

NEURONOWE PROGRAMOWANIE DYNAMICZNE W STEROWANIU BEHAVIORALNYM MOBILNYM ROBOTEM KOŁOWYM

Zenon HENDZEL*, Marcin SZUSTER*

*Katedra Mechaniki Stosowanej i Robotyki, Wydział Budowy Maszyn i Lotnictwa, Politechnika Rzeszowska,
Al. Powstańców Warszawy 8, 35-959 Rzeszów

zenhen@prz.edu.pl, mszuster@prz.edu.pl

Streszczenie: W pracy przedstawiono problematykę generowania bezkolizyjnych trajektorii ruchu mobilnego robota kołowego (MRK) z zastosowaniem algorytmów Neuronowego Programowania Dynamicznego (NDP – Neural Dynamic Programming) oraz układów z logiką rozmytą (FL – Fuzzy Logic). Proponowany hierarchiczny układ sterowania składa się z warstwy generowania trajektorii ruchu, opartej na odruchowej nawigacji MRK w nieznanym środowisku dwuwymiarowym ze statycznymi przeszkodami, oraz warstwy realizacji ruchu. Badania symulacyjne zaproponowanego algorytmu generowania bezkolizyjnych trajektorii oraz realizacji ruchu MRK Pioneer 2-DX przeprowadzono w wirtualnym środowisku obliczeniowym.

1. WPROWADZENIE

Rozwój robotyki mobilnej znacznie poszerzył obszary jej zastosowań, jednocześnie zwiększając poziom złożoności realizowanych zadań, co pociągnęło za sobą konieczność opracowania odpowiednich algorytmów sterowania realizacją ruchu platform mobilnych. Wzrost złożoności konstrukcji MRK, ilości dostępnej informacji o otaczającym środowisku, oraz stopnia zaawansowania jednostek obliczeniowych, otworzyły możliwość projektowania układów sterowania generujących trajektorie ruchu na bieżąco, w czasie ruchu MRK, z możliwością jej modyfikowania pod wpływem warunków środowiska, np. pojawiających się przeszkód. W zagadnieniach tych szerokie zastosowanie znalazły nowoczesne metody sztucznej inteligencji, obejmujące sztuczne sieci neuronowe (SN) oraz układy z logiką rozmytą (Drainkov i Saffiotti, 2001; Hendzel, 2004).

W pracy przedstawiono nowe ujęcie problematyki generowania bezkolizyjnych trajektorii ruchu MRK z zastosowaniem algorytmów NDP. Proponowany hierarchiczny układ sterowania składa się z warstwy generowania trajektorii ruchu, opartej na odruchowej nawigacji MRK w nieznanym środowisku dwuwymiarowym ze statycznymi przeszkodami, oraz warstwy realizacji ruchu. Zagadnienie odruchowej nawigacji MRK, w prezentowanym algorytmie generowania bezkolizyjnych trajektorii ruchu, obejmuje implementację realizacji dwóch elementarnych zachowań: „podążaj do celu” (GS – Goal Seeking) oraz „podążaj środkiem wolnej przestrzeni/omijaj przeszkodę” (OA – Obstacle Avoiding) z zastosowaniem algorytmów NDP w postaci struktury Zależnego od Sterowania Heurystycznego Programowania Dynamicznego (ADHDP - Action Dependent Heuristic Dynamic Programming) (Ferrari i Stengel, 2001; Hendzel, 2007; Hendzel i Szuster, 2009, Powell, 2007; Prokhorov i Lurch, 1997; Sutton i Barto, 1999). Połączenie dwóch elementarnych zachowań (CB – Combination

of Behaviours) odruchowej nawigacji GS oraz OA, w celu realizacji złożonego zadania generowania bezkolizyjnej trajektorii ruchu MRK w nieznanym środowisku, obejmującego podążanie do celu z omijaniem statycznych przeszkód, zrealizowano przy pomocy układu FL. Zadanie generowania sterowania dla ruchu nadążnego, w warstwie realizacji ruchu, zrealizowano z zastosowaniem SN do kompensacji nieliniowości MRK. Badania symulacyjne zaproponowanego algorytmu generowania bezkolizyjnych trajektorii oraz realizacji ruchu MRK Pioneer 2-DX przeprowadzono w wirtualnym środowisku symulacyjnym.

Wyniki badań przedstawione w artykule stanowią kontynuację wcześniejszych prac autorów związanych z opracowaniem układów generowania bezkolizyjnych trajektorii ruchu dla MRK oraz ich realizacji z zastosowaniem nowoczesnych metod sztucznej inteligencji, takich jak SN (Giergiel i inni, 2002), układy FL (Hendzel, 2004), oraz algorytmy oparte na uczeniu ze wzmocnieniem (Hendzel, 2007; Hendzel i Szuster, 2009). Artykuł składa się z sześciu części. Część pierwsza zawiera krótkie wprowadzenie, w części drugiej przedstawiono model MRK Pioneer 2-DX. Część trzecia zawiera syntezę warstwy realizacji ruchu nadążnego, w części czwartej przedstawiono algorytm NDP w postaci struktury ADHDP. Część piąta zawiera syntezę warstwy generowania bezkolizyjnej trajektorii ruchu MRK, część szósta podsumowuje projekt badawczy.

2. MOBILNY ROBOT DWUKOŁOWY

W pracy analizowany jest ruch dwukołowego MRK z trzecim kołem samonastawnym, w płaszczyźnie xy . Schemat MRK Pioneer 2-DX pokazano na Rys. 1.

W modelu matematycznym MRK uwzględniono dynamiczne równania ruchu Maggie’go (Giergiel i inni, 2002; Giergiel i Żylski, 2005) oraz własności dynamiczne układów wykonawczych, zbudowanych z silnika elektrycznego

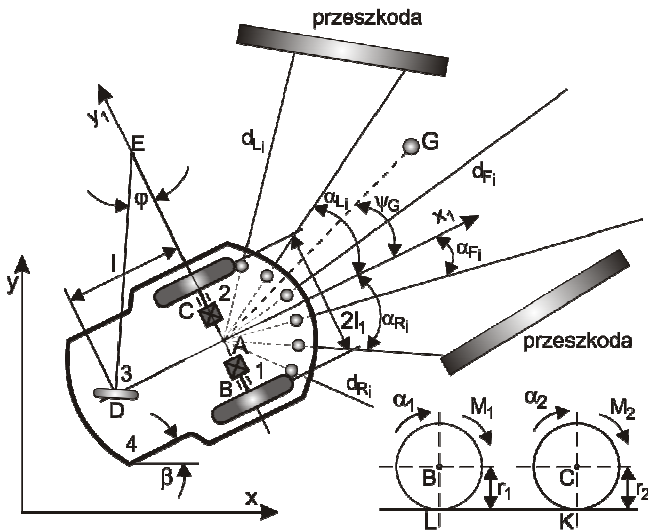
prądu stałego, przekładni oraz enkodera. Przyjęto wektor współrzędnych uogólnionych

$$\mathbf{q} = [\alpha^T, \dot{\alpha}^T]^T, \quad (1)$$

gdzie $\alpha = [\alpha_i]^T$, α_i to kąt obrotu własnego odpowiedniego koła napędowego MRK, $i=1,2$. Dynamiczne równania ruchu MRK przyjmują postać

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}(\dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{F}(\dot{\mathbf{q}}) + \boldsymbol{\tau}_d = \mathbf{u}, \quad (2)$$

gdzie budowa macierzy \mathbf{M} , \mathbf{C} oraz wektora \mathbf{F} wynika z zastosowania dynamicznych równań ruchu Maggie'go do opisu dynamiki MRK, oraz uwzględnienia dynamiki układów wykonawczych, $\boldsymbol{\tau}_d$ jest wektorem ograniczonych zakłóceń, natomiast \mathbf{u} oznacza wektor sterowań.



Rys. 1. Schemat MRK Pioneer 2-DX

3. WARSTWA REALIZACJI RUCHU

W zadaniu sterowania ruchem nadążnym MRK warstwy realizacji ruchu założymy, że znana jest zadana trajektoria ruchu $[\mathbf{q}_d, \dot{\mathbf{q}}_d, \ddot{\mathbf{q}}_d] \in \mathbf{R}^2$. Celem syntezy układu sterowania nadążnego z zastosowaniem SN do kompensacji nieliniowości MRK (Giergiel i inni, 2002), jest taki dobór struktury

sygnału sterowania oraz prawa adaptacji wag SN, dla którego realizowana trajektoria ruchu MRK będzie zbieżna do trajektorii zadanej.

Przyjęto błąd nadążania \mathbf{e} , uogólniony błąd nadążania \mathbf{s} oraz sygnał \mathbf{v} dane zależnościami

$$\mathbf{e} = \mathbf{q}_d - \mathbf{q}, \quad (3)$$

$$\mathbf{s} = \dot{\mathbf{e}} + \boldsymbol{\Lambda}\mathbf{e}, \quad (4)$$

$$\mathbf{v} = \dot{\mathbf{q}}_d + \boldsymbol{\Lambda}\mathbf{e}, \quad (5)$$

gdzie $\boldsymbol{\Lambda}$ to diagonalna macierz projektowa dodatnio określona. Wówczas równanie (2) opisujące dynamikę MRK możemy zapisać w postaci

$$\mathbf{M}\dot{\mathbf{s}} = -\mathbf{u} - \mathbf{C}(\dot{\mathbf{q}})\mathbf{s} + \mathbf{M}\dot{\mathbf{v}} + \mathbf{C}(\dot{\mathbf{q}})\mathbf{v} + \mathbf{F}(\dot{\mathbf{q}}) + \boldsymbol{\tau}_d. \quad (6)$$

Zdefiniowano nieliniową funkcję

$$\mathbf{f}(\mathbf{x}) = \mathbf{M}\dot{\mathbf{v}} + \mathbf{C}(\dot{\mathbf{q}})\mathbf{v} + \mathbf{F}(\dot{\mathbf{q}}), \quad (7)$$

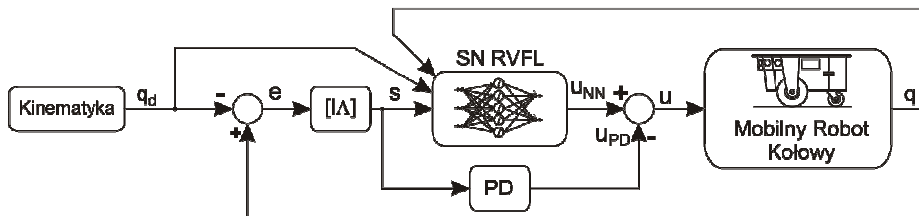
gdzie $\mathbf{x} = [\mathbf{v}^T, \dot{\mathbf{v}}^T, \dot{\mathbf{q}}^T]^T$, oraz przyjęto prawo sterowania z uwzględnieniem kompensacji nieliniowości MRK w postaci zależności

$$\mathbf{u} = \mathbf{f}(\mathbf{x}) + \mathbf{K}_D\mathbf{s}, \quad (8)$$

gdzie \mathbf{K}_D to diagonalna macierz projektowa dodatnio określona.

W praktyce prawo sterowania MRK (8) jest trudne do zrealizowania z powodu braku znajomości dokładnych wartości parametrów MRK oraz możliwości ich zmiany w czasie ruchu np. pod wpływem zmiany przewożonego obciążenia czy zmiany rodzaju nawierzchni, po której MRK się porusza.

W prezentowanym algorytmie realizacji ruchu MRK do aproksymacji nieliniowej funkcji $\mathbf{f}(\mathbf{x})$ zastosowano liniową ze względu na wagi SN Random Vector Functional Link (RVFL) z losowym doбором wag warstwy wejściowej, sigmoidalnymi bipolarnymi funkcjami aktywacji neuronów oraz adaptowanymi wagami warstwy wyjściowej. Na Rys. 2 przedstawiono schemat układu sterowania z zastosowaniem SN do kompensacji nieliniowości MRK.



Rys. 2. Schemat układu sterowania ruchem nadążnym

Estymatę nieliniowej funkcji $\mathbf{f}(\mathbf{x})$ realizowaną przez SN zapiszemy jako

$$\hat{\mathbf{f}}(\mathbf{x}) = \hat{\mathbf{W}}^T \mathbf{S}(\mathbf{V}^T \mathbf{x}), \quad (9)$$

gdzie $\hat{\mathbf{W}}$ to macierz wag warstwy wyjściowej, $\mathbf{S}(\cdot)$ to wek-

tor sigmoidalnych bipolarnych funkcji aktywacji neuronów opisanych zależnością

$$\mathbf{S}(\mathbf{x}) = \frac{2}{1 + \exp(-\beta \mathbf{V}^T \mathbf{x})} - 1, \quad (10)$$

gdzie β odpowiada za pochylenie funkcji sigmoidalnych, natomiast \mathbf{V} to macierz stałych wag warstwy wejściowej dobranych losowo w procesie inicjalizacji sieci.

Prawo sterowania (8) przyjmuje postać

$$\mathbf{u} = \hat{\mathbf{W}}^T \mathbf{S}(\mathbf{V}^T \mathbf{x}) + \mathbf{K}_D \mathbf{s}. \quad (11)$$

Stabilność zamkniętego układu sterowania zagwarantowano, korzystając z teorii stabilności Lapunowa, poprzez odpowiedni dobór prawa adaptacji wag SN

$$\dot{\hat{\mathbf{W}}} = \mathbf{\Gamma} \mathbf{S}(\mathbf{V}^T \mathbf{x}) \mathbf{s}^T, \quad (12)$$

gdzie $\mathbf{\Gamma}$ to stała macierz diagonalna, dodatnio określona, zawierająca współczynniki wzmocnienia uczenia neuronów.

4. NEURONOWE PROGRAMOWANIE DYNAMICZNE

NDP, zwane również Aproksymacyjnym Programowaniem Dynamicznym (ADP – Approximate Dynamic Programming) wywodzi się z tradycyjnej teorii optymalizacji Bellman'a (1957). Zastosowanie nowoczesnych metod sztucznej inteligencji w postaci SN umożliwiło zastosowanie metod opartych na programowaniu dynamicznym w obliczeniach prowadzonych on-line, od pierwszego etapu procesu iteracyjnego do etapu ostatniego (FDP – Forward Dynamic Programming), w przeciwieństwie do podejścia klasycznego (Bacward DP).

NPD obejmuje grupę algorytmów (Powell, 2007; Prokhorov i Wunch, 1997; Sutton i Barto, 1999), w których zakłada się dekompozycję problemu generowania optymalnego prawa sterowania na dwa zagadnienia: generowanie prawa sterowania realizowane przez adaptacyjną strukturę aktora oraz estymację funkcji kosztu realizowaną przez krytyka. Funkcja kosztu V w przypadku ogólnym zależy od stanu obiektu x_k oraz generowanego sterowania u_k w k -tym kroku procesu dyskretnego, i może być zapisana w postaci zależności

$$V(x_k, u_k) = \sum_{k=0}^n \gamma^k L_C(x_k, u_k), \quad (13)$$

gdzie n jest liczbą kroków procesu, γ jest tzw. współczynnikiem zapominania, który przyjmuje wartości z zakresu $0 < \gamma < 1$, $L_C(x_k, u_k)$ jest lokalnym kosztem w k -tym kroku.

W projektowanych układach przyjęto lokalny koszt $L_C(x_k, u_k)$ w ogólnej postaci jako

$$L_{Ci}(\mathbf{x}_{ik}, \mathbf{u}_{ik}) = \frac{1}{2} \{ \mathbf{e}_{ik}^T \mathbf{P}_{ii} \mathbf{e}_{ik} \} + \frac{1}{2} \{ \mathbf{u}_{ik}^T \mathbf{R}_{ii} \mathbf{u}_{ik} \} \quad (14)$$

gdzie $i=1,2$, \mathbf{P} oraz \mathbf{R} to dodatnio określone, stałe macierze diagonalne określające wpływ błędów oraz generowanych sterowań na wartość funkcji kosztu. Zadaniem algorytmu NDP jest wygenerowanie prawa sterowania które minimalizuje funkcję kosztu (13) dla lokalnego kosztu (14).

W prezentowanym hierarchicznym układzie sterowania algorytmy NDP w konfiguracji ADHDP zastosowano w warstwie generowania trajektorii ruchu do realizacji ele-

mentarnych zachowań „podażania do celu” (GS) oraz „podażania środkiem wolnej przestrzeni” (OA). Algorytm ADHDP jako jedyny z rodziny algorytmów NDP nie wymaga znajomości modelu sterowanego obiektu w procedurach uczenia struktur adaptacyjnych. Algorytm składa się z:

- krytyka - zrealizowanego w postaci SN, którego zadaniem jest estymacja funkcji kosztu

$$\hat{\mathbf{V}}_{ik}(\mathbf{x}_{Cik}, \hat{\mathbf{W}}_{Cik}) = \hat{\mathbf{W}}_{Cik}^T \mathbf{S}(\mathbf{x}_{Cik}), \quad (15)$$

gdzie $i=1,2$, $\hat{\mathbf{W}}_C$ to macierz wag warstwy wyjściowej, $\mathbf{S}(\mathbf{x}_C)$ to wektor funkcji aktywacji neuronów SN, \mathbf{x}_{Ci} - wektor wejść do i -tej sieci.

Wagi SN krytyka są adaptowane z zastosowaniem metody wstecznej propagacji błędu poprzez minimalizację błędu dyskretnych różnic czasowych (TD - Temporal Difference error)

$$\mathbf{e}_{TDk} = \mathbf{L}_C(\mathbf{x}_k, \mathbf{u}_k) + \hat{\mathbf{V}}(\mathbf{x}_{Ck+1}, \hat{\mathbf{W}}_{Ck+1}) - \hat{\mathbf{V}}(\mathbf{x}_{Ck}, \hat{\mathbf{W}}_{Ck}), \quad (16)$$

- aktora - zrealizowanego w postaci SN, którego zadaniem jest estymacja optymalnego prawa sterowania

$$\mathbf{u}_{Aik}(\mathbf{x}_{Aik}, \hat{\mathbf{W}}_{Aik}) = \hat{\mathbf{W}}_{Aik}^T \mathbf{S}(\mathbf{x}_{Aik}), \quad (17)$$

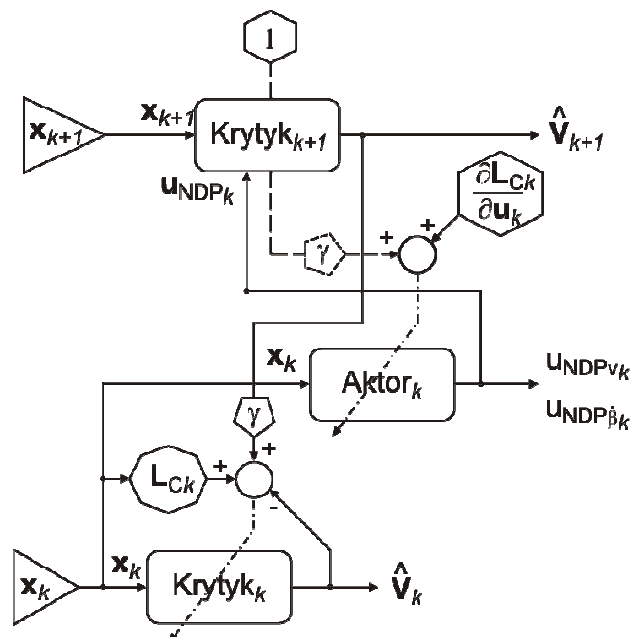
gdzie $i=1,2$, $\hat{\mathbf{W}}_A$ to macierz wag warstwy wyjściowej, $\mathbf{S}(\mathbf{x}_A)$ to wektor funkcji aktywacji neuronów SN, \mathbf{x}_{Ai} - wektor wejść do i -tej sieci aktora.

Wagi SN aktora są adaptowane z zastosowaniem metody wstecznej propagacji błędu poprzez minimalizację wskaźnika jakości

$$\mathbf{e}_{Ak} = \frac{\partial \mathbf{L}_C(\mathbf{x}_k)}{\partial \mathbf{u}_k} \mathbf{J}_{2 \times 1} + \gamma \frac{\partial \hat{\mathbf{V}}(\mathbf{x}_{Ck+1}, \hat{\mathbf{W}}_{Ck})}{\partial \mathbf{u}_k} \mathbf{J}_{2 \times 1}, \quad (18)$$

gdzie \mathbf{J} to wektor jedynekowy.

Schematycznie strukturę ADHDP przedstawiono na Rys. 3.

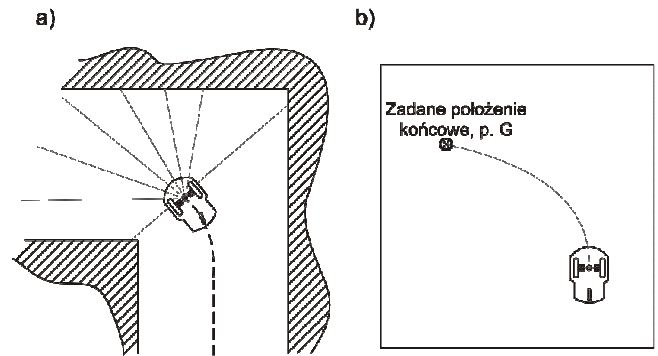


Rys. 3. Schemat struktury ADHDP

Struktura ADHDP składa się z SN aktora w kroku k -tym, oraz SN krytyka w kroku k -tym oraz $k+1$.

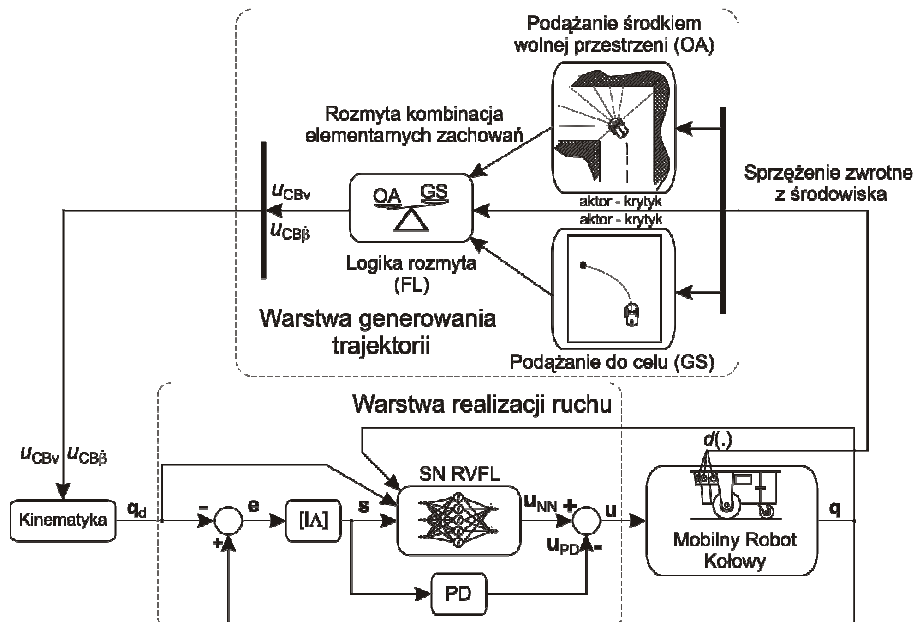
5. WARSTWA PLANOWANIA TRAJEKTORII RUCHU

W proponowanym hierarchicznym układzie sterowania warstwa generowania bezkolizyjnej trajektorii ruchu w nieznanym środowisku składa się z dwóch algorytmów realizujący elementarne zachowania nawigacji odruchowej: „podążania środkiem wolnej przestrzeni” (OA) i „podążania do celu” (GS), oraz algorytmu łączącego elementarne zachowania (CB) w celu realizacji złożonych zadań. Algorytmy realizujące elementarne zachowania nawigacji odruchowej zaimplementowano w postaci struktur NDP w konfiguracji ADHDP, natomiast algorytm służący do łączenia tych zachowań zbudowano w oparciu o układ z logiką rozmytą. Schematycznie zagadnienia realizacji elementarnych zachowań nawigacji odruchowej OA oraz GS przedstawiono na Rys. 4.



Rys. 4. a) Schemat realizacji elementarnego zachowania „podążania środkiem wolnej przestrzeni” (OA), b) schemat realizacji zachowania „podążania do celu” (GS)

Schemat proponowanego hierarchicznego układu sterowania przedstawiono na Rys. 5.



Rys. 5. Schemat hierarchicznego układu sterowania z warstwą generowania trajektorii ruchu

W nieruchomym globalnym układzie odniesienia (x, y) położenie MRK opisują współrzędne zdefiniowane jako $[x_A, y_A, \beta]^T$, gdzie współrzędne wybranego punktu A wynoszą (x_A, y_A) , natomiast β to kąt obrotu ramy MRK. Kinematykę robota opisuje równanie

$$\begin{bmatrix} \dot{x}_A \\ \dot{y}_A \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} v_A^* \cos(\beta) & 0 \\ v_A^* \sin(\beta) & 0 \\ 0 & \omega^* \end{bmatrix} \begin{bmatrix} u_{CBv} \\ u_{CB\beta} \end{bmatrix}, \quad (19)$$

gdzie v_A^* to maksymalna prędkość liniowa ruchu punktu A, ω^* to maksymalna prędkość kątowna obrotu własnego ramy mobilnego robota. W przypadku realizacji jednego z zachowań behawioralnych wektor u_{CB} jest zastępowany przez u_{GS} lub u_{OA} . MRK Pioneer 2-DX jest wyposażony

w 8 sonarów ultradźwiękowych (s_1, \dots, s_8) , dla których maksymalny zasięg pomiaru wynosi $d_{s_{max}}=6$ [m]. Sensory podzielono na grupy składające się z dwóch sąsiednich sensorów (odrzucając pierwszy oraz ostatni), rejestrują one odległości od przeszkód d_{L1}, d_{L2} dla grupy czujników s_2 oraz s_3 znajdujących się po lewej stronie ramy MRK, d_{F1}, d_{F2} dla grupy czujników znajdujących się z przodu (s_4, s_5), oraz d_{R1}, d_{R2} dla grupy czujników znajdujących się z lewej strony ramy (s_6, s_7). Wartość pomiarów z sensorów ultradźwiękowych mieści się w przedziale $d_{s_{min}} \leq d(.) \leq d_{s_{max}}$, przy czym $d_{s_{min}}=0.5$ [m]. Kąty odchylenia osi pomiarów czujników względem osi ramy MRK wynoszą odpowiednio $\alpha_{Li}, \alpha_{Fi}, \alpha_{Ri}$, dla $i=1,2$.

Zadaniem warstwy generowania trajektorii hierarchicznego układu sterowania jest wygenerowanie w czasie rzeczywistym współrzędnych wybranego punktu A ramy

MRK, zdefiniowanych jako $\mathbf{x}_d=[x_A, y_A, \beta]^T$. Współrzędne te są generowane przez jedną ze struktur warstwy generowania trajektorii w zależności od typu realizowanego zadania, na podstawie informacji ze środowiska w postaci pomiarów z sonarów ultradźwiękowych $d(\cdot)$. Rozważane są dwa elementarne zachowania zrealizowane przy pomocy algorytmów NDP: „podążaj środkiem wolnej przestrzeni” (OA) i „podążaj do celu” (GS), w których zadana trajektoria ruchu generowana jest poprzez zadanie odpowiednich wartości wektora $\mathbf{u}_{OA}=[u_{OA\dot{v}}, u_{OA\dot{\beta}}]^T$ lub $\mathbf{u}_{GS}=[u_{GSv}, u_{GS\beta}]^T$.

Rozpatruje się również kombinację powyższych elementarnych zachowań zrealizowaną przy pomocy układu FL, w celu realizacji złożonego zadania podążania do celu z omijaniem przeszkód, w którym trajektoria zadana generowana jest poprzez zadanie odpowiednich wartości wektora $\mathbf{u}_{CB}=[u_{CBv}, u_{CB\dot{\beta}}]^T$.

Sygnaly wejściowe do struktur warstwy generowania trajektorii ruchu układu sterowania znormalizowano do przedziału $\langle -1, 1 \rangle$ lub $\langle 0, 1 \rangle$ poprzez odpowiednią transformację: dla pomiaru odległości przez prawą grupę czujników $d_R^*=d_R/(d_L+d_R)$, przez lewą grupę czujników $d_L^*=d_L/(d_L+d_R)$, oraz przez przednią grupę czujników $d_F^*=d_F/d_{smax}$, gdzie $d_F=\min(d_{F1}, d_{F2})$, $d_L=\min(d_{L1}, d_{L2})$, $d_R=\min(d_{R1}, d_{R2})$.

5.1. Zadanie „podążania środkiem wolnej przestrzeni” (OA)

Elementarne zachowanie „podążania środkiem wolnej przestrzeni” polega na wyborze trajektorii ruchu MRK, która maksymalizuje odległość od przeszkód.

Zdefiniujemy błąd prędkości $e_{OA\dot{v}}$ oraz błąd podążania środkiem wolnej przestrzeni $e_{OA\dot{\beta}}$

$$e_{OA\dot{v}} = d_F^* - v_A/v_A^*, \quad (20)$$

$$e_{OA\dot{\beta}} = d_R^* - d_L^*, \quad (21)$$

gdzie v_A to realizowana prędkość ruchu p. A, $v_A^*=0.6$ [m/s] to maksymalna liniowa prędkość ruchu MRK.

Przyjęto funkcję kosztu lokalnego w postaci zależności

$$L_{Ci}(\mathbf{e}_{OAik}, \mathbf{u}_{OAik}) = \frac{1}{2} \{ \mathbf{e}_{OAik} \mathbf{P}_{OAii} \mathbf{e}_{OAik} \} + \frac{1}{2} \{ \mathbf{u}_{OAik} \mathbf{R}_{OAii} \mathbf{u}_{OAik} \} \quad (22)$$

gdzie $i=1,2$, $\mathbf{e}_{OA}=[e_{OA\dot{v}}, e_{OA\dot{\beta}}]^T$, \mathbf{P}_{OA} , \mathbf{R}_{OA} to diagonalne macierze projektowe dodatkowo określone.

Algorytm $ADHDP_{OA}$ składa się z dwóch struktur aktor-krytyk, z których każda zawiera SN aktora oraz krytyka. Sygnał sterowania generowany przez strukturę $ADHDP_{OA}$, $\mathbf{u}_{OANDP}=[u_{OA\dot{v}NDP}, u_{OA\dot{\beta}NDP}]^T$, poszerzono o sygnał regulatora proporcjonalnego, generującego sterowanie

$$\mathbf{u}_{OAPk} = \mathbf{K}_{OA} \mathbf{e}_{OAK}, \quad (23)$$

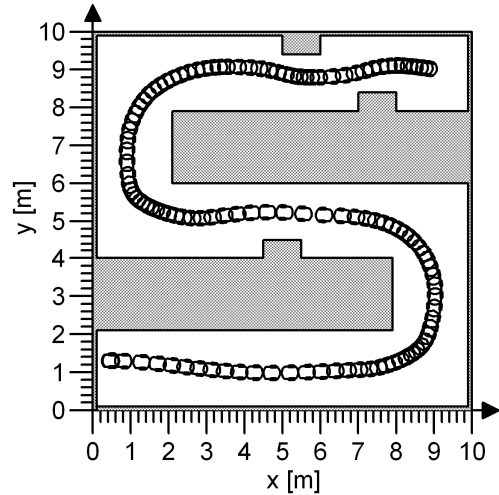
gdzie \mathbf{K}_{OA} to stała diagonalna macierz projektowa wzmoc-

nień regulatora proporcjonalnego.

Całkowity sygnał sterowania opisuje zależność

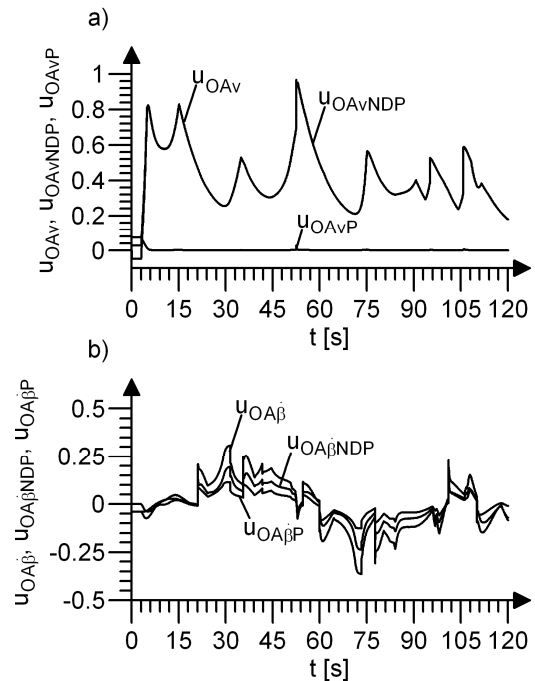
$$\mathbf{u}_{OAK} = \mathbf{u}_{OANDPk} + \mathbf{u}_{OAPk}. \quad (24)$$

Dla układu sterowania złożonego z warstwy generowania trajektorii ruchu dla zadania OA, oraz warstwy realizacji ruchu, w środowisku obliczeniowym przeprowadzono symulację ruchu MRK Pioneer 2-DX w zadanej przestrzeni roboczej, przedstawionej na Rys. 6.



Rys. 6. Realizacja elementarnego zadania podążania środkiem wolnej przestrzeni (OA) przez MRK

Sygnaly sterowania wygenerowane przez warstwę generowania trajektorii przedstawiono na Rys. 7.



Rys. 7. a) Przebieg wartości sygnałów sterowania $u_{OA\dot{v}}$, $u_{OA\dot{v}NDP}$ oraz $u_{OA\dot{v}P}$, b) przebieg wartości sygnałów sterowania $u_{OA\dot{\beta}}$, $u_{OA\dot{\beta}NDP}$, $u_{OA\dot{\beta}P}$

5.2. Zadanie „podążania do celu” (GS)

Elementarne zachowanie „podążania do celu” polega na wygenerowaniu trajektorii ruchu MRK, która gwarantuje osiągnięcie przez wybrany punkt A MRK zadanego położenia (p. G). Zadanie to może być widziane jako zagadnienie minimalizacji odległości $l_G = \|A, G\|$ oraz kąta ψ_G zdefiniowanego jako odchylenie osi ramy MRK od prostej przechodzącej przez p. A i G. Zadanie wyznaczenia sterowań $\mathbf{u}_{GS} = [u_{GSv}, u_{GS\dot{\beta}}]^T$, gwarantujących realizację zachowania GS, zrealizowano przy pomocy struktury NDP w postaci algorytmu ADHDP_{GS}.

Zdefiniujmy błąd prędkości e_{GSv} oraz błąd odchylenia osi ramy MRK od kierunku prostej przechodzącej przez p. A i G, $e_{GS\dot{\beta}}$

$$e_{GSv} = v_{Ad}(l_G) - v_A, \quad (25)$$

$$e_{GS\dot{\beta}} = \psi_G - \beta, \quad (26)$$

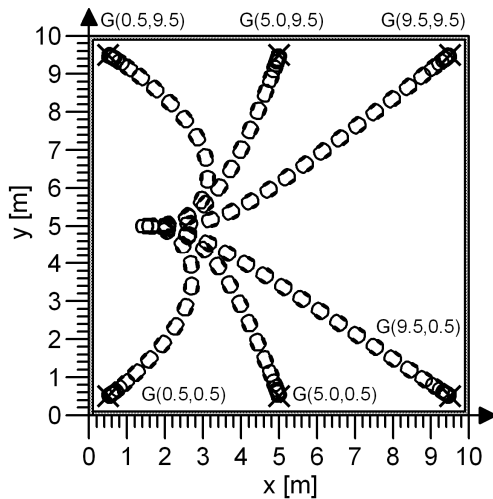
gdzie

$$v_{Ad}(l_G) = \frac{2}{1 + \exp(-l_G)} - 1. \quad (27)$$

Przyjęto funkcję kosztu lokalnego dla struktury NDP w postaci zależności

$$L_{Ci}(\mathbf{e}_{GSik}, \mathbf{u}_{GSik}) = \frac{1}{2} \{ \mathbf{e}_{GSik} \mathbf{P}_{GSii} \mathbf{e}_{GSik} \} + \frac{1}{2} \{ \mathbf{u}_{GSik} \mathbf{R}_{GSii} \mathbf{u}_{GSik} \} \quad (28)$$

gdzie $i=1,2$, $\mathbf{e}_{GS} = [e_{GSv}, e_{GS\dot{\beta}}]^T$, \mathbf{P}_{GS} , \mathbf{R}_{GS} to diagonalne macierze projektowe dodatnio określone.



Rys. 8. Realizacja elementarnego zadania podążania środkiem wolnej przestrzeni (OA) przez MRK

Algorytm ADHDP_{GS} składa się z dwóch struktur aktor-krytyk, z których każda zawiera SN aktora oraz krytyka. Sygnał sterowania generowany przez strukturę ADHDP_{GS}, $\mathbf{u}_{GSNDP} = [u_{GSvNDP}, u_{GS\dot{\beta}NDP}]^T$, poszerzono o sterowanie generowane przez regulator proporcjonalny

$$\mathbf{u}_{GSPk} = \mathbf{K}_{GS} \mathbf{e}_{GSk}, \quad (29)$$

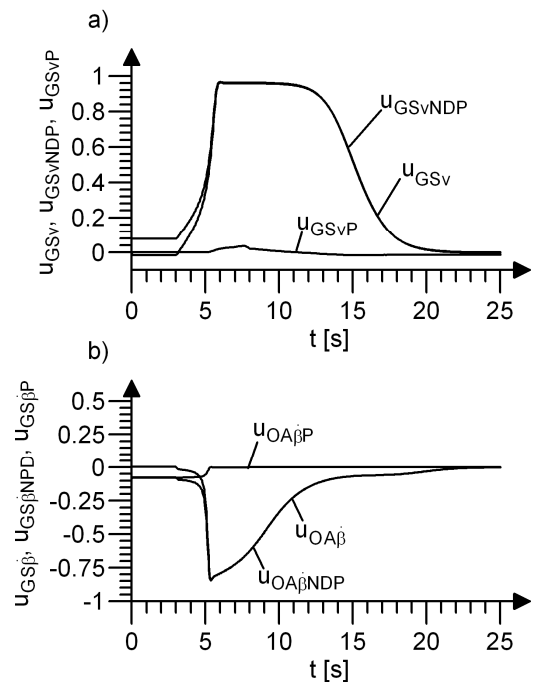
gdzie \mathbf{K}_{GS} to stała diagonalna macierz projektowa wzmocnień regulatora proporcjonalnego.

Całkowity sygnał sterowania w zadaniu GS opisuje zależność

$$\mathbf{u}_{GSk} = \mathbf{u}_{GSNDPk} + \mathbf{u}_{GSPk}. \quad (30)$$

Dla zadania GS, przeprowadzono symulację ruchu MRK Pioneer 2-DX w zadanej przestrzeni roboczej. Na Rys. 8 przedstawiono tor ruchu p. A MRK dla realizacji zadania GS dla wybranych współrzędnych p. G.

Sygnały sterowania wygenerowane przez warstwę generowania trajektorii ruchu MRK dla elementarnego zachowania GS i punktu G o współrzędnych G(0.5;0.5), przedstawiono na Rys. 9.



Rys. 9. a) Przebieg wartości sygnałów sterowania u_{GSv} , u_{GSvNDP} oraz u_{GSvP} , b) przebieg wartości sygnałów sterowania $u_{GS\dot{\beta}}$, $u_{GS\dot{\beta}NDP}$, $u_{GS\dot{\beta}P}$

5.3. Rozmyta kombinacja elementarnych zachowań (CB)

Zadaniem rozmytego układu nawigacji jest dobór współczynnika skalowania zachowań behawioralnych OA oraz GS, który zagwarantuje realizację złożonego zadania podążania do celu z omijaniem statycznych przeszkód w nieznanym środowisku dwuwymiarowym. W prezentowanym układzie rozmytej nawigacji, służącym do koordynacji elementarnych zachowań (CB), funkcje przynależności do zbiorów rozmytych zostały dostrojone metodą prób i błędów. Zastosowano model rozmyty Mamdaniego, z rozmytą implikacją typu MIN, trójkątnymi lub trapezowymi funkcjami przynależności do zbiorów rozmytych przesłanek i konkluzji, oraz metodą wyostrażania typu średnia z maksimum (Center Average Defuzzyfication).

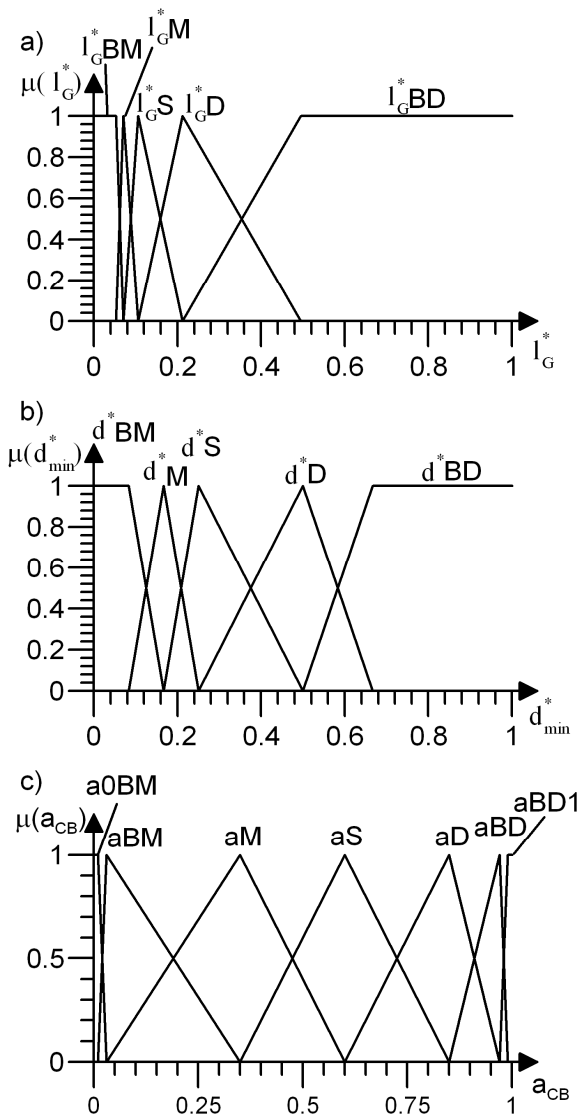
Układ z logiką rozmytą CB posiada bazę $m=25$ reguł o postaci

$$R_{CB}^m : JEZELI (l_G^* \text{ jest IBM}) I (d_{min}^* \text{ jest dBM}) TO (a_{CB} \text{ jest aBD1}), \quad (31)$$

gdzie l_G^* to znormalizowana odległość od celu, d_{min}^* to znormalizowana minimalna odległość od przeszkód, $a_{CB} \in \langle 0,1 \rangle$ jest współczynnikiem skalowania poszczególnych zachowań elementarnych OA oraz GS zgodnie z zależnością

$$\mathbf{u}_{CB} = \mathbf{u}_{OA} \cdot (1 - a_{CB}) + \mathbf{u}_{GS} \cdot a_{CB}. \quad (32)$$

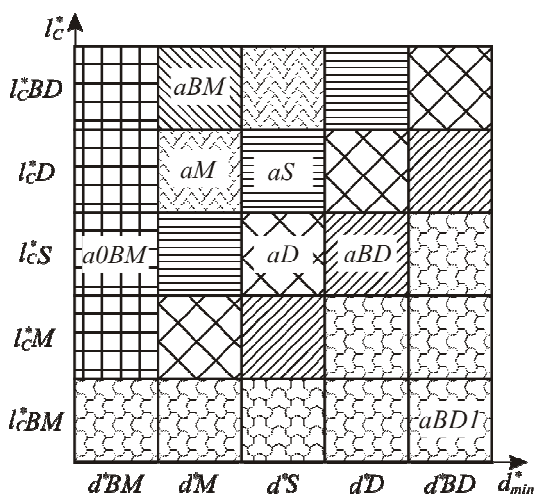
Zastosowanie bazy 25 reguł (31) jest związane z podziałem przestrzeni znormalizowanej odległości od celu l_G^* na pięć zbiorów rozmytych trójkątnych oraz trapezowych: „bardzo mała” (l_G^*BM), „mała” (l_G^*M), „średnia” (l_G^*S), „duża” (l_G^*D), „bardzo duża” (l_G^*BD). Funkcje przynależności do odpowiednich zbiorów rozmytych pokazano na Rys. 10.



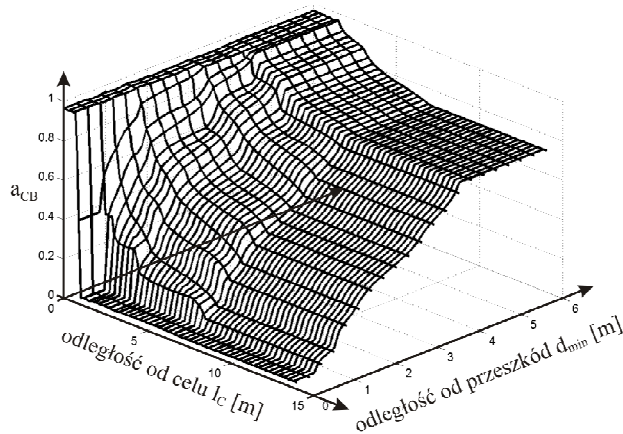
Rys. 10. a) Rozkład funkcji przynależności do zbiorów rozmytych przesłanek dla odległości od celu l_G^* , b) rozkład funkcji przynależności do zbiorów rozmytych przesłanek dla minimalnej odległości od przeszkód d_{min}^* , c) rozkład funkcji przynależności do zbiorów rozmytych konkluzji dla współczynnika skalowania a_{CB}

Przestrzeń znormalizowanej minimalnej odległości od przeszkód d_{min}^* podzielono na pięć zbiorów rozmytych trójkątnych oraz trapezowych: „bardzo mała” (d_{min}^*BM), „mała” (d_{min}^*M), „średnia” (d_{min}^*S), „duża” (d_{min}^*D), „bardzo duża” (d_{min}^*BD). Do podziału przestrzeni wyjścia z układu rozmytego - współczynnika skalowania elementarnych zachowań a_{CB} zastosowano siedem funkcji przynależności do zbiorów rozmytych trójkątnych oraz trapezowych, oznaczonych etykietami lingwistycznymi: „bardzo mały, bliski zera” (a_{0BM}), „bardzo mały” (a_{BM}), „mały” (a_M), „średni” (a_S), „duży” (a_D), „bardzo duży” (a_{BD}), „bardzo duży, bliski jeden” (a_{BD1}).

Ilustrację graficzną bazy reguł pokazano na Rys. 11. Dla bazy 25 reguł rozmytego układu koordynacji elementarnych zachowań CB w warstwie generowania trajektorii ruchu hierarchicznego układu sterowań, wyznaczono powierzchnię modelu $a_{CB}(l_G, d_{min})$, pokazaną na Rys. 12.



Rys. 11. Ilustracja bazy 25 reguł

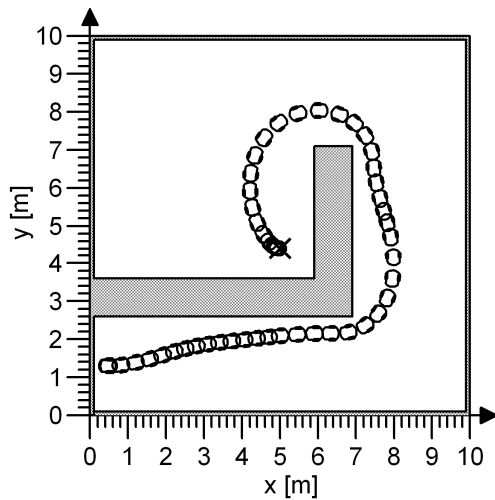


Rys. 12. Graficzna interpretacja powierzchni modelu rozmytego $a_{CB}(l_G, d_{min})$

Kształt powierzchni modelu rozmytego związany z doborem bazy reguł układu FL można interpretować jako wybór:

- elementarnego zachowania OA gdy odległość od przeszkód jest bardzo mała, bez względu na odległość od celu (poza sytuacją, gdy odległość do celu jest bardzo mała),

- elementarnego zachowania GS, gdy odległość od celu jest bardzo mała,
- płynne przełączanie zachowań OA i GS w zależności od odległości od celu oraz przeszkód.

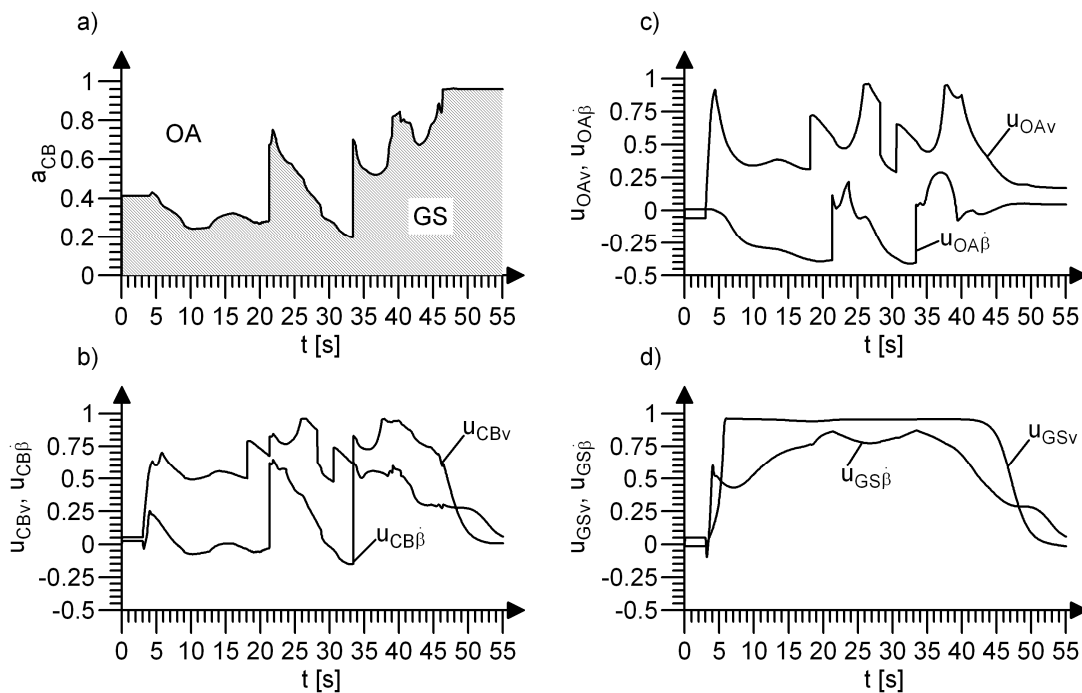


Rys. 13. Realizacja rozmytego połączenia dwóch elementarnych zachowań: GS oraz OA przez MRK

Zadanie generowania bezkolizyjnej trajektorii ruchu MRK w nieznanym środowisku z punktu startowego $A_0(0.5, 1.3)$ do punktu $G(5.0, 4.4)$, oraz realizację zadanej trajektorii przez zaprojektowany hierarchiczny układ sterowania przedstawiono na Rys. 13.

Przebieg sygnału a_{CB} skalującego sygnały sterowania zachowań behawioralnych w czasie ruchu MRK przedstawiono na Rys. 14.a). W początkowej fazie ruchu, gdy odległość od celu jest duża i odległość od przeszkód mała, dominujący udział w całkowitym sterowaniu warstwy generowania trajektorii u_{CB} (Rys. 14.b) ma sygnał generowany przez zachowanie OA u_{OA} (Rys. 14.c), jednakże gdy odległość do celu zmniejsza się, oraz odległość od przeszkód jest większa, decydującą rolę w sterowaniu ma zachowanie GS u_{GS} (Rys. 14.d).

Dobór odpowiedniej bazy reguł oraz funkcji przynależności do zbiorów rozmytych wejść układu FL pozwoliły na gładkie przełączenie realizacji elementarnych zachowań OA i GS podczas ruchu MRK w nieznanym środowisku, oraz realizację złożonego zadania podążania do celu z ominięciem przeszkód.



Rys. 14. a) Współczynnik skalowania elementarnych zachowań a_{CB} , b) przebiegi wartości sygnałów u_{CBv} oraz $u_{CB\beta}$, c) przebiegi wartości sygnałów $u_{OA\alpha}$ oraz $u_{OA\beta}$, d) przebiegi wartości sygnałów $u_{GS\alpha}$ oraz $u_{GS\beta}$

6. PODSUMOWANIE

W pracy zaproponowano zastosowanie hierarchicznego układu sterowania, składającego się z warstwy generowania bezkolizyjnej trajektorii ruchu w nieznanym otoczeniu ze statycznymi przeszkodami, oraz warstwy realizacji ruchu. Warstwa realizacji ruchu realizuje zadanie generowania sterowań w ruchu nadążnym, dla trajektorii zadanej wygenerowanej przez wyższą warstwę układu sterowania.

Zadanie generowania bezkolizyjnej trajektorii ruchu w nieznanym otoczeniu zostało dekomponowane na dwa elementarne zachowania: „omijania przeszkód” oraz „podążania do celu”, zrealizowane w postaci struktur NDP z zastosowaniem nowoczesnych metod sztucznej inteligencji. Każde z elementarnych zachowań jest realizowane poprawnie, pozwalając na generowanie bezkolizyjnej trajektorii ruchu (OA) lub trajektorii gwarantującej osiągnięcie założonego celu (GS).

Kombinację dwóch elementarnych zachowań zrealizowano przy pomocy układu z logiką rozmytą w oparciu o wiedzę eksperta, co pozwoliło na płynne przełączanie realizacji poszczególnych zachowań elementarnych w zależności od sygnału ze środowiska. Badania symulacyjne przeprowadzono na emulatorze MRK Pioneer 2-DX.

Analizowany hierarchiczny układ sterowania składający się z warstwy generowania bezkolizyjnej trajektorii ruchu oraz realizacji ruchu MRK pracuje poprawnie realizując założone zadania w nieznanym otoczeniu ze statycznymi przeszkodami. Analizowane rozwiązania zostaną zastosowane w nawigacji MRK w czasie rzeczywistym.

LITERATURA

1. **Bellman R.**, (1957), *Dynamic programming*, Princeton University Press, New York.
2. **Drainkov D., Saffiotti A.** (2001), *Fuzzy logic techniques for autonomous vehicle navigation*, Springer-Verlag, New York.
3. **Ferrari S., Stengel R.F.** (2004), Model-based adaptive critic design, W: *Handbook of Learning and Approximate Dynamic Programming*, Willey - IEEE Press, New York, 65-95.
4. **Giergiel J., Żylski W.** (2005), Description of motion of a mobile robot by Maggie's Equations, *J. Theor. and App. Mech.*, Vol. 43, No 3, 511–521.
5. **Giergiel M., Hendzel Z., Żylski W.** (2002), *Modelowanie i sterowanie mobilnych robotów kołowych*, PWN, Warszawa.
6. **Hendzel Z.** (2004), Fuzzy reactive control of wheeled mobile robot, *J. Theor. App. Mech.*, Vol. 42, No 3, 503-517.
7. **Hendzel Z.** (2007), An adaptive critic neural network for motion control of a wheeled mobile robot, *Nonlin. Dyn.*, Vol. 50, No 4, 849-855.
8. **Hendzel Z., Szuster M.** (2009), Discrete action dependant heuristic dynamic programming in wheeled mobile robot control, W: *V International Conf. on Mechatronic Systems and Materials 2009*, Wilno, 114-119.
9. **Powell W.B.** (2007), *Approximate dynamic programming: Solving the curses of dimensionality*, Willey-Interscience, Princeton.
10. **Prokhorov D., Wunch D.C.** (1997), Adaptive critic designs, *IEEE Trans. on Neural Net.*, Vol. 8, No 5, 997-1007.
11. **Sutton R.S., Barto A.G.** (1999), *Reinforcement learning: an introduction*, MIT Press, Cambridge, 1999.

NEURAL DYNAMIC PROGRAMMING IN BEHAVIOURAL CONTROL OF A WHEELED MOBILE ROBOT

Abstract: In this paper an innovative approach to a collision-free trajectory generating for a wheeled mobile robot (WMR) with Neural Dynamic Programming (NDP) and Fuzzy Logic (FL) algorithms, is proposed. The presented hierarchical control system consists of a trajectory generating algorithm based on reactive navigation of the WMR in unknown 2D environment with static obstacles, and a tracking control system. A strategy of reactive navigation is developed including two main behaviours: obstacle avoiding behaviour (OA) and goal-seeking behaviour (GS) realized in a form of NDP algorithms. These simple, individual behaviours are combined by the fuzzy combiner of behaviours (CB), that determines influence of the individual behaviours according to environment conditions. A computer simulations of the proposed control algorithm have been conducted in virtual environment for the WMR Pioneer 2-DX.