

Sztuczna inteligencja i inżynieria wiedzy

laboratorium

Ćwiczenie 4. Analiza obrazu – określanie podobieństwa obrazów

opracowanie: M. Paradowski, H. Kwaśnicka

Cel ćwiczenia

Zapoznanie się z podstawowymi metodami określania podobieństwa obrazów, poprzez własnoręczną implementację oraz zbadanie ich właściwości.

Realizacja ćwiczenia

- 1) Zapoznanie się z gotowymi narzędziami do wyznaczania punktów kluczowych na obrazie.
- 2) Zapoznanie się z algorytmem *analizy spójności sąsiedztwa pary punktów kluczowych*.
- 3) Zapoznanie się z metodą *Random Sample Consensus (RANSAC)*.
- 4) Implementacja metody określania podobieństwa obrazów:
 - a) Implementacja metody wczytującej dane o punktach kluczowych dla zadanego obrazu.
 - b) Implementacja metody wyznaczania par punktów kluczowych na dwóch obrazach.
 - c) Implementacja metody analizy spójności sąsiedztwa pary punktów kluczowych.
 - d) Implementacja metody wyznaczania dominującej transformaty geometrycznej pomiędzy obrazami (metoda *RANSAC*).
- 5) Zaimplementowana metoda ma mieć możliwość **wizualizacji par punktów kluczowych** powstałych na etapie (2) oraz (3) z poprzedniego punktu.
- 6) Przebadanie efektywności i skuteczności działania metody zgodności sąsiedztwa dla:
 - a) Różnych wartości parametru liczności sąsiedztwa,
 - b) Różnych wartości parametru spójności sąsiedztwa.
- 7) Przebadanie efektywności i skuteczności działania metody *RANSAC* dla:
 - a) Transformaty afinicznej oraz perspektywicznej,
 - b) Różnych wartości progu akceptacji danych,
 - c) Różnej liczby iteracji metody.
- 8) Prezentacja najciekawszych wyników (zarówno pozytywnych jak i negatywnych)
- 9) Dyskusja wyników
- 10) Przygotowanie sprawozdania zawierającego powyższe punkty (pomijając punkt związany z implementacją).

Ogólny schemat metody z wykorzystaniem analizy zgodności sąsiedztwa

WEJŚCIE: Obrazy I oraz J

1. Wyznacz punkty kluczowe dla obrazów I i J.
2. Dla każdego punktu kluczowego na obrazach I i J wyznacz jego sąsiedztwo.
3. Wyznacz pary punktów kluczowych metodą wzajemnego najbliższego sąsiada.
4. Zweryfikuj wszystkie pary punktów kluczowych przy użyciu metody analizy spójności sąsiedztwa, odrzuć pary punktów niespełniające kryterium spójności.

WYJŚCIE: Zbiór spójnych par punktów kluczowych.

Ogólny schemat metody z wykorzystaniem RANSAC

WEJŚCIE: Obrazy I oraz J

5. Wyznacz punkty kluczowe dla obrazów I i J
6. Wyznacz pary punktów kluczowych metodą wzajemnego najbliższego sąsiada
7. Wyznacz dominujące przekształcenie geometryczne na znalezionych parach punktów kluczowych przy użyciu metody RANSAC
8. Zweryfikuj, które pary punktów kluczowych pasują do znalezionej przekształcenia geometrycznego

WYJŚCIE: Zbiór par punktów kluczowych pasujący do przekształcenia geometrycznego.

Gotowe narzędzia do wyznaczania punktów kluczowych na obrazie

Punkty kluczowe (określane również terminem regiony kluczowe) opisują lokalną charakterystykę wizualną danego obrazu. Mają one najczęściej postać elips lub okręgów. Każdy punkt kluczowy opisany jest pojedynczym wektorem cech wyznaczonym dla obszaru obrazu, który ten punkt obejmuje. W Internecie dostępne jest bardzo wiele narzędzi do wyznaczania różnego rodzaju punktów kluczowych na obrazie. W przypadku tego ćwiczenia najbardziej interesujące są punkty kluczowe wykryte detektorem *Harris-Affine* oraz opisane deskryptorem *SIFT*. Gotowe narzędzie do wyznaczania tego typu punktów kluczowych można znaleźć pod adresem:

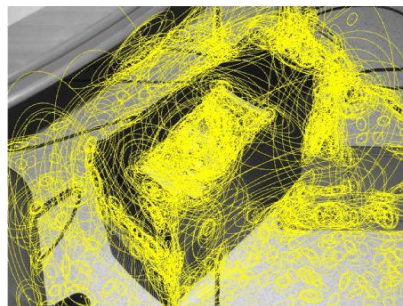
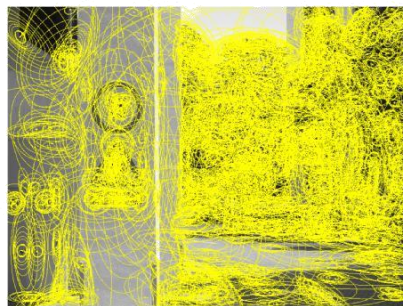
http://www.robots.ox.ac.uk/~vgg/research/affine/det_eval_files/extract_features2.tar.gz

Wyznaczenie punktów kluczowych na obrazie „obraz.png” uzyskujemy poprzez wywołanie następującego polecenia (Linux):

```
./extract_features_32bit.ln -haraaff -sift -i obraz.png -DE
```

Na stronie WWW podanej w spisie literatury dostępna jest również wersja aplikacji pod Windows. Jej uruchomienie wymaga jednak pobrania dodatkowych bibliotek Cygwin. Należy zwrócić jednak uwagę na fakt, iż wskazana aplikacja to najnowsza wersja zalecanego detektora punktów kluczowych.

W wyniku tego wywołania otrzymujemy plik „obraz.png.haraaff.sift” zawierający wykryte punkty kluczowe. Parametr -DE jest opcjonalny i odpowiada za rysowanie wyznaczonych punktów kluczowych. Wyznaczone punkty kluczowe mają postać eliptyczną. Przykłady wyznaczonych punktów kluczowych przedstawione zostały na poniższych obrazach:



Pliki *.haraff.sift mają następującą postać:

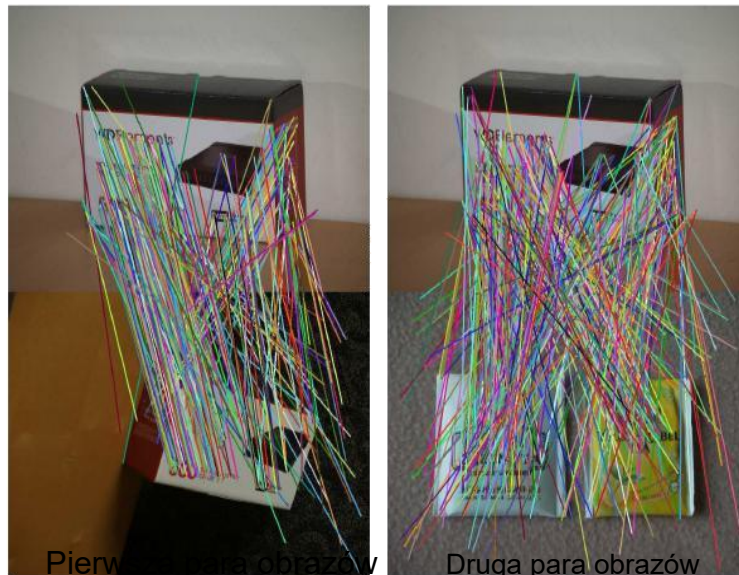
```
[liczba cech dla każdego punktu kluczowego, zawsze 128 dla .haraff.sift] [liczba punktów kluczowych]
[współrzędna x] [współrzędna y] [param A] [param B] [param C] [cecha 1] [cecha 2] ... [cecha 128]
[współrzędna x] [współrzędna y] [param A] [param B] [param C] [cecha 1] [cecha 2] ... [cecha 128]
...
[współrzędna x] [współrzędna y] [param A] [param B] [param C] [cecha 1] [cecha 2] ... [cecha 128]
```

Parametry „**współrzędna x**” oraz „**współrzędna y**” określają współrzędne środka punktu kluczowego (elipsy) na obrazie. Dla uproszczenia doboru parametrów metody współrzędne można znormalizować do przedziału (0, 1). Parametry „**A**”, „**B**” oraz „**C**” określają kształt punktu kluczowego (elipsy) i są w naszym przypadku pomijalne. Parametry „**cecha n**” stanowią reprezentację wizualną danego punktu kluczowego i służą do wyznaczania par punktów kluczowych.

Wyznaczanie par punktów kluczowych

Do wyznaczania par punktów kluczowych zastosować należy algorytm **wzajemnego najbliższego sąsiada**. Dla każdego punktu kluczowego na obrazie A wyznaczany jest jego najbliższy sąsiad (w sensie reprezentacji wizualnej punktów kluczowych – a więc wszystkie wartości „**cecha n**”) na obrazie B. Następnie dla każdego punktu na obrazie B wyznaczamy najbliższego sąsiada (również w sensie wizualnym) na obrazie A. Para punktów kluczowych jest generowana tylko wtedy, gdy punkt A wskazuje na punkt B oraz punkt B wskazuje na punkt A (wzajemność).

Przykłady par punktów kluczowych przedstawione zostały na poniższych obrazach. Każda para narysowana jest w postaci odcinka o innym kolorze. Metoda wzajemnego najbliższego sąsiada zwraca relatywnie dużą liczbę par punktów kluczowych. Pomimo tego, że jest to jedna z najlepszych jakościowo metod, znacząca część par punktów kluczowych stanowi szum informacyjny.



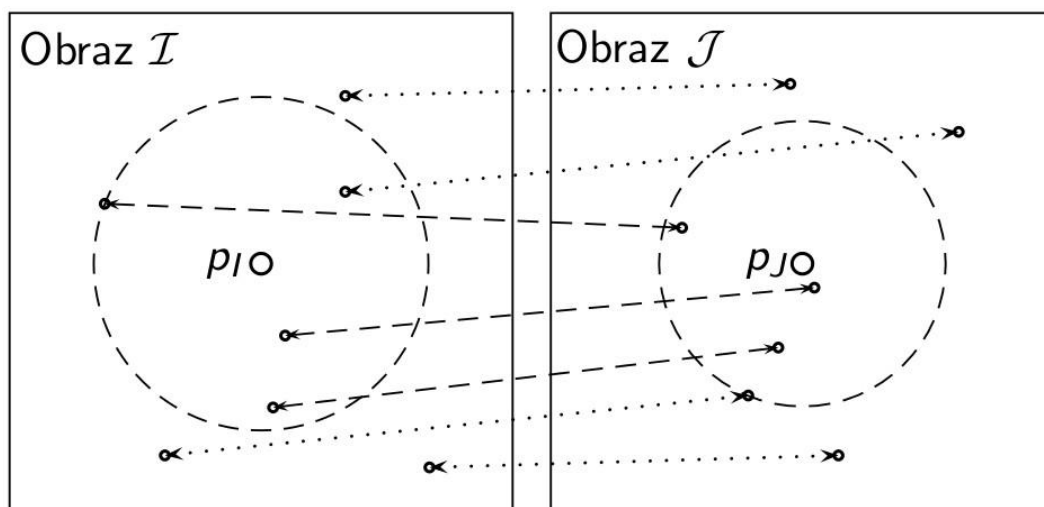
Wyznaczone w ten sposób pary punktów kluczowych są niewystarczające do poprawnego określenia podobieństwa obrazów. Konieczna jest dodatkowa weryfikacja sensowności wyznaczonych par. Do tego celu posłużyć mogą cechy **geometryczne** obrazów.

Algorytm analizy spójności sąsiedztwa punktów kluczowych

Na wejściu algorytm analizy spójności sąsiedztwa przyjmuje zbiór par punktów kluczowych wraz z informacją o sąsiedztwie każdego z punktów tej pary. **Wyjściem** algorytmu jest podzbiór par spełniający założone kryterium spójności. Pozostałe pary są odrzucone.

Sąsiedztwem punktu kluczowego P_i na obrazie I nazwiemy zbiór (o zadanej liczności) punktów kluczowych na obrazie I , których odległość od punktu P_i jest najmniejsza i większa od 0. Jednocześnie para punktów kluczowych Q_i - Q_j **sąsiaduje z inną parą punktów** P_i - P_j kluczowych jeżeli punkt Q_i należy do sąsiedztwa punktu P_i , a punkt Q_j należy do sąsiedztwa punktu P_j .

Sąsiedztwo pary punktów kluczowych P_i - P_j jest **sąsiedztwem spójnym** jeśli para punktów kluczowych P_i - P_j sąsiaduje z **co najmniej** n innych par punktów kluczowych.



Na powyższym przykładowym rysunku przyjęto licznosc sąsiedztwa punktów kluczowych równą 4 (dla punktu P_i oraz P_j rozważane są 4 najbliższe im punkty). Punkty P_i oraz P_j stanowią parę. W zbiorze wszystkich dostępnych par odnajdywane są tylko te pary, których punkty należą zarówno

do sąsiedztwa P_i oraz P_j . Maksymalna możliwa liczba tych par zbudowana przez metodę wzajemnego najbliższego sąsiada jest równa wielkości sąsiedztwa i wynosi 4. Na przedstawionym przykładzie odnaleziono 3 pary. Spójność pary punktów P_i - P_j oceniona jest zatem na $\frac{3}{4}$.

Algorytm analizy spójności sąsiedztwa iteruje po wszystkich parach punktów kluczowych. Dla każdej pary z osobna badana jest jej spójność sąsiedztwa. Jeśli para punktów kluczowych jest spójna to jest ona akceptowana. Pary niespójne są odrzucane.

Wartości parametrów algorytmu zależą od stopnia złożoności (liczby punktów kluczowych) analizowanych obrazów. Przy założeniu liczby punktów kluczowych około 2500, liczebność sąsiedztwa zazwyczaj nie powinna być mniejsza o 25-50 (zbyt mało danych). Zbyt duże sąsiedztwa (np. 200) również nie są zalecane (sąsiedztwo przestaje mieć znaczenie). Próg spójności sąsiedztwa zazwyczaj powinien być większy około od 50% wielkości sąsiedztwa.

Metoda RANSAC

Na wejściu metoda RANSAC przyjmuje zbiór par punktów kluczowych. **Wyjściem** metody RANSAC jest przekształcenie geometryczne, które „najlepiej pasuje” do możliwie największej liczby par punktów kluczowych. Metoda RANSAC jest pseudolosową metodą iteracyjną. W przypadku tego ćwiczenia zastosować należy przekształcenie **afiniczne** oraz **perspektywiczne**. Błąd przekształcenia (*modelerror*) wyznaczamy jako odległość euklidesową punktu docelowego i punktu wyznaczonego przez przekształcenie geometryczne.

RANSAC estymuje model najlepiej pasujący do danych D

```
1: bestmodel  $\leftarrow$  nil
2: bestscore  $\leftarrow$  0
3: for  $i = 1 \rightarrow \text{iter}$  do
4:   model  $\leftarrow$  nil
5:   repeat
6:      $s \leftarrow$  randomly choose  $n$  samples from  $D$ 
7:     model  $\leftarrow$  calculatemodel( $s$ )
8:   until model = nil
9:   score  $\leftarrow$  0
10:  for all  $\text{data} \in D$  do
11:    error  $\leftarrow$  modelerror(model, data)
12:    if error < maxerror then
13:      score  $\leftarrow$  score + 1
14:    end if
15:  end for
16:  if score > bestscore then
17:    bestscore  $\leftarrow$  score
18:    bestmodel  $\leftarrow$  model
19:  end if
20: end for
21: return bestmodel
```

Modele przekształceń obrazu stosowane w metodzie RANSAC

W metodzie RANSAC należy zastosować dwa modele przekształcenia punktów kluczowych z jednego obrazu w drugi obraz: transformatę **afiniczną** oraz **perspektywiczną**. Transformata afiniczna jest trójpunktowa, co oznacza, że trójkąt na obrazie A jest przekształcany jednoznacznie w trójkąt na obrazie B. Transformata perspektywiczna jest czteropunktowa, co oznacza, że czworokąt

na obrazie A jest przekształcany jednoznacznie w czworokąt na obrazie B. Transformację afiniczną można zatem wyznaczyć z trzech par punktów kluczowych, transformację perspektywiczną z czterech par punktów kluczowych.

Transformację afiniczną można wyznaczyć z trzech par punktów $(x_1, y_1) - (u_1, v_1)$, $(x_2, y_2) - (u_2, v_2)$ oraz $(x_3, y_3) - (u_3, v_3)$ na podstawie poniższego równania:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 \\ x_2 & y_2 & 1 & 0 & 0 & 0 \\ x_3 & y_3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 \\ 0 & 0 & 0 & x_3 & y_3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Transformata perspektywiczna. Transformację perspektywiczną można wyznaczyć z czterech par punktów na podstawie równania:

$$\mathbf{H} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -u_2x_2 & -u_2y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -u_3x_3 & -u_3y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -u_4x_4 & -u_4y_4 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -v_2x_2 & -v_2y_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -v_3x_3 & -v_3y_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -v_4x_4 & -v_4y_4 \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Do obliczeń na macierzach można w tym ćwiczeniu użyć gotowego pakietu.

Przyspieszanie działania metody RANSAC – heurystyka odległości par punktów

RANSAC jest metodą pseudolosową. Liczba możliwych do wyboru trójek (transformata afiniczna) lub czwórek (transformata perspektywiczna) jest bardzo duża. Znacząca większość kombinacji par punktów jest jednak niepoprawna. Istnieje wiele metod przyspieszania działania metody RANSAC, w tym prosta *heurystyka* oparta na odległościach współrzędnych (x, y) na obrazie A oraz (u, v) na obrazie B) par punktów kluczowych. Podkreślić należy, iż przedstawiona poniżej metoda jest heurystyką, a więc nie musi zadziałać w każdym przypadku.

Pierwsza para z trójki (lub czwórki) jest wybierana całkowicie losowo. Następne pary wybierane są również losowo, jednak z uwzględnieniem współrzędnych pierwszej i kolejnych par. Druga, trzecia (i ewentualnie czwarta) para powinny znajdować się dość blisko pierwszej pary (np. bliżej niż R). Uzasadnienie tej heurystyki jest dość proste – jeśli na dwóch obrazach występuje ten sam obiekt, to bardziej prawdopodobne jest wylosowanie punktów należące do obiektu blisko od siebie, niż daleko od siebie. Z drugiej strony pary nie powinny leżeć zbyt blisko siebie (np. dalej niż r), gdyż wtedy błędy obliczeniowe (położenia punktów kluczowych są tylko przybliżeniem!) będą zbyt wielkie. Przykładowe warunki, jakie można nałożyć na losowane pary (tutaj: trójkę par) są następujące:

$$\begin{aligned} r^2 &< (x_1 - x_2)^2 + (y_1 - y_2)^2 < R^2 & \text{oraz} & & r^2 &< (u_1 - u_2)^2 + (v_1 - v_2)^2 < R^2 \\ r^2 &< (x_1 - x_3)^2 + (y_1 - y_3)^2 < R^2 & \text{oraz} & & r^2 &< (u_1 - u_3)^2 + (v_1 - v_3)^2 < R^2 \\ r^2 &< (x_2 - x_3)^2 + (y_2 - y_3)^2 < R^2 & \text{oraz} & & r^2 &< (u_2 - u_3)^2 + (v_2 - v_3)^2 < R^2 \end{aligned}$$

Sugerowane wartości parametrów: r – około 1% wielkości obrazu, R – około 30% wielkości obrazu

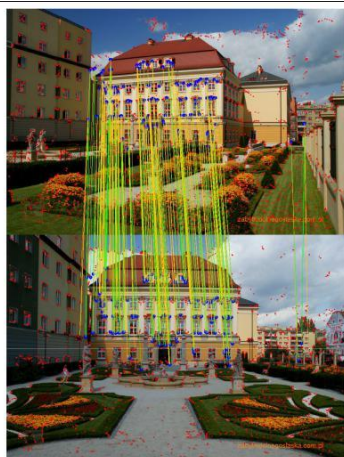
Dobór parametrów metody

Dobór parametrów metody jest istotnym elementem ćwiczenia i ma być zrealizowany przez osobę wykonującą ćwiczenie. Należy wykonać badania sprawdzające jakość i czas działania metody w zależności od wybranych parametrów. **W sprawozdaniu należy zasugerować „odpowiednio dobre” parametry metody.**

Przykładowe wyniki

Poniżej przedstawione zostały przykładowe wyniki uzyskane przez metodę RANSAC dla transformaty perspektywicznej i afinicznej. Większa liczba uzyskanych par punktów kluczowych oznacza lepszą detekcję obiektu. Przy wyraźnej detekcji dużych obiektów liczba par w przykładzie przekracza 100. Dla obrazów, na których nie obserwujemy podobieństwa liczba ta oscyluje w okolicach 20. W przypadku obrazów opisanych znacząco różną liczbą punktów kluczowych (np. wyższa rozdzielczość) wartości te mogą być inne!

Na **czzerwono** oznaczone zostały wszystkie punkty kluczowe, na **niebiesko** oznaczono zaakceptowane punkty kluczowe. **Zielone linie** oznaczają zaakceptowane pary punktów kluczowych. **Żółte linie** wskazują lokalizacje punktów kluczowych na drugim obrazie wyznaczone przez znaną transformację (tylko RANSAC).



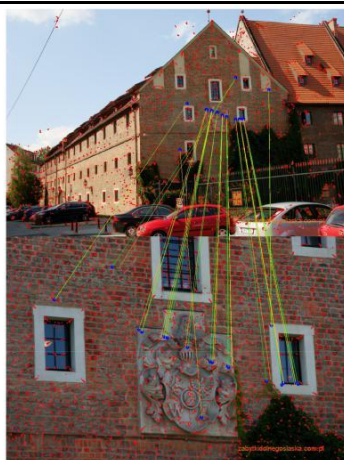
transformata perspektywiczna,
liczba znalezionych par = 196



transformata afiniczna,
liczba znalezionych par = 179



transformata afiniczna,
liczba znalezionych par = 45



transformata afiniczna,
liczba znalezionych par = 29



transformata perspektywiczna,
liczba znalezionych par = 22



transformata afiniczna,
liczba znalezionych par = 14

Zdjęcia Wrocławia zamieszczamy dzięki uprzejmości autorów strony internetowej: <http://www.zabytkidolnegoslaska.com.pl/>

Obrazy stosowane do testów

Obrazy stosowane do testów powinny mieć taką rozdzielczość, aby można na nich było odnaleźć odpowiednią liczbę punktów kluczowych (zazwyczaj powyżej 1000-2000 punktów). Sensowną rozdzielczością jest np. 800x600. Zbyt duża rozdzielczość (np. powyżej 2000 pikseli w osi) spowoduje wygenerowanie bardzo dużej liczby punktów kluczowych, co spowoduje znaczące wydłużenie czasu działania metody.

Do badań należy wziąć minimum 5 par wykonanych własnoręcznie zdjęć. Zdjęcia powinny przedstawiać przedmioty codziennego użytku, na przykład ulubioną książkę, kubek, myszkę komputerową. Każdy z wybranych przedmiotów należy sfotografować w dwóch różnych ujęciach i scenerii. Można wykonać badania zaimplementowanych algorytmów na drugim pakiecie 10 zdjęć, obrazujących bardziej skomplikowane sceny, na przykład zabytki architektury lub ujęcia większych fragmentów zabudowy miasta.

Ocena ćwiczenia

2pkt	Implementacja metody wyznaczania par punktów kluczowych
3pkt	Implementacja metody wyznaczania spójności sąsiedztwa oraz dominującego przekształcenia geometrycznego (RANSAC)
3pkt	Przebadanie i dyskusja skuteczności działania: <ul style="list-style-type: none">- algorytmu spójności sąsiedztwa (parametry: liczność sąsiedztwa, wartość progu spójności sąsiedztwa)- metody RANSAC dla transformaty afinicznej (parametry: liczba iteracji, progu błędu transformaty oraz heurystyki odległości par punktów).- metody RANSAC dla transformaty perspektywicznej
2pkt	Prezentacja wyników w postaci par obrazów z naniesionymi parami punktów kluczowych przed i po zastosowaniu algorytmu spójności sąsiedztwa oraz metody RANSAC

Literatura

1. Materiały do wykładu na BOARD: Mariusz.Paradowski/Systemy.Wizyjne/wykład 3 (wyszukiwanie, lokalizacja, rozpoznawanie).
2. <http://en.wikipedia.org/wiki/RANSAC>
3. http://en.wikipedia.org/wiki/Scale-invariant_feature_transform
4. <http://www.robots.ox.ac.uk/~vgg/research/affine/detectors.html>
5. <http://cmp.felk.cvut.cz/~chum/papers/Chum-PhD.pdf>
6. <http://www.zabytkidolnegoslaska.com.pl/>