Bożena Frischmuth, Witold Kosiński
Sektion Mathematik
Wilhelm-Pieck Universität Rostock
oraz
Pracownia Modelowania Systemów
Komputerowego Wspomagania
ORT IPPT PAN Warszawa

O RODZINIE ZNIEKSZTAŁCEŃ W PROCESIE ROZPOZNAWANIA

1. Wstęp

Celem tego artykułu jest przedstawienie pewnego alternatywnego spojrzenia na metody budowania algorytmów systemu wizyjnego, które służą do identyfikacji elementów obrazu.

Człowiek w wielu przypadkach potrafi nie tylko rozpoznać przedmiot widziany w dowolnym położeniu, lecz także przedmiot widziany jako odbicie, np. na bombce choinkowej lub na zafalowanej powierzchni wody, na kliszy negatywu filmu; czasem można nawet rozpoznać przedmiotu z jego cienia. W każdym z wymienionych przypadków mamy do czynienia z rozpoznawaniem zniekształconych obrazów. Spostrzeżenia te mają swoje praktyczne znaczenie przy próbach sformalizowanego opisu procesu percepcji wizyjnej człowieka i jego komputerowym modelowaniu. Dodatkowym argument dla poruszania tego tematu jest fakt, że zagadnienia zniekształceń nieliniowych występują w układzie optycznym kamery systemu wizyjnego [1].

W przypadku ogólnych nieliniowych odwzorowań 3 czy 2 wymiarowej przestrzeni na płaszczyznę, bądź jej podzbiór - zbiór peli C tj. piksli lub plamek, jak proponuje się w innym miejscu w tym tomie) wydaje się, że identyfikacja obrazów na podstawie niezmienników przekształceń (czytaj - zniekształceń) jest niemożliwa z braku dostatecznej ilości takich niezmiennych własności, względnie z zupełnego braku niezmienników przekształceń nieliniowych. Również bezpośrednie zastosowanie teorii przestrzeni tolerancyjnych [2] nie wydaje się tutaj być odpowiednim podejściem. Będziemy natomiast stosować pewne uogólnienia metody generacji wymienionej w [3].

Idea proponowanego podejścia jest następująca. Zakładamy istnienie pewnej rodziny zniekształceń $\langle T_\theta |\ \theta \in \Theta \rangle$ o parametrze θ, w której dla zadanego parametru θ i zadanego obiektu A pojedyncze T_θ generuje "teoretyczny" obraz obiektu $T_\theta(A)$. Parametr θ może przyjmować wartości wektorowe, jeśli funkcja zniekształcenia jest wyznaczona przez więcej niż jeden parametr.

Rodzina T_θ powinna zwykle zawierać wszystkie transformacje afiniczne i dylatacje, jeśli obiekty pojawiaja się w polu widzenia kamery w różnych położeniach i w zmiennej od niej odległości. Przesunięcia sztywne z obrotami realizować będzie podzbiór macierzy trzy na trzy, dodatnie skalary będą parametryzować dylatacje (przeskalowania) obrazu. Perspektywa też powinna być uwzględniana w rodzinie T_θ.

Dla θ przebiegającego całą swoją dziedzinę zmienności Θ teoretyczny obraz $T_\theta(A)$ przebiega całą swoją orbitę. Wyznaczymy parametr θ tak, by odległość (odpowiednio zdefiniowana) między teoretycznym a widzianym obrazem była minimalna. Tę minimalną odległość uważamy za odległość orbity od widzianego obrazu. Następnie wybierzemy orbitę z minimalną odległością od widzianego obrazu, przy czym zbiór wszystkich orbit odpowiada katalogowi zarejestrowanych (czytaj - zapamiętanych) obiektów przez system wizyjny. Należy zwrócić tutaj uwagę na analogię do pewnych metod estymacji parametrów oraz do wyboru modeli w teorii regresji nieliniowej [4].

W przypadku wizji dwuwymiarowej (tzn. gdy mamy do czynienia z obiektami płaskimi) wydaje się naturalne założenie o istnieniu T_θ^{-1}. Można wtedy "wyczyścić" widziany obraz ze zniekształceń i szumów przy pomocy T_θ^{-1}, przy czym θ zostaje wyznaczony jak wyżej, a następnie zastosować zwykłe metody. W szczególności można, stosując T_θ do wewnętrznej i zewnętrznej aproksymacji wzorca [2], uzyskać tolerancję dla dopuszczalnej odległości między orbitą a widzianym obrazem.

W niniejszej pracy zajmujemy się wyłącznie wizją dwuwymiarową, a w szczególności identyfikujemy obiekty do rozpoznania z ich zdjęciami "wzorcowymi" oraz zakładamy odwracalność transformacji T_θ. Zadanie jest przez to symetryczne i polega na tym, żeby za pomocą transformacji dwa obrazy nałożyć jak najlepiej na siebie.

Pewna trudność polega na tym, że dyskretna siatka peli przy transformacjach nieliniowych nie przechodzi na siebie. W następnym punkcie przedstawiamy aparat potrzebny do pokonania tej trudności.

2. Generacja orbity

Niech $P = [0,n_x] \times [0,n_y] \cap \mathbb{Z}^2 \subset \mathbb{R}^2$, gdzie $\mathbb{Z}$ oznacza zbiór liczb całkowitych, będzie zbiorem peli zawartym w pewnym prostokącie Π; przez F oznaczamy zbiorem kolorów względnie stopni szarości. Dowolne odwzorowanie $W: P \longrightarrow F$ nazywamy obrazem, a zbiór wszystkich obrazów F^P oznaczamy przez $\mathcal{W}$. Następnie wprowadzamy operator prolongacji $\Omega : \mathcal{W} \longrightarrow F^{\mathbb{R}^2}$ oraz operatory restrykcji $R_\pi : F^{\mathbb{R}^2} \longrightarrow F^\pi$ oraz $R_p : F^{\mathbb{R}^2} \longrightarrow \mathcal{W}$, przy czym zakładamy

$$R_p\, \Omega = \mathrm{id}, \quad R_p R_\pi = R_p .$$

Niech zadane będą odwzorowania T_{θ_1}, T_{θ_2}, $\theta \in \Theta$,

$$T_{\theta_1} : \mathbb{R}^2 \xrightarrow[\text{na}]{1-1} \mathbb{R}^2,$$

$$T_{\theta_2} : F \xrightarrow[\text{na}]{1-1} F .$$

Odwzorowanie $T_\theta = (T_{\theta_1}, T_{\theta_2})$ działa na zadaną funkcję $\chi : \mathbb{R}^2 \longrightarrow F$ następująco:

$$(T_{\theta}\chi)(x) = T_{\theta_2}(\chi(T_{\theta_1}^{-1}(x))).$$

Przez analogię do pojęcia funktora w teorii kategorii [5], odwzorowanie T_{θ_1} można nazwać przyporządkowaniem obiektowym funktora kowariantnego (odwzorowania) T_{θ}. Łatwo się przekonać, że jeśli

$$T_{\theta_1} = T_{\theta''_1} \circ T_{\theta'_1} \text{ i } T_{\theta_2} = T_{\theta''_2} \circ T_{\theta'_2}, \text{ to } T_{\theta}\chi = T_{\theta''}(T_{\theta'}\chi).$$

Za pomocą prolongacji i restrykcji jesteśmy teraz w stanie określić również działanie T_{θ} na funkcję $w \in W$, mianowicie

$$T_{\theta}w = R_p T_{\theta} \Omega w.$$

Zauważmy jednak, że z relacji $R_p\Omega = \mathrm{id}$ nie wynika $\Omega R_p = \mathrm{id}$, a w związku z tym na ogół nie zachodzi $T_{\theta''}(T_{\theta'}w) = (T_{\theta''} \circ T_{\theta'})w$. Istotnie

$$T_{\theta''}(T_{\theta'}w) = R_p T_{\theta''} \Omega R_p T_{\theta'} \Omega w \neq (T_{\theta''} \circ T_{\theta'})w.$$

Będziemy nazywać umownie zbiór

$$O_w := (T_{\theta}w : \theta \in \Theta) \subset W$$

orbitą obrazu w. Podana definicja orbity jest zależna od wyboru operatorów prolongacji i restrykcji, a nie tylko od samego nieliniowego zniekształcenia T_{θ}. Przydatność tej definicji w dużej mierze będzie zależała od tego wyboru.

Pewne uogólnienie powstaje, gdy dopuszczamy do tego, by prolongacja mogła przyjmować wartości z pewnego rozszerzenia zbioru F, np. interpolacja obrazu czarno-białego może przyjmować poza punktami dyskretnej siatki dowolną wartość z przedziału [0,1]. Z tego uogólnienia będziemy zawsze wtedy korzystać, gdy zbiór kolorów jest dyskretny. Odcinek [0,1] pozostanie zakresem 8-bitowej skali szarości (kontrastu) F', jeśli przeprowadzimy unormowanie $F = \frac{1}{255} F'$.

3. Identyfikacja parametru

Zakładamy obecnie, że w zbiorze W zadana jest pewna miara odległości ρ, lub też rodzina miar ρ_θ,

$$\rho(w_1,w_2) \geq 0,\ \rho(w_1,w_2)=0 \Leftrightarrow w_1 = w_2.$$

Oznaczmy teraz

$$f_{w,w_o}(\theta) = \rho_\theta(T_\theta w,w_o).$$

Zauważmy, że jeśli $w_o=T_{\theta^*}(w)$, to $f_{w,w_o}(\theta)$ jako funkcja θ przyjmuje w punkcie $\theta=\theta^*$ swoje minimum równe zero.

Uzasadnia to uznanie za wiarygodne, że obrazy w oraz w_o^* przedstawiają ten sam obiekt, jeśli

$$\inf_{\theta\in\Theta} f_{w,w_o}(\theta) < \varepsilon$$

dla pewnej liczby ε, którą dobiera się w zależności od w i w_o.

Praktycznie, będziemy stosować pewien algorytm A, który przyporządkowuje trójce argumentów (w,w_o,δ) parę $(\theta,f_{w,w_o}(\theta)) \equiv (A_1(w,w_o,\delta),A_2(w,w_o,\delta))$, przy czym δ jest stałą rządzącą dokładnością obliczenia wartości θ, dla której $f_{w,w_o}(\theta)$ osiąga minimalną wartość. Jeśli $A_2(w,w_o,\delta)$ będzie odpowiednio duże, to odrzucimy hipotezę, że w i w_o przedstawiają identyczne obiekty, w przeciwnym razie zaliczymy w do kandydatów na obiekt do rozpoznania.

Duże znaczenie przypada wyborowi funkcji odległości $\rho(w_1,w_2)$. Przyjmujemy

$$\rho_\theta(w_1,w_2)= \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (w_{1ij}-w_{2ij})^2 m_{ij}(\theta,w_1)$$

z pewnymi wagami m_{ij}. W przypadku $m_{ij}=1$ otrzymujemy zwykłą odległość euklidesową, niezależnie od parametru θ. Wybierając natomiast funkcje wagowe tak, by $m_{ij}(\theta,w)$ było duże w otoczeniu nośnika w, a bliskie zeru poza tym, uzyskujemy możliwość analizy fragmentów.

Zależność m_{ij} od θ umożliwia uwzględnienie faktu, że niektóre fragmenty obrazu w weszły na pole widzenia Π tylko dzięki działaniu deformacji T_θ i powinny być traktowane z większą nieufnością, np.

$$m_{ij}(\theta, w) = [\mathrm{dist}(T_\theta^{-1}(\xi), \mathbb{D}) + 1]^{-1}$$

4. Uwagi końcowe

Przygotowujemy obecnie przykłady numeryczne na podstawie różnych przekształceń geometrycznych T_θ zakładając $T_{\theta_z} = \mathrm{id}$. Należy przy tym zwrócić uwagę na niejednoznaczność przy wyznaczaniu parametru θ oraz na związek między geometrią obrazów a właściwościami funkcji $f_{w,w_o}(\theta)$.

LITERATURA

1. T. Gawlik, M. Nieniewski : *Podstawowe algorytmy bazy danych systemu wizyjnego z kamerą D-CAM.* w : W. Kosiński (Red.), *Komputerowe modelowanie percepcji i rozumowania.* t.1, Prace IPPT 31, 128-144, 1987.
2. S. Matysiak, W. Nagórko, Cz. Woźniak : *O dyskretnej reprezentacji obiektów materialnych w przestrzeniach tolerancyjnych* . tamże str. 145-153.
3. L. Chmielewski : *Robotyka, systemy wizyjne, systemy ekspertowe a sztuczna inteligencja.* tamże str. 6-72.
4. G. E. P. Box and H. L. Lucas : *Design of experiments in non-linear estimations.* Biometrika 46, 77-90, 1959.
5. Z. Semadeni, A. Wiweger, *Wstęp do teorii kategorii i funktorów,* Wyd. 2-gie rozszerzone, PWN, Warszawa, 1978.