## Rekonstrukcja geometrii

Celem rekonstrukcji geometrii jest otrzymanie jak najwięcej informacji o kształcie badanego przedmiotu. Reprezentowane są on najczęściej w postaci chmury punktów 3D, a to, w jaki sposób SfM pozwala na wyznaczenie tej chmury, jest tematem tego rozdziału.

### Punkt charakterystyczny

Kiedy rano szukamy swojego ulubionego kubka do kawy, próbujemy rozpoznać go w otoczeniu dużej liczby innych przedmiotów. Podświadomie porównujemy wtedy widziane przez nas przedmioty z obrazem kubka który mamy w pamięci. W zasadzie, nie zestawiamy ze sobą obrazów w sensie jeden do jednego. Porównujemy jedynie cechy szczególne. Cechami takimi mogą być: kolor, kształt lub rozmiar. Jeżeli na obu obrazach występują te same lub podobne znaki charakterystyczne wtedy z dużym prawdopodobieństwem znaleźliśmy swój kubek. Podobna sytuacja ma miejsce kiedy próbujemy przypomnieć sobie imię widzianej przed chwilą osoby. Próbujemy wtedy znaleźć w pamięci obraz osoby o podobnych cechach szczególnych. Dla ludzie takimi znakami mogą być: wzrost, budowa ciała, kolor oczu lub tatuaż.

Analogiczne podejście znalazło zastosowanie w szeroko rozumianej komputerowej analizie i przetwarzaniu obrazów. Bez względu na to czy mowa o śledzeniu pewnych obszarów zainteresowania, czy identyfikacji osób lub przedmiotów w sekwencji obrazów. Wszędzie tam, gdzie zachodzi potrzeba porównania dwóch obrazów, porównuje się znalezione na nich cechy charakterystyczne, nazywane także punktami charakterystycznymi.

Trudno o dokładną definicję punktu charakterystycznego. Jest to termin lepiej rozumiany intuicyjnie, stąd często zamiast dokładnej definicji w literaturze można natknąć się na ich przykłady. Ogólnie rzecz biorąc są to obszary na zdjęciu, posiadające unikalne, indywidualne atrybuty, które sprawiają, że łatwo jest je znaleźć i porównać z innymi. Punkty te powinna cechować powtarzalność, w praktyce oznacza to, że na dwóch zdjęciach przedstawiających ten sam obiekt ale wykonanych z różnej perspektywy, znalezione punkty kluczowe na pierwszym zdjęciu powinny zostać znalezione również na drugim. Ilość punktów charakterystycznych powinna być odpowiednio duża, a jednocześnie każdy z nich powinien posiadać jak najwięcej indywidualnych atrybutów, które pozwolą na prawidłowe znalezienie odpowiadających im punktów na innym zdjęciu.



Rysunek 1.1.1.1 Przykłady kandydatów na punkty charakterystyczne

W celu lepszego zrozumienia czym są punkty charakterystyczne/kluczowe warto posłużyć się powyższym rysunkiem. Zawiera on zdjęcie budynku oraz sześć jego fragmentów, które są kandydatami na punkty charakterystyczne.

Na fragmencie *A* nie występuje lokalnie duża różnica jasności, oprócz jednolitego koloru nie wyróżnia go nic co pozwalałoby na szybkie znalezienie jego dokładnego położenia. Fragment B przedstawia pewien wycinek elewacji. Jego zlokalizowanie może okazać się kłopotliwe, ponieważ duża część zdjęcia zawiera identyczny regularny wzorzec. Fragmenty C i D wydają się lepszymi kandydatami niż poprzednie A i B, ponieważ zawierają krawędzie. Znacząco ogranicza to obszar poszukiwań, jednak i w tym przypadku dopasowanie ich do konkretnego miejsca na obrazie może nie być proste. Ostatnie dwa E oraz F które prezentują naroża. Podobnych fragmentów na zdjęciu jest niewiele. Są to najlepsi kandydaci z przedstawionego zestawu. Ich wygląd oraz indywidualne atrybuty sprawiają, że dopasowanie ich do konkretnego miejsca na obrazie jest stosunkowo proste.

### Detekcja punktów charakterystycznych

Detekcja punktów charakterystycznych jest jednym z elementarnych zadań podczas analizy i przetwarzania obrazu. Narzędzia które służą do tego celu nazywa się detektorami cech. Istnieje bardzo wiele algorytmów przeznaczonych do wykrywania punktów charakterystycznych. Niektóre z nich odznaczają się większą ilością wykrytych punktów, inne krótszym czasem wyszukiwania, itd. Każdy z nich ma swoje zalety oraz wady i stosowany jest w zależności od stawianych wymagań.   
Przykładowymi algorytmami tego typu są detektory:

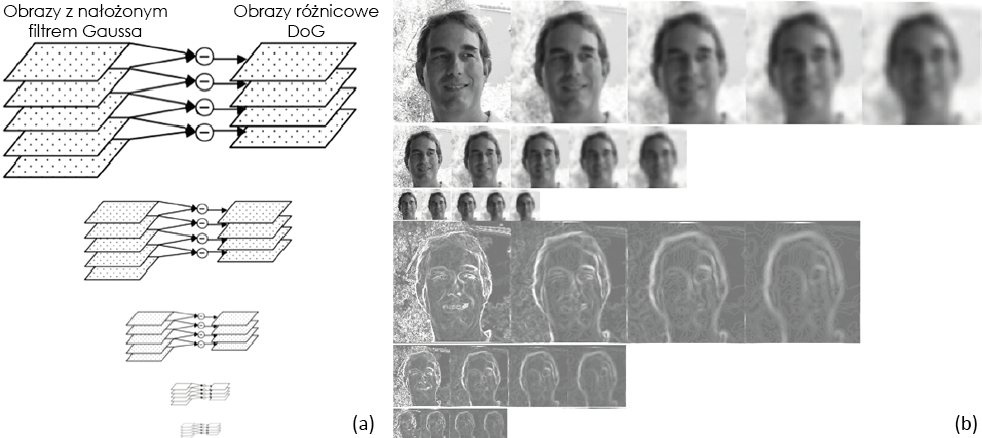
* Harris
* SIFT
* SURF
* FAST
* BRIEF
* ORB

Użytym w pracy został detektor SIFT. Cechuje go niezmienniczość względem rotacji, translacji i skali, a także duży współczynnik poprawnie dopasowanych par [23]. Z tego powodu dalsza część tego rozdziału będzie dotyczyła tylko jego implementacji. Pozostałe algorytmy zostały szczegółowo opisane m.in. w [23] [24].

Algorytm SIFT w pierwszym etapie detekcji cech jest dokonuje budowy piramidy obrazów. Jej konstrukcja polega na zastosowaniu filtru Gaussa z różnym parametrem rozmycia 𝜎 dla obrazu wejściowego w różnej skali, tj. dla różnych rozdzielczości.

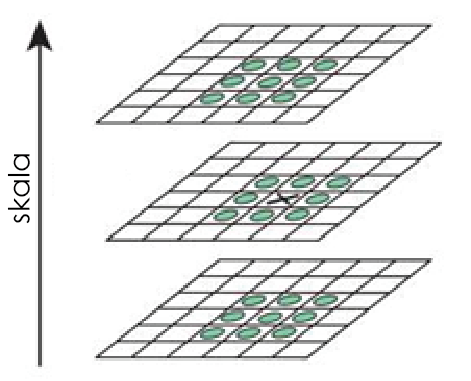
,gdzie I(x,y) - obraz wejściowy  
L(x,y,𝜎) – obraz wyjściowy  
 G(x,y,𝜎) – operator Gaussa =

Piramida obrazów podzielona jest na oktawy. Numer oktawy informuje o rozmiarze obrazu. Im jest on wyższy tym analizowany obraz jest mniejszy. W jednej oktawie znajduje się z kolei kilka warstw, czyli obrazów o różnym poziomie rozmycia.



Rysunek 1.1.2.1 Piramida obrazów SIFT. Rysunek (a) przedstawia ogólny schemat, natomiast (b) jego praktyczną implementacje. Obrazy umieszczone na górze są obrazami wejściowymi z nałożonym filtrem Gaussa. Obrazy na dole stanowią odpowiadające im obrazy różnicowe DoG.

Następnym krokiem jest obliczenie obrazów różnicowych DoG (ang. Difference of Gaussian). Są one wynikiem odejmowania odpowiadających sobie pikseli dla dwóch wybranych obrazów wejściowych z nałożonym filtrem Gaussa. Dla wszystkich obrazów DoG w każdej oktawie z wyjątkiem pierwszego i ostatniego poszukuje się lokalnych ekstremów. Każdy pixel nie brzegowy porównuje się z jego ośmioma sąsiednimi w tej samej skali i dziewięcioma w skali o jeden wyżej i niżej. W przypadku kiedy wartość pixela jest większa lub mniejsza od wartości wszystkich sąsiednich, punkt zostaje zakwalifikowany jako prawdopodobny punkt charakterystyczny.



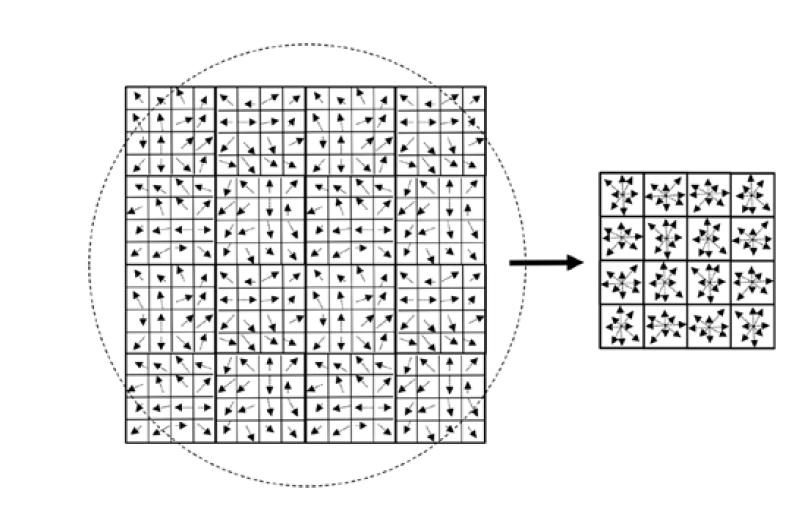
Rysunek 1.1.2.2 Poszukiwanie lokalnych ekstremów, jako możliwych położeń punktów charakterystycznych

Na koniec usuwa się te które mogą być niestabilne, tj. leżą na krawędziach lub nie spełniają warunku progu kontrastu

### Obliczenie deskryptorów

Podczas detekcji cech z użyciem detektora SIFT wyznaczana jest ich lokalizacja na zdjęciu, orientacja i skala. Nie jest to jednak wystarczająca ilość informacji, aby dokonać ich porównania i znalezienia odpowiadających sobie par. Dlatego dla każdego punktu oblicza się deskryptor, który jest jego „odciskiem palca”. Deskryptor to 128-wymiarowy wektor opisujący punkt charakterystyczny i jego otoczenie. Ogólna zasada jego obliczenia polega na:

1. Analizie obszaru o wymiarach 16x16 pikseli, zwierającego dany punkt charakterystyczny



Rysunek 1.1.3.1 Proces obliczania deskryptora dla punktu charakterystycznego

1. Obliczeniu gradientu dla każdego piksela w tym obszarze
2. Podzieleniu analizowanego obszaru na 16 bloków o wymiarach 4x4 piksele
3. Obliczeniu dla każdego bloku histogramu gradientu w 8 możliwych kierunkach
4. Połączeniu powyższych histogramów i otrzymaniu z nich 128 (4x4x8) wymiarowego wektora.

### Dopasowanie punktów charakterystycznych

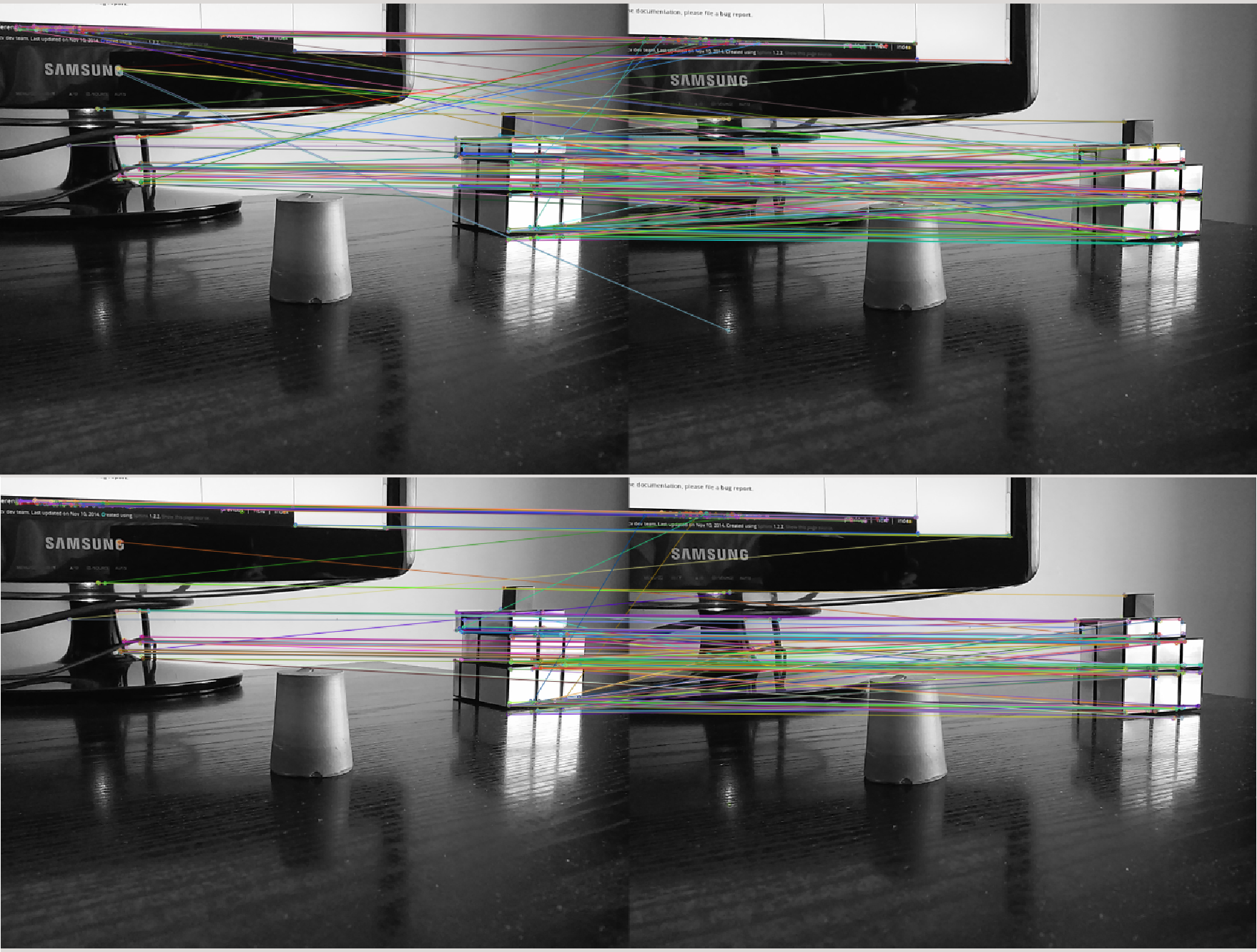
Z szeregu różnych implementacji etapu dopasowywania, często wykorzystywany jest model w którym początkowo punkty dopasowuje się metodą każdy z każdym, a następnie znalezione pary poddaje się testom. Użycie testów ma na celu usunięcie jak największej ilości złych dopasowań..

W przypadku omawianego deskryptora SIFT, wstępne dopasowanie punktów polega na obliczeniu odległości pomiędzy deskryptorem z i-tego i z j-tego zdjęcia. Im mniejsza odległość, tym większe prawdopodobieństwo, że opisują one ten sam punkt charakterystyczny. Każdy z deskryptorów to 128 wymiarowy wektor, dlatego aby wyznaczyć odległości między nimi można użyć normy euklidesowej , która jest pierwiastkiem z sumy kwadratów współrzędnych. Tak więc, dla dwóch wektorów i ich norma euklidesowa równa jest

(3.5.1)

,gdzie *n* dla deskryptora SIFT wynosi 128.

Jednym z testów sprawdzających czy punkty zostały poprawnie dopasowane jest test symetrii. Polega on na znalezieniu dla każdego punktu charakterystycznego na pierwszym zdjęciu jego najlepszego odpowiednika na drugim. W kolejnym kroku, postępując analogicznie, dla każdego punktu na drugim zdjęciu szuka się odpowiednika na pierwszym. Na koniec sprawdza się czy w obu przypadkach znalezione dopasowania są identyczne.

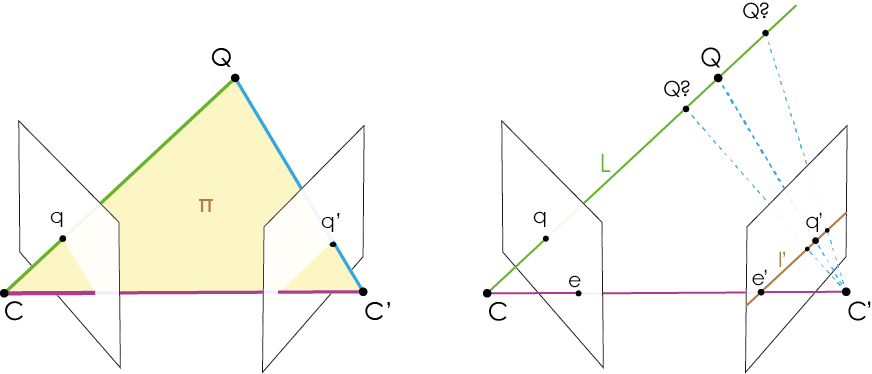


Rysunek 1.1.4.1 Test symetrii. Kolorowe linie łączą dopasowane punkty charakterystyczne na dwóch obrazach.   
Rysunek na górze przedstawia wstępnie dopasowane pary, a rysunek na dole - po przeprowadzeniu testu symetrii.

Kolejnym, niezwykle ważnym testem jest weryfikacja geometryczna. Do zrozumienia jej działania niezbędna jest znajomość geometrii epipolarnej.

Geometria epipolarna

Geometria epipolarna zajmuje się opisem geometrycznych zależności, jakie zachodzą między dwoma widokami wybranego przedmiotu z dwóch różnych pozycji.



Rysunek 1.1.4.2 Geometria epipolarna między dwoma widokami.   
Płaszczyzna epipolarna π zdefiniowana jest przez punkt 3D Q oraz centrum rzutowania obu kamer. Linia łącząca C i C’   
nazywa się linią bazową, natomiast położenie środka jednej z kamer na obrazie drugiej jest punktem epipolarnym e i e’.

Rozważmy sytuację, w której dwie kamery, o środkach w punktach C i C’ obserwują pewien punkt przestrzenny Q. Z poprzedniej części wstępu teoretycznego wiadomo, że obraz tego punktu powstanie w miejscach, w których promienie rzutujące przebiją rzutnie obu kamer, czyli w punktach q i q’. Wszystko to zdefiniowano wcześniej jako rzutowanie.

Nic nie stoi na przeszkodzi, żeby postępując analogicznie odwrócić ten proces, tj. wykonać rzutowanie wsteczne. Znając współrzędne punktu q na obrazie pierwszej kamery, można wyprowadzić równanie prostej L, która teoretycznie powinna połączyć C i Q, tak jak ilustruje to rysunek 1.1.1.1 b.   
Co zrozumiałe, nie jest możliwe określenie w ten sposób, gdzie dokładnie na tej prostej znajduje się Q. Skoro jednak L należy do płaszczyzny epipolarnej π, a ta przecina płaszczyznę obrazu drugiej kamery w miejscu linii epipolarnej l’, to odpowiadający punkt dla q musi znajdować się właśnie na l’.

Jeżeli wewnętrzne parametry kamer nie są znane, tj. kiedy obie kamery nie zostały wcześniej skalibrowane, ograniczenie to można zapisać w postaci

( 1.. )

, gdzie F jest macierzą fundamentalną, która pozwala przypisać każdemu punktowi q na pierwszym obrazie odpowiadającą mu linię epipolarną l’ na drugim. Macierz ta jest wymiaru 3x3 i zawiera informację o wzajemnym położeniu i orientacji obu kamer w układzie współrzędnych globalnych. Zawiera ponadto podstawowe informacje o ich wewnętrznych parametrach i z tego powodu łączy położenie pary dopasowanych punktów, których położenie określone jest w układzie współrzędnych pikselowych obrazu UWPO każdej z kamer.

Inną właściwością macierzy fundamentalnej jest

( 2. )

Powyższe równanie pozwala w prosty sposób zweryfikować, czy wybrana para punktów jest poprawnie dopasowana w tym sensie, czy punkty te leżą na odpowiadających sobie liniach epipolarnych.

W przypadku kiedy obie kamery zostały skalibrowane, rolę F zajmuje macierz esencjonalna E. Zawiera ona informacje jedynie o wzajemnym położeniu kamer, dlatego wyraża związek pomiędzy dwoma punktami zdefiniowanymi w układzie współrzędnych metrycznych obrazu UWMO. Przejście pomiędzy F i E wyraża się wzorem.

( 2.2 )

,gdzie K i K’ są macierzami kalibracji odpowiednio dla pierwszej i drugiej kamery. E jest szczególnym przypadkiem macierzy fundamentalnej stąd

( 2.2 )

Ostatecznie więc, dysponując wiedzą na temat wzajemnej geometrii obu kamer, a także o położeniu punktu charakterystycznego na obrazie pierwszej z nich, dokonuje się odrzucenia błędnych dopasowań, jeżeli nie leżą one na odpowiadającej im linii epipolarnej.

### Wyznaczenie macierzy projekcji

Przedstawione w poprzednim rozdziale macierz fundamentalna i esencjonalna odgrywają ważną rolę w rekonstrukcji geometrii z wykorzystaniem techniki SfM. Zawierają one informacje o wzajemnym położeniu i orientacji kamer rejestrujących obiekt z dwóch różnych perspektyw. Informacje te można wykorzystać do uzupełniania macierzy projekcji P, a w konsekwencji do wyznaczenia chmury punktów 3D.

Wyznaczenie macierzy F lub E sprowadza się do rozwiązania równania (2.2) dla co najmniej 7 par poprawnie dopasowanych punktów. W praktyce jednak korzysta się ze znormalizowanego algorytmu dla 8 par i metody najmniejszych kwadratów, gdyż dają one lepsze wyniki i są prostsze w obliczeniach. Problem znalezienia macierzy zarówno fundamentalnej jak esencjonalnej jest nie trywialny i bez dodatkowych mechanizmów łatwo w tym miejscu o błędne wyznaczanie współczynników F i/lub E. Szczególnie ważne jest tutaj podanie najlepiej pasujących par punktów, stąd często spotykanym rozwiązaniem jest implementacja algorytmu RANSAC (ang. RANdom Sample Consensus). W tym przypadku jego działania można opisać w kilku krokach.

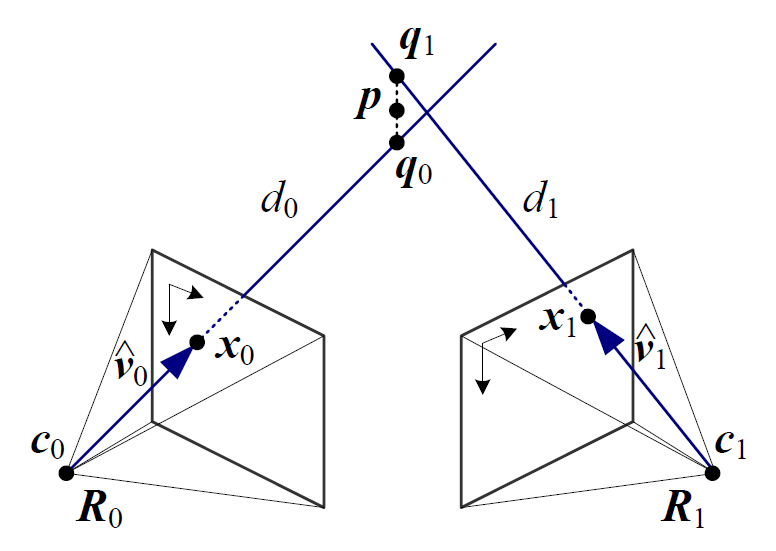
1. Ze zbioru wszystkich dopasowanych par punktów algorytm losowo wybiera, w zależności od zastosowanej metody wyznaczania macierzy 7 lub 8 par.
2. Na podstawie wybranej grupy punktów dokonuje się obliczenia macierzy F lub E
3. Dla wyznaczonej macierzy sprawdza się stopień dopasowania, tj. jaki procent wszystkich par punktów spełnia ograniczenie epipolarne (2.2)
4. Poprzednie kroki powtarza się w zadaną ilość razy
5. Ostatecznie wybiera się grupę tych punktów dla których obliczona macierz F lub E pasuje najlepiej do całego zbioru dopasowanych par.

### Triangulacja

Ostatnim krokiem w drodze do rekonstrukcji geometrii jest triangulacja, czyli wyznaczenie współrzędnych przestrzennych punktów na podstawie conajmniej dwóch ich rzutów. Niezbędna w tym celu jest znajomość geometrii kamery, tj. macierzy projekcji dla wszystkich rzutów.

Teoretycznie jest to zadanie proste dla każdej poprawnie dopasowanej pary punktów, która spełnia ograniczenie epipolarne. Znając położenie q i q’, parametry wewnętrzne i zewnętrzne obu kamer możliwe jest wyprowadzenie dla tych punktów prostych i które powinny pokrywać się z promieniami rzutującymi. Innymi słowy powinny przeciąć się dokładnie w punkcie Q w przestrzeni.

W praktyce jest to wysoce niemożliwe z wielu powodów. Położenie każdego punktu na obrazie ma charakter dyskretny, a nie ciągły. Wynika to z faktu, że płaszczyzną obrazu w aparacie jest macierz elementów światłoczułych o skończonych prostokątnych wymiarach. Ponadto do obliczeń zostają wprowadzone błędy wynikłe z wielu zaokrągleń, błędnie określonych położeń punktów charakterystycznych i inne. Z powyższych względów stosuje się różne alternatywne rozwiązania. Jednym z najprostszych jest metoda średniego punktu. Polega ona na znalezieniu punktu w przestrzeni który znajduje się najbliżej prostych i .



Rysunek 1.1.6.1 Metoda średniego punktu