



Národní konference s mezinárodní účastí

INŽENÝRSKÁ MECHANIKA 2002

13. – 16. 5. 2002, Svratka, Česká republika

LEARNING BASED CONTROL SYSTEM OF FOUR-LEGGED ROBOT

Tomáš Březina¹, Pavel Houška², Vladislav Singule³

Abstract: Possible discretization technique of the continuous state space of four-legged robot using simultaneous compositions of behaviors is described in this contribution. Compositions are generated by the instances of two basic controllers. The aim is to automatically develop the gait policy. Possible composition strategies are implemented through undeterministic state machine. In the machine design stage the number of both machine states and transitions could be essentially reduced using formal considerations only. Furthermore machine transitions are evaluated by selected alternative of Q-learning to incrementally improve the estimation of the future benefit from single transitions. This technique results in simultaneous activations of controller compositions instances with the highest estimation of the future benefit.

Keywords: Control Composition, Q-learning, Walking of Robot

1 Úvod

Základním cílem výzkumu kráčivých robotů je realizace autonomního stroje, který bude schopen pohybu v neznámém neregulárním terénu. Základním předpokladem autonomnosti je adaptabilita řídicího systému.

V literatuře je vymezena řada problémů významných pro stavbu a řízení kráčivých robotů. Jde především o stabilitu, univerzálnost a výpočetní složitost řízení, efektivnost plánování a interpretaci senzorických dat. Řešení se pohybují v mnoha oblastech, zejména v oblasti konstrukce nohy [1, 2] a v různých způsobech řízení včetně plánování (použití řízení odvozeného od chování [3, 4, 5], hybridního a reaktivního řízení [1, 6, 7, 8], dynamického řízení [9, 10, 11, 4], silového řízení [12, 13, 14], geometrického plánování [15, 16, 17, 18], atd.).

Využití opakovaně posilovaného učení slibuje dosažení potřebné adaptability řídicího systému bez potřeby modelování složitých a nepředvídatelných případů chování systému. Obecný model opakovaně posilovaného učení pracuje s agentem umístěným v prostředí. Agent je propojen s prostředím přes vjemy a akce v diskrétních časových krocích. Prostor je představováno diskrétními stavy a agent disponuje diskrétními akcemi. Vjemy agenta jsou tvořeny aktuálním stavem prostředí a okamžitou odměnou nebo pokutou vztahenou k tomuto stavu. Odměna či pokuta je generována posilovací funkcí. Na základě aktuálního stavu agent provede akci, která obecně způsobí přechod stavu prostředí do nového stavu. Úkolem agenta je naučit se chovat v daném prostředí optimálně. Důležitým předpokladem pro přijatelnou rychlost a úspěšnost učení je minimalizace počtu diskrétních stavů prostředí. Z různých paradigmat opakovaně posilovaného učení je velmi atraktivní Q-učení [19]. Pro Q-učení je teoreticky zaručeno dosažení v jistém smyslu optimálního chování (výsledkem učení je generování posloupnosti akcí měnících optimálně stav prostředí) při dostatečně velkém (teoreticky nekonečném) množství průchodů procedury všemi

¹ Tomáš Březina, Ing. RNDr. CSc., ÚAI FSI VUT Brno, Technická 2, 616 69 Brno, ČR,
e-mail: brezina@uai.fme.vutbr.cz

² Pavel Houška, Ing., ÚMT FSI VUT Brno, Technická 2, 616 69 Brno, ČR,
e-mail: houska@umt.fme.vutbr.cz

³ Vladislav Singule, Doc. Ing. CSc., ÚVSSaR FSI VUT Brno, Technická 2, 616 69 Brno, ČR,
e-mail: singule@zam.fme.vutbr.cz

dvojicemi stav/akce. Pro efektivní a úspěšné použití Q-učení tak získává zásadní význam vhodná diskretizace stavového prostoru a vhodná volba množiny akcí. Rovněž je významné, aby obě množiny byly co nejmenší. Problém je studován na bezdráhovém řízení čtyřnohého robotu v rovinném terénu.

2 Použitý přístup

Přístup uvedený v tomto příspěvku ukazuje na možnost řešení problému prudkého nárůstu alternativ chování redukcí množství kompozic chování beze ztráty složitosti výsledného chování. Používá malé množiny základních řídicích členů, které nejsou přímo závislé na úloze (řídicí báze) s jednoduchými cíly činnosti, které jsou schopny pokrýt rozsáhlou třídu úloh. Instance základních řídicích členů jsou tvořeny těmito řídicími členy s přiřazenými senzory (vstupními zdroji) a stupni volnosti robotu (výstupními zdroji).

Výsledné chování je získáváno simultánní aktivací různých kombinací instancí základních řídicích členů na základě kompoziční strategie, která již na úloze přímo závislá je. Výsledky aktivace instancí základních řídicích členů (tj. toho, zda bylo či nebylo dosaženo jejich cílů) definují stav robotu.

Robot může simultánní aktivací změnit svůj stav na jeden nový stav z množiny nových stavů, které jsou touto aktivací dosažitelné. Výsledek aktivace je tak možno chápat jako nedeterministický přechod do množiny nových stavů.

Pokud množina nových stavů obsahuje stav, který znamená nestabilní konfiguraci robotu, nesmí být příslušné aktivace použito v žádné kompoziční strategii. Robot rovněž nemůže přejít do stavu, který znamená nestabilní konfiguraci robotu a dále nemůže přejít do stavu, který je fyzikálně (mechanicky) nedosažitelný. Takto lze na základě čistě formálních úvah podstatně omezit jak počet stavů, tak počet přechodů a tím i počet možných kompozičních strategií.

Pro potřeby Q-učení jsou stavy robotu ztotožněny se stavy prostředí a simultánní aktivace řídicích členů s akcemi. Výsledkem Q-učení je nalezení optimálních simultánních aktivací řídicích členů v různých stavech robotu, tj. nalezení konkrétní optimální kompoziční strategie.

Zajímavá metodika omezení počtu diskrétních stavů je uvedena v [20] a např. [21, 22].

3 Základní řídicí členy

Pro řešení úlohy řízení robotu byla zvolena koncepce se dvěma základními řídicími členy konfigurace došlapu Φ_1 a konfigurace noh Φ_2 .

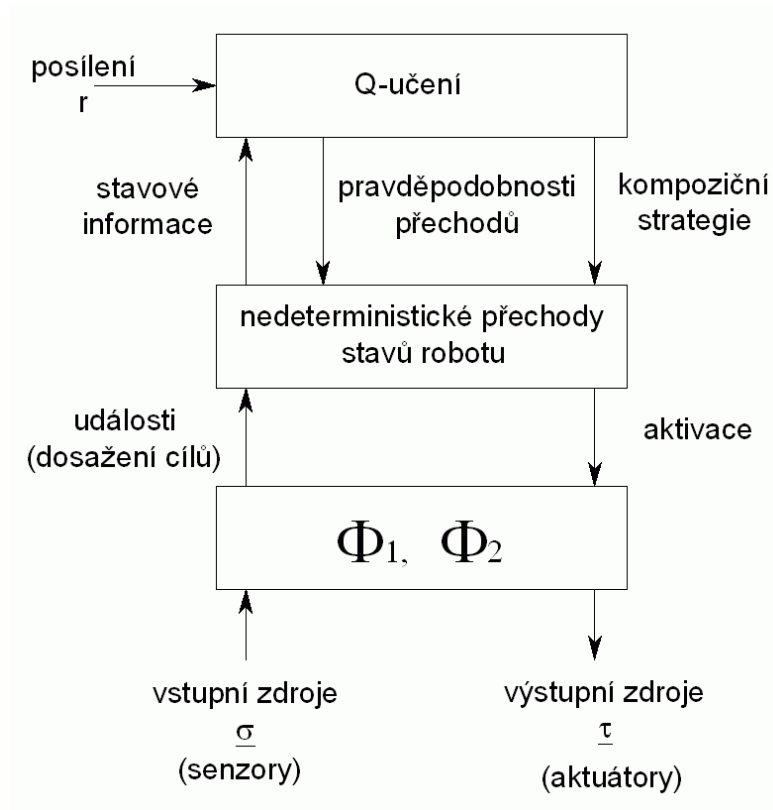
- a) Řídicí člen konfigurace došlapu Φ_1 je definován cílem dosáhnout změnou polohy došlapu jedné ze zvolených tří noh (staticky) stabilního postavení robotu na těchto třech nohách.
- b) Řídicí člen konfigurace noh Φ_2 je definován cílem dosáhnout změnou polohy a orientace těžiště robotu kinematicky optimálního postavení noh robotu. To znamená dosáhnout takové polohy aktuátorů noh, aby byly co nejbližší středním polohám mezi svými krajními polohami.

Celkové uspořádání architektury řídicího systému je uvedeno na obr. 1.

3.1 Instance základních řídicích členů

Základní řídicí funkce je realizována činností základního řídicího členu, který má přiřazenu konkrétní množinu vstupních zdrojů (senzorů) a výstupních zdrojů (aktuátorů). Hovoříme o instanci základního řídicího členu, zkráceně o základní instanci (specifikace zdrojů – viz obr. 2). Cílem základní instance je cíl odpovídajícího základního řídicího členu s konkrétními senzory a

aktuátory. Aktivní je ta základní instance, která aktuálně realizuje dosažení svého cílového stavu. Aktivací základní instance se rozumí převedení instance do aktivního stavu.



Obr. 1. Uspořádání architektury řídicího systému

Základní instance nabývá logické hodnoty 1 právě když aktuální konfigurace robotu vyhovuje cíli instance, jinak nabývá hodnoty 0. Je označena $\Phi_{i \frac{s}{t}}$, kde Φ_i je základní řídicí člen a $\underline{s}$ a $\underline{t}$ jsou přiřazené množiny vstupních a výstupních zdrojů.

- Základní instance řídicího členu konfigurace došlapu jsou ve tvaru $\Phi_{1 \frac{a,b,c}{a}}$, kde $a \neq b \neq c$ jsou čísla noh robotu; ve skutečnosti instance používá polohy koncových bodů noh s těmito čísly. Možných instancí je 12: $\Phi_{1 \frac{1,2,3}{*}}$, $\Phi_{1 \frac{0,2,3}{*}}$, $\Phi_{1 \frac{0,1,3}{*}}$, $\Phi_{1 \frac{0,1,2}{*}}$, kde $*$ zastupuje libovolné číslo nohy ze senzorů příslušné instance.
- Jediná základní instance řídicího členu konfigurace noh je ve tvaru $\Phi_{2 \frac{0,1,2,3}{x,y,j}}$, kde 0, 1, 2, 3 jsou opět čísla noh robotu a x, y, j poloha a orientace těžiště robotu.

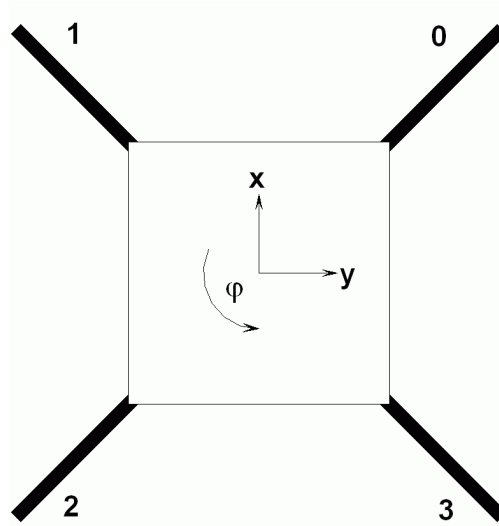
Pro zápis indikace dosažení cíle základní instancí řídicího členu je použito operátoru přiřazení „ $\leftarrow$ “.

Výraz $x_i \leftarrow \Phi_{j \frac{s_k}{t_k}}$ znamená přiřazení hodnoty booleovské proměnné x_i podle toho, zda aktuální konfigurace robotu vyhovuje cíli základní instance $\Phi_{j \frac{s_k}{t_k}}$.

Dosahování cílů uvažovaných základních instancí lze indikovat hodnotami následujících pěti booleovských proměnných:

$$\begin{aligned}
x_1 &\leftarrow \Phi_{1*}^{\underline{1,2,3}}, \quad x_2 \leftarrow \Phi_{1*}^{\underline{0,2,3}}, \quad x_3 \leftarrow \Phi_{1*}^{\underline{0,1,3}}, \quad x_4 \leftarrow \Phi_{1*}^{\underline{0,1,2}}, \\
x_5 &\leftarrow \Phi_{2 \underline{x,y,j}}^{\underline{0,1,2,3}},
\end{aligned} \tag{1}$$

přičemž proměnné x_1 až x_4 indikují stabilní postavení robotu na třech nohách a proměnná x_5 indikuje kinematicky optimální postavení čtyřech noh robotu. Tyto proměnné lze chápat jako složky diskrétního stavu robotu $x = (x_1, x_2, x_3, x_4, x_5)$. Činnost robotu pak lze popsat přechody mezi těmito stavy.



Obr. 2. Specifikace zdrojů

3.2 Kompozitní instance

Aby bylo možno dosáhnout složitějšího chování robotu jsou povoleny i simultánní aktivace více základních instancí. O takto získané instanci se hovoří jako o kompozitní instanci.

K popisu interakcí základních instancí je zaveden operátor „podmíněnosti“ „<“. Operátor omezuje řídicí proces realizovaný podřízenou základní instancí tak, aby nebylo dotčeno dosažení cíle nadřízenou základní instancí.

Zápis $\Phi_{i \underline{t_j}}^{\underline{s_j}} < \Phi_{k \underline{t_l}}^{\underline{s_l}}$ znamená, že instance $\Phi_{i \underline{t_j}}^{\underline{s_j}}$ je simultánně podmíněna instancí $\Phi_{k \underline{t_l}}^{\underline{s_l}}$, tj. že

- instance $\Phi_{i \underline{t_j}}^{\underline{s_j}}$ i $\Phi_{k \underline{t_l}}^{\underline{s_l}}$ byly aktivovány současně,
- instance $\Phi_{k \underline{t_l}}^{\underline{s_l}}$ omezuje řídicí proces realizovaný $\Phi_{i \underline{t_j}}^{\underline{s_j}}$ tak, aby nebylo dotčeno dosažení cíle instancí $\Phi_{k \underline{t_l}}^{\underline{s_l}}$.

Lze konstruovat i kompozitní instance z více než dvou základních instancí. Kompozitní instance dosáhne svého cíle, právě když dosáhne cíle každá z jejích instancí.

3.3 Stav a přechody

Možné přechody stavů robotu lze určit zcela formálně na základě vztahů mezi vstupními a výstupními zdroji instancí. Bude-li se např. instance $\Phi_{10}^{\underline{0,2,3}}$ pokoušet nastavit hodnotu proměnné x_2

na 1 změnou polohy došlapu nohy 0, mohou se změnit i hodnoty proměnných x_2 až x_5 , protože poloha koncového bodu nohy 0 je vstupním zdrojem instancí odpovídajících těmto proměnným. Ze stejného důvodu ale nebude ovlivněna hodnota proměnné x_1 . Je-li kráčivý robot ve stavu (10000), pak aktivace instance $\Phi_{10}^{0,2,3}$ povede na jeden ze stavů (1****), kde * udává, že odpovídající proměnná může mít libovolnou logickou hodnotu, tedy 0 nebo 1, i když není jasné, kterého z těchto stavů bude konkrétně dosaženo. Výsledek aktivace instance tak může být modelován jako nedeterministický přechod v diskrétním prostoru stavů robotu. Přechod je ohodnocen odpovídající kompozitní instancí.

3.4 Redukce počtu stavů a počtu přechodů

Při redukci počtu možných stavů a přechodů se ponechávají jen ty stavy a přechody, které vyhovují jednoduše formulovanému požadavku stability robotu a fyzikální dosažitelnosti cílů řízení.

a) Stabilní stavy robotu a přípustné kompozitní instance

Základním požadavkem je, aby byl kráčející robot stabilní. Omezíme-li se na statickou stabilitu, může robot zaujímat pouze taková postavení, ve kterých bude stabilní alespoň na jedné trojici. noh Robot tak může dosahovat pouze stavů, pro které booleovský výraz $s_1 + s_2 + s_3 + s_4$ nabývá hodnoty 1.

Úvahami o stabilitě lze efektivně zredukovat i počet přípustných kompozitních instancí tak, aby garantovaly pouze přechody do stabilních stavů. Např. kompozitní instance $\Phi_{11}^{0,1,3} < \Phi_{13}^{1,2,3}$ není přípustná, protože základní instance $\Phi_{13}^{1,2,3}$ může modifikovat hodnoty všech proměnných mimo x_4 a základní instance $\Phi_{11}^{0,1,3}$ může modifikovat hodnoty všech proměnných mimo x_2 . Existuje tak možnost modifikace hodnot všech proměnných a tím ti možnost přechodu do stavu (00000). Naproti tomu kompozitní instance $\Phi_{12}^{1,2,3} < \Phi_{2\ x,y,j}^{0,1,2,3} < \Phi_{12}^{0,1,2}$ přechod do stabilního stavu nevylučuje, protože může změnit hodnoty všech proměnných s výjimkou x_3 . Přechod do stabilního stavu je však garantován pouze ve stavech robotu (*1*). S uvažováním požadavku stability nelze zkonstruovat kompozitní instanci, jejíž cíl by byl indikován nastavením více než 3 proměnných z x_1 až x_4 .

b) Fyzikálně (mechanicky) dosažitelné stavy robotu a přípustné kompozitní instance

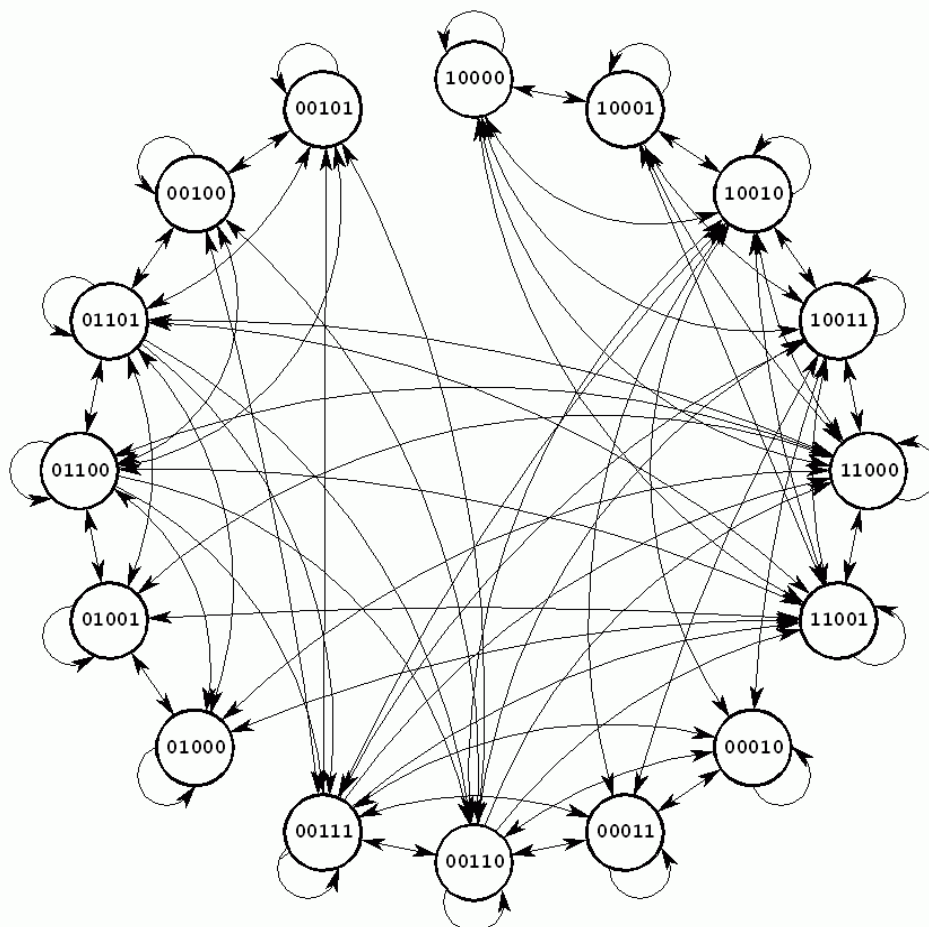
U čtyřnohého kráčivého robotu je vyloučeno současné dosažení stabilního postavení na protějších trojicích noh. Není možné, aby těžiště robotu padlo současně např. do trojúhelníků s vrcholy z noh (1,2,3) a (0,1,3), tj. připadají v úvahu pouze takové stavy robotu, pro které je booleovský výraz $(x_1 \rightarrow \bar{x}_3) \cdot (x_3 \rightarrow \bar{x}_1) \cdot (x_2 \rightarrow \bar{x}_4) \cdot (x_4 \rightarrow \bar{x}_2)$ roven 1. Tento výraz lze zjednodušit na $\overline{x_1 \cdot x_3 + x_2 \cdot x_4}$, který nabývá hodnoty 1 pro všechna mechanicky dosažitelná postavení robotu.

S uvažováním uvedené podmínky není přípustná např. kompozitní instance $\Phi_{10}^{0,2,3} < \Phi_{10}^{0,1,2}$ která sice ze stavů (1****) zaručuje přechod do stabilního stavu (nemodifikuje x_1), ale nemůže nikdy dosáhnout svého cíle $x_2 = x_4 = 1$. Rovněž není možné zkonstruovat kompozitní instanci, jejíž cíl by byl indikován nastavením více než 2 proměnných z x_1 až x_4 .

Uvedenému požadavku stability robotu a fyzikální dosažitelnosti cílů řízení vyhovuje

- 16 stavů robotu
- 100 kompozitních instancí, které mohou být složeny z nejvýše 3 základních instancí.

Protože ani počet přípustných stavů robotu, ani počet přípustných kompozitních instancí není příliš vysoký, lze uvažovat o efektivním použití Q-učení k získání optimální kompozitní instance pro každý z přípustných stavů. Graf přípustných přechodů stavů robotu je zobrazen na obr.3.



Obr.3: Graf přípustných přechodů stavů robotu

4 Modifikace Q-učení

Cílem učení je určit vhodnou kompoziční strategii, tedy odhadnout budoucí prospěch z provedení každého z přechodů jednoho stavu do druhého a poté provádět pouze ty přechody stavů, u kterých byl odhadnut maximální prospěch. Stavovým prostorem X pro Q-učení je prostor přípustných stavů robotu. Množinou akcí A je množina všech přípustných kompozitních instancí.

Původní iterační pravidlo Q-učení je modifikováno tak, aby místo dvojic

(stav robotu, kompozitní instance)

používalo trojic

(stav robotu, kompozitní instance, další stav robotu).

Tento rozklad Q-hodnoty umožňuje ohodnotit dílčími Q-hodnotami přímo přechody stavů robotu. Výsledné pravidlo pak nabývá tvaru:

$$Q_t(x, a, y) = Q_{t-1}(x, a, y) + b \left(r_t + g \max_{b \in A} Q_{t-1}(y, b) - Q_{t-1}(x, a, y) \right), \quad (2)$$

kde g je srážkový koeficient, b představuje parametr učení, t index iterace a

$$Q(x, a) = \sum_{y \in X} (p(x, a, y) Q(x, a, y)). \quad (3)$$

Výraz $p(x, a, y)$ označuje pravděpodobnost, že použití kompozitní instance a ve stavu x povede na stav y a proto

$$\sum_{y \in X} p(x, a, y) = \begin{cases} 1 & \text{pro } a \text{ přípustné ve stavu } x \\ 0 & \text{jinak} \end{cases}. \quad (4)$$

Výraz $Q(x, a, y)$ představuje Q-hodnotu přechodu ze stavu x do stavu y prostřednictvím kompozitní instance a . Iteračního pravidla pro aktualizaci $Q(x, a, y)$ se použije, kdykoliv dojde k přechodu (x, a, y) .

Do ohodnocení přechodů stavů robotu jsou přidány i pravděpodobnosti přechodů z jednoho stavu do druhého. Aktualizace pravděpodobností přechodu se provádí pomocí čítače počtu průchodů přechodem.

Optimální kompozitní instance a_{opt} pro každý stav x robotu je určena obvyklým způsobem jako

$$a_{opt} = \arg \max_{a \in A} Q(x, a). \quad (5)$$

Posloupnost posílení r_t je v současnosti definována tak, že hodnotí postup robotu směrem k zadanému uzlovému bodu, dosažený během posledního přechodu.

5 Závěr

Q-učení může být vhodnou metodou pro řízení složitých kinematických struktur v reálném čase pomocí konečných automatů. Uvedená metodika představuje originální přístup jak k podstatnému omezení počtu stavů konečného automatu, tak jeho přechodů. Tato omezení jsou nutná pro úspěšné a efektivní použití Q-učení. Přitom je zaručeno, že kompoziční strategie, která byla získána učením, nemůže vést na jiné než stabilní konfigurace robotu.

Předpokládají se průkazné experimenty po realizaci laboratorního modelu. V současné době probíhá tvorba modelů jednotlivých noh robotu, modelování jeho chůze a ověřování výše uvedeného přístupu k realizaci systému řízení robotu. Dosud provedené simulace ukazují na robustnost řídicího systému a potvrzují reálnost tohoto přístupu.

Poděkování

Príspevek byl zpracován za podpory pilotního projektu ÚT AV ČR č. 52020 „Řízení kráčivého robotu s využitím metod umělé inteligence“.

Literatura

- [1] Hirose, S.: A study of design and control of quadruped walking vehicle. *Int. J. Robotics Res.*, (3)2:113-133, 1984.
- [2] Arikawa, A., Hirose, S.: Study of walking robot for 3 dimensional terrain. In *ICRA, IEEE*, June, 1995.
- [3] Brooks, R. A.: A robust layered control system for mobile robot. *ICRA* (2)1:14-23, March, 1986.
- [4] Raibert, M.: *Legged Robots that Balance*. MIT Press, Cambridge, MA, 1986.
- [5] Celaya, E., Porta, J. M.: Force-based control of six-legged robot on abrupt terrain using the subsumption architecture. In *Int. Conf. on Adv. Rob.*, Sant Felin de Guixols, Spain, 1995.
- [6] Nagy, P., Whittaker, W., Desa, S.: A walking prescription for statically-stable walkers based on walker/terrain interaction. In *ICRA, IEEE*, Nice, France, 1992.
- [7] Pack, D., Kang, H.: An omnidirectional gait control using a graph search method for quadruped walking robot. In *ICRA, IEEE*, vol. 1, pp. 988-993, Nagoya, Japan, May 1995.
- [8] Wettergreen, D., Pangels, H., Bares, J.: Behavior-based gait execution for the Dante II walking robot. In *International Conference on Intelligent Robots and Systems, IEEE*, pp. 274-279, Pittsburgh, PA, August, 1995.
- [9] Guihand, M., Gorce, P., Fontaine, J.G., M'Sirdi, N.: A solution to control the dynamic behavior of a pneumatic quadruped robot. In *ICRA, IEEE*, pp. 1000-1005, Nagoya, Japan, May, 1995.
- [10] Takanishi, A., Yamaguchi, J., Iwata, M.: Dynamic quadruped walking stabilized with trunk motion. In *International Conference on Intelligent Robots and Systems, IEEE*, pp. 165-172, Pittsburgh, PA, August, 1995.
- [11] Yoneda, K., Iiyama, H., Hirose, S.: Dynamic and static fusion of quadruped walking vehicle on winding path/ In *ICRA*, pp. 143-148, Nice, France, May, 1992.
- [12] Adachi, H., Koiachi, N., Nakamura, T.: Adaptive gait for quadruped robot using force sensor. In *Proc. Int. Aut. Syst.* Pp. 54-63, Pittsburgh, PA, Feb, 1993.
- [13] Klein, D. P. C., Olson, K.: Use of force and attitude sensors for locomotion of legged vehicle over irregular terrain. *Int. J. Robotics Res.*, 2(2), 1983.
- [14] Yoneda, K., Iiyama, H., Hirose, S.: Sky-hook suspension control of a quadruped walking vehicle. In *ICRA, IEEE*, pp. 999-1004, San Diego, CA. May, 1994.
- [15] McGhee, R., Iswadi, G.: Adaptive locomotion of multilegged robot over rough terrain. *IEEE Trans. Syst. Man Cyber.*, 9(4):176, 1979.
- [16] Song, S. M., Waldron, K., J.: *Machines that Walk: The adaptive suspension Vehicle*. MIT Press, Cambridge, MA, 1989.
- [17] de Santos, P., Jimenez, M.: Generation of discontinuous gaits for quadruped walking vehicles. *J. Robotics Sys.* 12(9):599-611, 1995.
- [18] Jeong, K., Yang, T., Oh, J.: A study on the support pattern of quadruped walking robot for aperiodic motion. In *Proc. IROS*, pp. 308-313, Pittsburgh, PA, August, 1995.
- [19] Watkins, C.J.C.H., Dayan, P.: Q-learning. *Machine learning.* 8(3), 1992, pp. 279-292.
- [20] Huber, M., Grupen, R.A.: A hybrid discrete event dynamic systems approach to robot control. Technical report 96-43, CMPSCI Dept., Univ. of Mass., Amherst, October, 1996.
- [21] Huber, M., Grupen, R.A.: Feedback Control Structure for On-line Learning Tasks. *Robotics and Autonomous Systems*, 22(3-4):303-315, Elsevier, 1997.
- [22] Huber, M., Grupen, R.A.: A control structure for learning locomotion gaits. In *ISORA'97*, Anchorage, AK, May 1998, TSI Press.