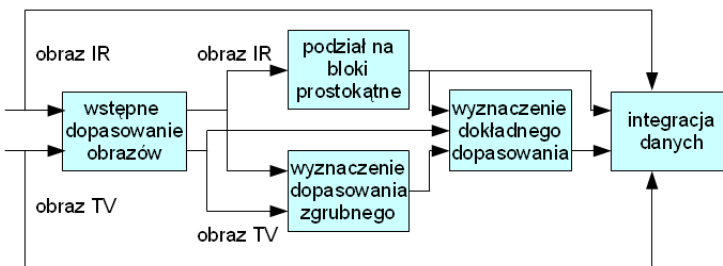
3.4. Integracja danych – etap IV

Ostatni etap obejmuje integrację danych pochodzących z poprzednich etapów działania algorytmu. Dla prostokątnych bloków IR wraz z wartościami ich dopasowania do obrazu referencyjnego TV konieczna jest taka modyfikacja bloków, aby w obrazie po fuzji wypełnić przerwy pomiędzy blokami powstałe w wyniku wyznaczonych różnych wartości dopasowania ich w pionie. Realizuje się to przez skalowanie i interpolację bloków umieszczonych powyżej przerwy (rys. 5). Po wykonaniu tych operacji obrazy IR i TV są gotowe do przeprowadzenia właściwej fuzji obrazów.

Diagram przepływu danych dla wszystkich czterech etapów algorytmu ilustruje rys. 6.



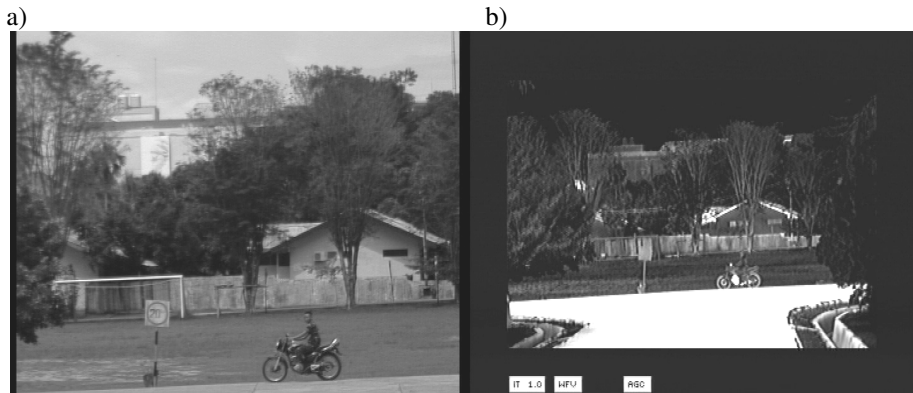
Rys. 5. Ilustracja zasady działania algorytmu integracji danych



Rys. 6. Przepływ danych na kolejnych etapach algorytmu

4. Testy i przykładowe wyniki

Na rys. 7–9 widoczna jest scena przedstawiająca motocyklistę, drzewa oraz zabudowania. Na obrazach można wyróżnić kilka planów. Najbliżej znajduje się motocyklista, dalej dom i drzewa, a najdalej zabudowania. Rysunek 7 przedstawia przykładowe obrazy pochodzące z głowicy kamer TV i IR.

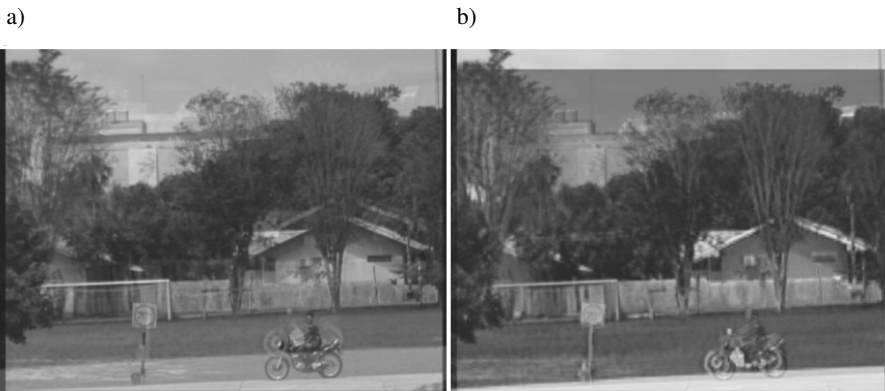


Rys. 7. Obrazy po akwizycji: a) obraz TV; b) obraz IR

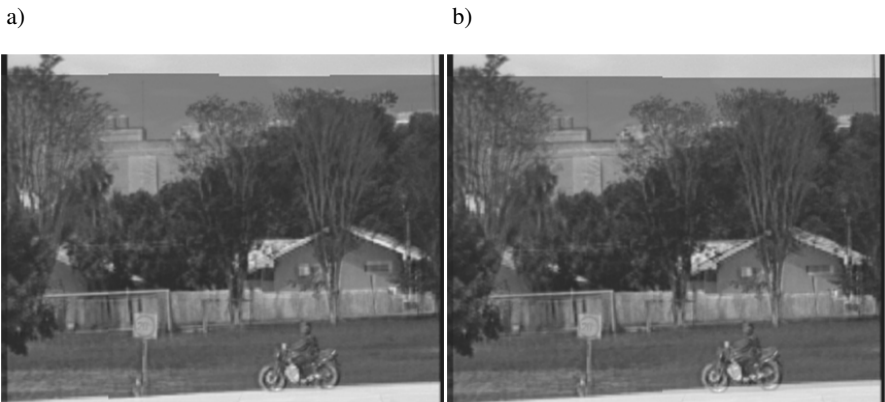
W celu uwidocznienia różnic w translacji pomiędzy obrazami IR i TV do fuzji obrazów zastosowano metodę średniej ważonej. Rysunek 8a przedstawia obrazy po etapie wstępnego dopasowania, czyli bez zastosowania parametrów translacji. Rysunek 8b przedstawia wyniki dopasowania zgrubnego, czyli z zastosowaniem jednej wartości translacji dla pary obrazów. Dla motocyklisty można zauważyć lekkie niedopasowanie w porównaniu z obiektami położonymi dalej (drzewa i zabudowania). W celu poprawienia dopasowania konieczne jest zastosowanie opisanej blokowej metody dopasowania. Rysunek 9a przedstawia wyniki dopasowania metodą z podziałem na 16 bloków. Porównując wyniki zgrubnego dopasowania z wynikami dopasowania zrealizowanego opracowanym algorytmem, można zauważyć jednocześnie poprawne dopasowanie obiektów znajdujących się na planie bliskim (motocyklista) oraz obiektów położonych daleko (budynki) od obserwatora. Zachowane zostaje poprawne dopasowanie obiektów na środkowym planie (dwuspadzisty dach).

Podobne rezultaty uzyskano, stosując metodę blokowego dopasowania z wykorzystaniem detektora krawędzi Canny'ego [5] i maksymalizacji miary podobieństwa SAD. Mimo że uzyskano wynik nieco gorzej dopasowany (rys. 9b) niż przy zastosowaniu korelacji fazowej, w obu przypadkach cały obraz i wszystkie plany zostały lepiej dopasowane z wykorzystaniem blokowej realizacji niż przy dopasowywaniu obrazów w całości.

Wykorzystanie detektora krawędzi Canny'ego było testowane w fazie opracowywania końcowego algorytmu blokowego dopasowania. Jako przeznaczona do sprzętowej implementacji została jednak wybrana blokowa metoda korelacji fazowej z uwagi na niższą złożoność obliczeniową. Pozwoli to na poszerzenie rozwiązań opisanych w [1]. Tym niemniej metoda podana w [5], łącząca algorytm korelacji fazowej z detekcją krawędzi Canny'ego, warta jest dalszych badań odnośnie do możliwości jej realizacji w czasie rzeczywistym.



Rys. 8. Obrazy nałożone na siebie metodą średniej ważonej na potrzeby wizualizacji: a) po dopasowaniu wstępnym; b) po dopasowaniu zgrubnym



Rys. 9. Obrazy nałożone na siebie metodą średniej ważonej na potrzeby wizualizacji: a) po dopasowaniu dokładnym (metodą blokową z korelacją fazową); b) po dopasowaniu metodą blokową z wykorzystaniem przeszukiwania obszarowego i detektora Canny'ego

Podsumowanie

Metoda prezentowana w niniejszym artykule opiera się na klasycznym algorytmie korelacji fazowej i prostym, ale skutecznym dopasowywaniu obrazów

blokami. Różne inne rozwiązania z wykorzystaniem metod blokowych można spotkać w literaturze. W [8] opisana została metoda optymalnego podziału obrazu na prostokątne bloki. Do wyznaczenia dopasowania wykorzystano metodę informacji wzajemnej, a do wyboru fragmentów obrazu do podziału zastosowano metodę opartą o wyznaczanie entropii. Dla obrazów z systemu monitorowania opisanego w rozdziale 3 nie było jednak możliwe zaproponowanie podobnego rozwiązania. Ze względu na niską jakość obrazu IR i inne parametry obrazowania, metody bazujące na MI i entropii nie dawały pożądaných rezultatów.

W [2] zastosowano ciekawą metodę wykrywania dopasowania dla różnych bloków obrazu, w której autorzy podchodzą do podziału obrazu na bloki w sposób podobny do zaprezentowanego w niniejszym artykule. Jednakże z powodu dużego zaszumienia obrazów IR i TV oraz z powodu różnic wynikających z modalności, wykorzystanie techniki szacowania jakości dopasowania na podstawie wartości maksimum korelacji fazowej nie znalazło zastosowania w przypadku analizowanych sekwencji wideo pochodzących z głowicy pomiarowej. Są to jednak metody, które mogą pozwolić na wyznaczenie nowych dróg rozwoju algorytmu opisanego w artykule.

Praca wykonana częściowo w ramach projektu rozwojowego MNiSW nr O R00 0019 07.

Bibliografia

1. Antoniewicz A., Kondej M., Putz B.: Implementacja algorytmu dopasowania obrazów IR i TV w układzie FPGA. Rozdz. 27 w: Projektowanie, analiza i implementacja systemów czasu rzeczywistego, pod red. L. Trybusa i S. Samoleja, WKiŁ 2011, 329–340 (w druku).
2. Argyriou V., Vlachos T.: Motion Estimation Using Quad-Tree Phase Correlation, IEEE Conf. on Image Proc. 2005, 1–4 (I–1081).
3. Heather, J.P., Smith, M.I.: Multimodal Image Registration with Applications to Image Fusion. 7th International Conference on Information Fusion, 2005, vol. 1, 1–8.
4. Mitchell H.B.: Image Fusion. Theories, Techniques and Applications. Springer, 2010.
5. Orchard J.: Globally optimal multimodal rigid registration: an analytic solution using edge information. IEEE Conf. on Image Proc., 2007, 485–488.
6. Reddy B.S., Chatterji B.N.: An FFT-based technique for translation, rotation and scale-invariant image registration, IEEE Transactions on Image Processing, 1996, vol. 5, 1266–1271.
7. Scharstein D., Szeliski R.: A Taxonomy and evaluation of dense Two-Frame Stereo Correspondence Algorithms. SMBV '01 Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision, 2001.

8. Zhaoying L., Fugen Z., Xiangzhi B., Hui W., Dongjie T.: Multi-modal image registration by mutual information based on optimal region selection. *Information Networking and Automation (ICINA)*, 2010, 249–253.
9. Zitová B., Flusser J., Sroubek F.: *Image Registration: A Survey and Recent Advances*. IEEE Conf. on Image Proc. 2005 Tutorial, 1–55.

Recenzenci:

Liliana BYCZKOWSKA-LIPIŃSKA

Andrzej KASIŃSKI

Multi-modal image registration algorithm for a monitoring system, based on block matching

Key words

Image registration, image fusion, surveillance system, phase correlation, FFT.

Summary

Sophisticated video surveillance systems use many cameras for watching over the same area. Image fusion allows combining two or more images into a single image containing the most relevant information. One of the most important phases of image fusion is image registration. In this article, we present a block based image registration algorithm for multi-modal images, using the example of TV and thermal (IR) camera images acquired by a monitoring head. For this type of head, the proposed algorithm searches only for translation parameters to align both images; scale and rotation parameters are assumed to be constant, and distortion is neglected. The rough translation parameters are calculated by the classic phase correlation method for image registration. Then, the same method is used to vertically align corresponding rectangular blocks of both images. Inaccurate alignment parameters are detected by the analysis of these parameters in some antecedent time probes and adequately corrected. Data integration by filling gaps between image blocks constitutes the last phase of the presented algorithm. This algorithm delivers good registration effects for images with several near and distant planes of images and preserves a low computation complexity enabling real-time hardware implementation.