AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Śledzenie obiektów

Tomasz Kryjak

Wydział EAIiIB
Katedra Automatyki i Robotyki

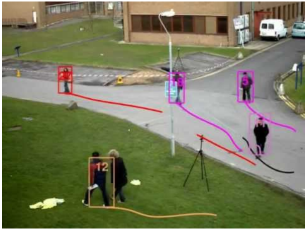
28.05.2018

Śledzenie

- Co to jest śledzenie ?

Śledzenie

- Co to jest śledzenie ?
- Estymacja trajektorii obiektu poruszającego się na scenie (tj. przypisanie etykiet do poszczególnych obiektów).
- Oczywiście ten sam obiekt, ma mieć w trakcie obecności na scenie cały czas tą samą etykietę.



A co to jest obiekt ?

Obiekt – coś co jest interesujące w rozważanej aplikacji (ludzie, samoloty, samochody, krople wody itp.)

Czy śledzenie to temat ważny ?

Czy śledzenie to łatwe zadanie w automatycznych systemach wizyjnych ?

Problemy

- nagły, szybki ruch obiektu,
- zmiana wyglądu obiektu,
- zmiana wyglądu sceny,
- obiekty, które nie są bryłami sztywnymi i zmieniają w trakcie śledzenia swój kształt (patrz człowiek),
- przesłaniania (obiekt/obiekt, obiekt/scena),
- ruch kamery,
- wymaganie działania w czasie rzeczywistym,
- zmiany oświetlenia na scenie,
- szum,
- redukcja informacji związana z przejściem z 3D do 2D (szeroko rozumiana).

Jak sobie z nimi radzić ?

Primo. Zastosować dobry algorytm.

Ograniczenia/założenia:

- ruch obiektów jest płynny,
- stała prędkość, stałe przyspieszenie,
- wiedza a priori o liczbie i rozmiarze obiektów, wyglądzie, kształcie.

Algorytmów śledzenia obiektów jest wiele ...

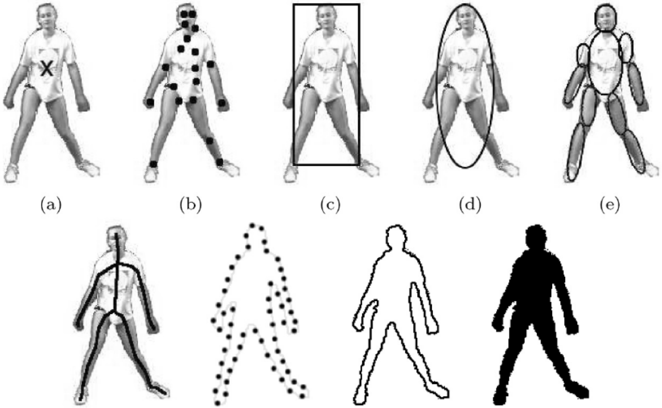
Pytania różnicujące:

- Jaka reprezentacja obiektu jest najlepsza ?
- W jaki sposób modelować ruch, wygląd czy kształt obiektu ?
- Jakie cech obrazu powinny zostać użyte ?

Reprezentacja obiektu - kształt

- punkt (środek ciężkości)
- proste figury geometryczne (prostokąt, elipsa) – łatwo modelować ruch np. transformacja afiniczna. Zwykle do brył sztywnych, ale nie tylko.
- sylwetka, kontur - złożone obiekty nie będące bryłami sztywnymi,
- przeguby (przykładowo części ciała człowieka)
- szkielet.

Reprezentacja obiektu - kształt



(a) (b) (c) (d) (e)

(f) (g) (h) (i)

Źródło: Object tracking, A. Yilmaz, O. Javed, M. Shah, ACM Computing Surveys, Vol. 38, No. 4, Article 13

Reprezentacja obiektu - wygląd

- rozkłady gęstości prawdopodobieństwa wyglądu obiektu.
  - parametryczne: Gaussian, Mixture of Gaussian.
  - nieparametryczne: okno Parzen'a, histogramy.
- wzorce
- aktywne modele wyglądu:
  - jednoczesne modelowanie kształtu i wyglądu,
  - kształt jest określony jako zbiór punktów charakterystycznych (każdy ma wektor cech np. kolor, teksturę, jasność, amplitudę gradientu).
  - w fazie uczącej model jest przygotowywany na podstawie zbioru uczącego (np. PCA).
- modele wielowidokowe.

Wykorzystywane cechy obrazu

Bardzo ważna sprawa. Cechy powinny w sposób (względnie) unikalny opisywać dany obiekt.
Przykłady:

- kolor,
- krawędzie (wykrycie konturów) – mniej wrażliwe od koloru na zmiany oświetlenia,
- przepływ optyczny,
- tekstura.

Detekcja śledzonych obiektów

Co najmniej na jednej ramce, czasem na każdej (śledzenie przez detekcję) trzeba wykryć obiekty, które mamy zamiar śledzić
Przykłady metod:

- detekcja punktów charakterystycznych (Harris, SIFT, SURF, KLT),
- odejmowanie tła (segmentacja obiektów pierwszoplanowych),
- segmentacja (inne metody wyodrębniania interesujących nas obiektów),
- uczenie maszynowe (np. klasyfikator neuronowy),
- "ręczne wskazanie".

Algorytm *mean shift*

Śledzenie oparte na analizie rozkładu prawdopodobieństwa pewnej cechy na obrazie. Przykładowo barwa (H – hue) z przestrzeni HSV. Inaczej mówiąc – prawdopodobieństwo, że obiekt znajduje się w danej lokalizacji.

Algorytm iteracyjny – znajdywanie maksimum funkcji gęstości (największego prawdopodobieństwa).

Mean shift – średnie przesunięcie. Nazwa dość dobrze charakteryzuje działanie. W pewnym obszarze obliczany jest środek ciężkości punktów, a następnie aktualny środek obszaru jest przesuwany do znalezionego środka ciężkości.

Mean shift

Mamy zbór punktów $x_i, i = 1...N$ oraz środek obszaru y_0. Wektor *mean shift* dany jest wzorem:

$$M(y) = \frac{1}{N}\sum_{i=1}^{N} x_i - y_o \tag{1}$$

Uwaga. Takie rozwiązanie zakłada, że waga poszczególnych punktów w rozważanym obszarze jest taka sama, co niekoniecznie może być pożądane.

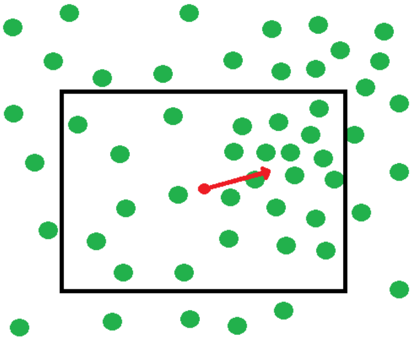
Mean shift

Ale zawsze możemy wprowadzić wagi ...

$$M(y) = \frac{1}{N}\sum_{i=1} N \frac{w_i(y_0)\cdot x_i}{w_i(y_0)} - y_o \quad (2)$$

Przykładowo waga może być zależna od odległości rozpatrywanego punktu od środka y_0 (im dalej tym punkt mniej ważny).

Interpretacja graficzna *mean shift*



Mean shift

Maksymalnie uproszczony opis:

1. wybór obiektu śledzonego (np. ręczny),
2. ekstrakcja cech wzorca (funkcja gęstości rozkładu prawdopodobieństwa),
3. ekstrakcja cech kandydata,
4. określenie podobieństwa wzorca i kandydata (potrzebujemy funkcję),
5. wyliczenie nowych współrzędnych kandydata,
6. stop lub powrót do 3.

Wybieramy obiekt do śledzenia

Czy wszystkie piksele w oknie powinny być tak samo ważne ?

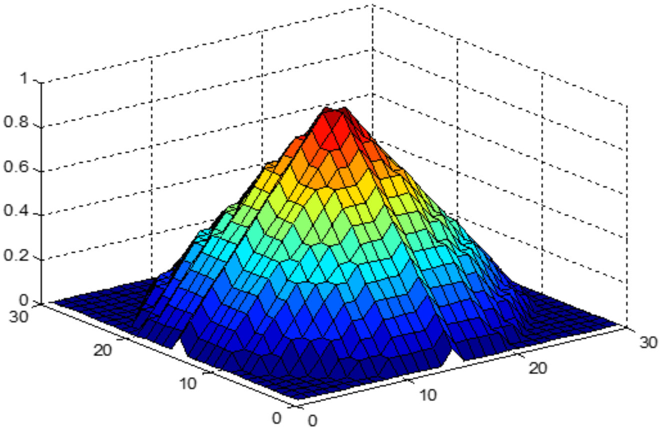
Zakłada się, że ważniejsze są piksele w środku. Dlatego wyznacza się jądro (*kernel*) np. Gaussa, Epanechnikova, kwadratowe, cosinusowe, czy trójkątne.

Jądro trójkątne:

$$K(u) = 1 - ||u||, \quad ||u|| < 1 \tag{3}$$

u stosunek odległości danego punktu od środka obszaru do długości promienia okręgu wpisanego w dany obszar.

Jądro o rozmiarze 30×30



Źródło: K. Mazur – praca inżynierska

Oszacowanie gęstości jądra

Mamy N punktów x_i w przestrzeni R^2 (w naszym obszarze zainteresowania).

Wykorzystując wcześniej przedstawione jądro, możemy zapisać funkcję gęstości w punkcie (x) jako:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K(||x - x_i||^2) \tag{4}$$

Różniczkujemy:

$$\nabla f(x) = \frac{2}{n}\sum_{i=1}^{n}(x - x_i)K'(||x - x_i||^2) \tag{5}$$

Oszacowanie gęstości jądra

Oznaczamy $-K'(x)$ poprzez $g(x)$. Wtedy:

$$\nabla f(x) = \frac{2}{n}\sum_{i=1}^{n}(x_i - x)g(||x - x_i||^2) \tag{6}$$

Przekształcając:

$$\nabla f(x) = \frac{2}{n}\sum_{i=1}^{n}[g(||x - x_i||^2)][\frac{\sum_{i=1}^{n} x_i g(||x - x_i||^2)}{\sum_{i=1}^{n} g(||x - x_i||^2)} - x] \tag{7}$$

Pierwszy człon to gradient jądra, a drugi to tzw. wektor *mean-shift* (wspomniany wcześniej).

Odległość pomiędzy dwoma funkcjami gęstości

Jeśli mamy wzorzec (obszar z ramki $t-1$) i kandydata (obszar z ramki t), to musimy je porównać. Może to być tzw. współczynnik Bhattacharyya dany jako:

$$\rho(y) = \sum_{u=1}^{m} \sqrt{(p_u(y) q_u)} \tag{8}$$

gdzie: q – wzorzec (histogram), $p(y)$ – kandydat w punkcie y (też histogram).

Jak to działa ?

Na początku dokonujemy inicjalizacji śledzenia tj. określamy jądra (i jego gradient) oraz określamy funkcję gęstości prawdopodobieństwa. Stosujemy histogram HSV, przy czym uwzględniamy wagi wynikające z funkcji jądra.

Formalnie:

$$q_u = C \sum_{i=1}^{N} K(||x_i||^2)\delta[b(x_i) - u] \tag{9}$$

gdzie: u – kolejna wartość barwy, C – współczynnik normalizujący (suma współczynników jądra). Finalnie otrzymujemy znormalizowany

histogram.

A później śledzimy ...

Wyznaczamy funkcję gęstości prawdopodobieństwa dla kandydata (znormalizowany histogram):

$$p_u(y) = C_h \sum_{i=1}^{N} K(||y - x_i||^2)\delta[b(x_i) - u] \tag{10}$$

gdzie: u – kolejna wartość barwy, C_h – współczynnik normalizujący (suma współczynników jądra), y – środek obszaru.

Porównujemy histogramy

$$\rho(y) = \sum_{u=1}^{m} \sqrt{(p_u(y) q_u)} \tag{11}$$

Szereg Taylora (dla otoczenia y_0)

$$\rho(p(y,q)) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{(p_u(y_0) q_u)} + \frac{1}{2} \sum_{u=1}^{m} p_u(y) \sqrt{\frac{q_u}{p_u(y_0)}} \tag{12}$$

Pierwszy człon nie zależy od y, czyli nie jest dla nas ciekawy, bo ostatecznie chodzi nam o maksymalizację wyrażenia w zależności od y.

Podstawiamy wzór na $p_u(y)$

$$\rho(p(y,q)) \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{(p_u(y_0)q_u)} + \frac{C_h}{2}\sum_{i=1}^{N} w_i K(||u - x_i||^2) \tag{13}$$

gdzie:

$$w_i = \delta[b(x_i) - u]\sqrt{\frac{q_u}{p_u(y_0)}} \tag{14}$$

Zatem maksymalizujemy skrajnie prawy człon. Do tej maksymalizacji używamy metody *mean shift*.

Wzór ostateczny ...

$$y_1 = \frac{\sum_{i=1}^{N} x_i w_i g(||y - x_i||^2)}{\sum_{i=1}^{N} w_i g(||y - x_i||^2)} \qquad (15)$$

A kiedy kończmy ?

- $||y1 - u_0||^2 < \epsilon$,
- maksymalna liczba iteracji.

Algorytm KLT

KLT to jeden ze starszych algorytmów śledzenia obiektów. Nazwa pochodzi od nazwisk twórców: Takeo Kanade, Bruce'a Lucasa oraz Carlo Tomasi. Jego pierwsza wersja została zaprezentowana w pracy „An Iterative Image Registration Technique with an Application to Stereo Vision" z roku 1981, a pewne ulepszenia zostały opisane w artykule „Detection and Tracking of Point Features" z roku 1991.

Idea opiera się na śledzeniu punktów charakterystycznych tj. wyliczaniu tzw. rzadkiego przepływu optycznego. Operacje przeprowadzane są na obrazach w odcieniach szarości.

Dobre cechy do śledzenia

Jeśli założymy, że chcemy śledzić położenie pikseli z ramki na ramkę to na obrazie znajdziemy „lepszych” i „gorszych” kandydatów.

Intuicyjnie. Na pewno złe będą jednorodne powierzchnie. Podobnie na krawędziach trudno dokonać jednoznacznego przypisania.

Najlepszym kandydatem wydają się być **narożniki**.

Czy można to wykazać ? (1)

Mamy obraz I, zbiór dopuszczalnych przekształceń $W(x,p)$ (x – lokalizacja, p – parametry przekształcenia) oraz fragment obrazu z poprzedniej ramki $T(x)$ – otoczenie punktu x.

błąd (SSD) -> $\sum_x [I(W,x,p) - T(x)]^2$

inkrementalne przesunięcie -> $\sum_x [I(W,x,p+\delta p) - T(x)]^2$

linearyzacja -> $\sum_x [I(W,x,p+\Delta p) + \nabla I \frac{\partial W}{\partial p} \Delta p - T(x)]^2$

$\nabla I = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}]$, $W(x,p)$ – Jakobian przekształcenia.

Czy można to wykazać ? (2)

Obliczamy pochodną z wyrażenia po Δp:
$2 \cdot \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [I(W, x, p + \Delta p) + \nabla I \frac{\partial W}{\partial p} \Delta p - T(x)]$

Przyrównujemy do zera:
$\Delta p = H^{-1} \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [T(x) - I(W, x, p + \Delta p)]$

gdzie H – Hessian: $H = \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [\nabla I \frac{\partial W}{\partial p}]$.

Czy można to wykazać ? (3)

Jeśli nasze śledzenie sprowadza się do iteracyjnego obliczania Δp, to stabilność procedury zależy głównie od obliczania odwrotności Hessianu.

Zatem pożądane jest aby: $\lambda_1 >> 0$ oraz $\lambda_2 >> 0$.

A co to oznacza w praktyce ??

Przykład

Załóżmy, że nasze przekształcenie to prosta translacja.
Wtedy:

$$W(x,p) = \begin{bmatrix} x_1 + p_1 \\ x_2 + p_2 \end{bmatrix} \tag{16}$$

$$\frac{\partial W}{\partial p} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{17}$$

$$H = \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [\nabla I \frac{\partial W}{\partial p}] = \sum_x \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{18}$$

$$H = \sum_x \begin{bmatrix} (\frac{\partial I}{\partial x})^2 & \frac{\partial I^2}{\partial x \partial y} \\ \frac{\partial I^2}{\partial x \partial y} & (\frac{\partial I}{\partial y})^2 \end{bmatrix} \tag{19}$$

Interpretacja H

Analizujemy wartości własne λ_1 i λ_2:

- $\lambda_1 \sim \lambda_2$ i obie małe – obszar jednorodny,
- $\lambda_2 >> \lambda_2$ – krawędź pozioma,
- $\lambda_1 >> \lambda_2$ – krawędź pionowa,
- $\lambda_1 \sim \lambda_2$ i obie duże – narożnik.

Uwaga. Proszę skojarzyć metodę z detekcję narożników Harrisa.