

Aplikácia metódy TLD na sledovanie objektov v stereovíznom obraze

¹Michal PUHEIM, ²Marek BUNDZEL

¹² Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská republika

¹michal.puheim@hotmail.com, ²marek.bundzel@tuke.sk

Abstrakt — Tracking-Learning-Detection (TLD) je algoritmus použiteľný na sledovanie ľubovoľného objektu vo videu v reálnom čase. Táto metóda je známa svojím vysokým výkonom a robustnosťou. Cieľom predkladanej práce je ukázať aplikáciu TLD bežiacu na robotovi Nao využívajúcu jeho hlavu so stereovíznymi kamerami. Táto aplikácia by mala skvalitniť informácie, ktoré je robot schopný získať zo svojich kamerových senzorov a zvýšiť jeho schopnosť vykonávať rôzne aktivity a úlohy v ľubovoľnom prostredí. Pre ukážku možného použitia sme implementovali regulátor založený na neurónovej sieti, ktorý bol použitý na ovládanie pohybu ruky robota Nao podľa vnímanej lokácie objektu sledovaného pomocou stereovíznej aplikácie TLD.

Kľúčové slová — Tracking, Learning, Detection, TLD, robot, Nao, sledovanie objektu, stereovízia, neurónová sieť, regulátor

I. ÚVOD

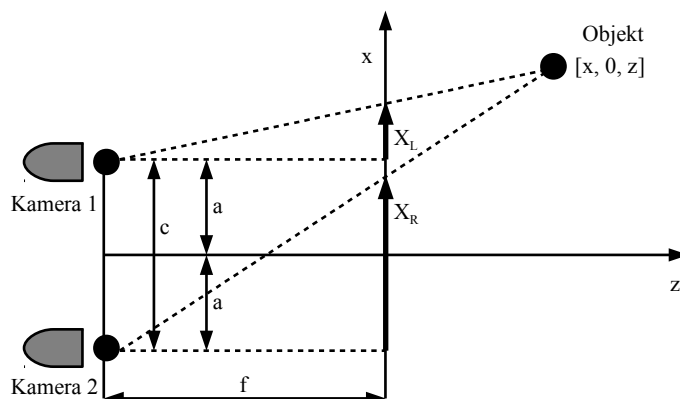
Trojrozmerné stereovidenie v súčasnosti nachádza mnoho aplikácií v oblasti zábavného priemyslu, prenosu informácií a autonómnych systémov. Stereovízia je veľmi dôležitá v oblastiach robotiky, napríklad pri extrakcii informácií o relatívnej pozícii 3D objektov v blízkom okolí autonómnych systémov. Ďalšie aplikácie v oblasti robotiky zahŕňajú rozpoznávanie objektov, kde informácia o hĺbke v obraze umožňuje systému separovať prekryté komponenty, napríklad jednu stoličku umiestnenú na popredí ďalšej. V prípade použitia snímania obrazu jednou kamerou by ich robot nebol schopný rozlíšiť ako samostatné objekty.

Predkladaná práca [1] navrhuje aplikáciu metódy Tracking-Learning-Detection (TLD) [2]-[4] v stereovízii na robotovi Nao. TLD je metóda schopná detegovať a sledovať objekt v po sebe nasledujúcich snímkach videa. Pozícia sledovaného objektu je určená súradnicami rámca ohraničujúceho objekt, ktorými sú x , y , šírka, výška. To nám síce dáva predstavu o pozícii objektu na obraze kamery, avšak akákoľvek ďalšia informácia o jeho umiestnení v reálnom svete chýba. Tento problém pramení z dôvodu, že metóda TLD spracúva iba dvojrozmerný obraz a teda žiadne informácie o treťom rozmere pre ňu ani nie sú dostupné. S cieľom získať túto chýbajúcu informáciu je možné použiť ďalší TLD detektor na kamere s iným umiestnením v priestore. Pomocou informácie o ohraničujúcom rámci objektu na obrazoch oboch kamier a informácie o vzájomnej polohe týchto kamier sme schopní relatívne presne určiť pozíciu objektu v trojrozmernom priestore. Tento poznatok potom môžeme ďalej použiť s cieľom zvýšiť povedomie robota Nao o okolitom svete, určiť jeho polohu, navigovať robota medzi viacerými viacerými referenčnými bodmi atď.

Navrhnutý systém je overovaný v dvoch aplikačných oblastiach. Prvou je určovanie vzdialenosti pomocou navrhnutého stereovízneho TLD systému. Druhou je kombináciu tohto systému s regulátorom ruky založeným na neurónovej sieti, ktorého cieľom je ovládanie ruky robota v snahe dotknúť sa sledovaného objektu.

II. PRINCÍP URČOVANIA VZDIALENOSTI V STEREOVÍZNYCH KAMEROVÝCH SYSTÉMOCH

Majme dve kamery v stereovíznej konfigurácii, ktorých optické osi sú posunuté voči centrálnej optickej osi, ako je ukázané na obr. 1.



Obr. 1 Odhad súradnice z použitím stereo kamier s paralelnými optickými osami. Vzdialenosť medzi kamerami je $c = 2a$, kde a je vzdialenosť kamery od optickej osi. X_R a X_L sú súradnice objektu tak ako sú videné pravou a ľavou kamerou. Ak f je ohnisková vzdialenosť, potom súradnica z môže byť vypočítaná pomocou X_R a X_L .

Predpokladajme, že f je ohnisková vzdialenosť (rovnaká pre obe kamery), c je vzdialenosť medzi kamerami a a je vzdialenosť kamier od stredovej optickej osi, pričom $c = 2a$. Ďalej predpokladajme, že X_R je súradnica x objektu videná pravou kamerou a X_L je súradnica x objektu videná ľavou kamerou. Použitím podobnosti trojuholníkov získavame nasledovné rovnosti:

$$\frac{X_L}{f} = \frac{(x-a)}{z}, \quad \frac{X_R}{f} = \frac{(x+a)}{z}. \quad (1)$$

Zlúčením týchto rovníc a elimináciou x získame vzorec, ktorý môžeme následne použiť na určenie súradnice z :

$$z = \frac{-2af}{X_L - X_R}. \quad (2)$$

Táto rovnica inak povedané hovorí, že pri znalosti parametrov kamier f a c , vieme využiť rozdiel vo videní toho istého objektu oboma kamerami s cieľom určiť vzdialenosť objektu od obrazovej roviny (tj. vzdialenosť objektu od kamier).

III. TRACKING-LEARNING-DETECTION

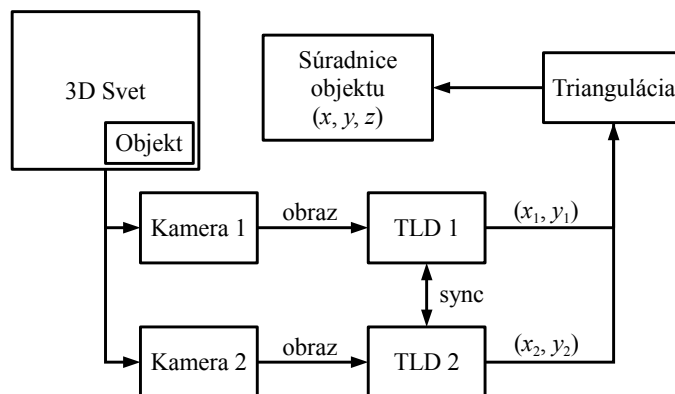
Tracking-Learning-Detection (TLD) [2]-[4] je metóda určená na dlhodobé sledovanie ľubovoľného objektu v neohraničenom prostredí. Veľkou výhodou systému je fakt, že nepotrebuje samostatnú učiacu fázu (tzv. offline učenie), stačí iba určiť sledovaný objekt v prvej snímke videa. Ďalšie učenie nových a pozmenených zhládov objektu prebieha priamo počas sledovania objektu (tzv. online učenie). Systém TLD je zostavený z troch hlavných komponentov. Sú to:

- Tracker – zabezpečuje krátkodobé sledovanie trajektórie objektu. Je založený na metóde Lucasa-Kanadeho [5] a je použitý na generovanie príkladov použitých pri učení detektora.
- Detektor – má formu tzv. náhodného lesa („randomised forest“) [6] a umožňuje inkrementálnu aktualizáciu rozhodovacích medzí a sekvenčnú evaluáciu počas behu v reálnom čase.
- Učiaci algoritmus – použité je tzv. P-N učenie [2], ktoré používa tracker na generovanie pozitívnych (P) a tiež negatívnych (N) príkladov, ktoré sú následne použité na zlepšenie modelu detektora.

Cieľom metódy je sledovanie objektu trackerom a súčasne simultánne učenie s cieľom vybudovať detektor, ktorý bude v prípade zlyhania trackera schopný objekt znovu detegovať. Detektor je budovaný na základe informácie z prvej snímky videa a taktiež na náklade informácií z trackera. Keďže obe komponenty nevyhnutne robia chyby, stabilita systému je dosiahnutá ich vzájomným anulovaním. Naučený detektor umožňuje re-inicializáciu trackera vždy, keď dôjde k znovuobjaveniu predtým pozorovaného zhládu objektu.

IV. NÁVRH STEREOVÍZNEHO SYSTÉMU TLD

Cieľom navrhnutého systému je využitie metódy TLD pri použití stereovíznej hlavy robota Nao. Takto navrhnutý systém po svojom dokončení umožňuje robotovi sledovať ľubovoľne vybraný objekt, a tiež určiť jeho pozíciu v priestore (viď obr. 2).



Obr. 2 Zjednodušená bloková schéma stereovízneho systému využívajúceho metódu TLD. Objekt v trojrozmernom svete je snímaný dvoma kamerami. Obraz z kamier je na vstupe dvoch synchronizovaných TLD systémov, ktorých výstupmi sú súradnice objektu na obrazoch kamier. Nakoniec sa pomocou triangulácie vypočítajú súradnice objektu v reálnom svete.

A. Spôsob využitia metódy TLD v systéme

Sledovaný objekt je najprv manuálne označený užívateľom zadáním ohraničujúceho rámca. Pomocou tohto rámca je inicializovaný jeden z TLD systémov a následne je užívateľovi ponúknutá možnosť trénovať tento detektor (napr. pohybom označeného objektu, zmenou jeho vzdialenosti, resp. otočením). Tento tréningový mód nie je nevyhnutný, ale v konečnom dôsledku zlepšuje úspešnosť a presnosť sledovania objektu počas ďalšej činnosti stereovízneho systému.

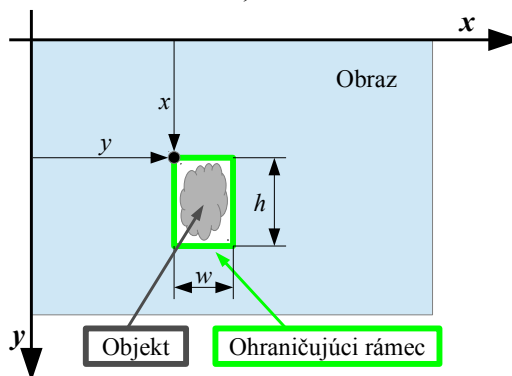
Po inicializácii počiatočného TLD detektora (a voliteľnom učení) sa tento detektor skopíruje, čím dostaneme dva samostatné TLD systémy, každý pre jednu kameru. Pri návrhu stereovízneho systému teda počítame s dvoma paralelne bežiacimi TLD, kde každý sleduje objekt na jednej z kamier v stereovíznej konfigurácii. Návrh tiež ráta s možnosťou pravidelnej synchronizácie modelov oboch detektorov s cieľom predísť prílišnému rozdielu medzi modelmi objektu týchto detektorov.

B. Určenie súradníc a vzdialenosti sledovaného objektu

Každý z paralelne bežiacich TLD systémov poskytuje na výstupe štyri hodnoty, definujúce ohraničujúci rámec objektu na obraze z kamery. Sú to:

- x – horizontálna poloha ľavého horného rohu ohraničujúceho rámca v obraze,
- y – vertikálna poloha ľavého horného rohu ohraničujúceho rámca v obraze,
- w – šírka ohraničujúceho rámca,
- h – výška ohraničujúceho rámca,

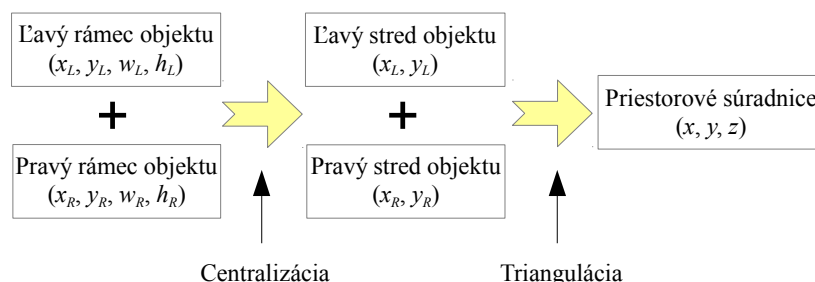
pričom je potrebné poznamenať, že súradnicový systém pri riešení úloh počítačového videnia spravidla začína v ľavom hornom rohu obrazu, viď obr. 3.



Obr. 3 Výstupom TLD systému je rámec ohraničujúci sledovaný objekt. Tento rámec je definovaný štyrmi hodnotami (x, y, w, h) , teda horizontálnou a vertikálnou polohou ľavého horného rohu rámca, jeho šírkou a výškou.

Pomocou týchto hodnôt následne vypočítame stredy (x_c, y_c) ohraničujúcich rámcov pre obe TLD systémy:

$$x_c = x + \frac{w}{2}, \quad y_c = y + \frac{h}{2}. \quad (3)$$



Obr. 4 Postup určenia priestorových súradníc objektu pri použití dvoch paralelných TLD systémov sledujúcich rovnaký objekt.

Po určení stredov objektov môžeme pristúpiť k určeniu súradníc v trojrozmernom priestore. Nech (x_L, y_L) sú stredové súradnice objektu na obraze sledovanom ľavou kamerou a obdobne (x_R, y_R) sú stredové súradnice objektu na obraze sledovanom pravou kamerou. Tiež si uvedomme, že hodnoty x_L, y_L, x_R, y_R sú uvedené v pixloch. Ak w_i je šírka obrazu (v pixloch), h_i je výška obrazu (v pixloch), f je ohnisková vzdialenosť kamier (v pixloch) a c je vzdialenosť medzi kamerami (v metroch), potom hodnoty priestorových súradníc (x, y, z) v metroch určíme nasledovne:

$$z = \frac{cf}{x_L - x_R}, \quad x = \frac{c}{2} + z \frac{\left(x_L - \frac{w_L}{2}\right)}{f}, \quad y = -z \frac{\left(y_L - \frac{h_L}{2}\right)}{f}. \quad (4)$$

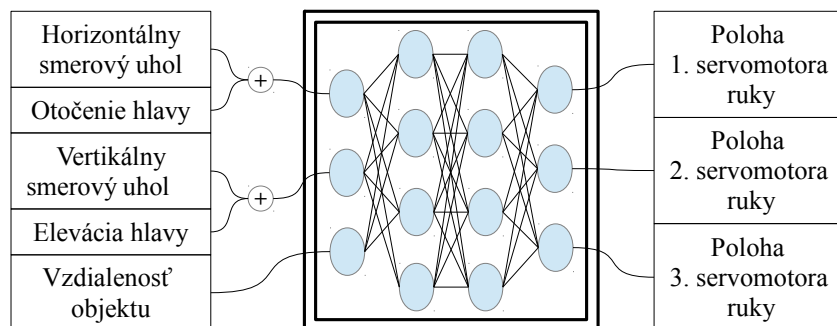
Absolútnu vzdialenosť objektu d potom určíme ako veľkosť vektora určeného súradnicami (x, y, z) :

$$d = \sqrt{x^2 + y^2 + z^2}, \quad (5)$$

V. NÁVRH REGULÁTORA RUKY ZALOŽENÉHO NA NEURÓNOVEJ SIETI

Za účelom overenia možností navrhnutého stereovízneho systému sme sa rozhodli navrhnúť špeciálny regulátor pohybu ruky robota. Jeho cieľom je umožniť robotovi dotknúť sa sledovaného predmetu na základe informácií o polohe objektu získanej z výstupov stereovízneho systému. Na tento účel bola použitá dopredná viacvrstvomá neurónová sieť.

Vstupmi neurónovej siete sú dostupné informácie o objekte získané stereovíznym TLD systémom a tiež informácie o polohe krčných servomotorov (viď obr. 5). Namiesto karteziánskych súradníc objektu sú použité polárne súradnice. Teda vstupnými informáciami sú vzdialenosť objektu od hlavy robota a tiež smer objektu vztiahnutý voči telu robota. Smer je určený dvoma smerovými uhlami – horizontálnym a vertikálnym, pričom každý z nich zahŕňa zložku otočenia hlavy robota spolu s posunutím objektu voči centru zorného poľa.



Obr. 5 Regulátor pohybu ruky robota založený na doprednej neurónovej sieti. Vstupy sú informácie o otočení hlavy robota a informácie o objekte získané stereovíznym TLD systémom. Výstupmi sú polohy servomotorov ruky nastavením ktorých sa ruka robota dostane do kontaktu s objektom.

Výstupmi neurónovej siete sú polohy trojice krokových servomotorov na ruke robota. V prípade správne natrénovanej NS sa nastavením servomotorov do danej polohy ruka robota dostane do kontaktu s objektom. Na tréning NS je však potrebné dostatočné množstvo tréningových dát. Spôsob ich získavania nebudeme z dôvodu rozsiahlosti problematiky rozoberať na tomto mieste, bližší popis je uvedený v samotnej diplomovej práci.

VI. EXPERIMENTY

Po dokončení systému sme prišli k otestovaniu jeho potenciálnych aplikačných možností. Celkovo boli vykonané dva typy pokusov. Prvý sa zaoberá presnosťou merania vzdialenosti objektu pomocou navrhnutého stereovízneho systému. Druhý sa venuje otestovaniu navrhnutého neuroregulátora ruky.

A. Určovanie vzdialenosti objektu

V prvom experimente sme overovali samotnú presnosť merania vzdialenosti. Sledovaný objekt sme 20 krát umiestnili do rôznych vzdialeností od robota, a pre každú z nich sme porovnávali výstupy systému s reálnymi hodnotami vzdialenosti. Vzdialenosť sme merali v rozsahu 0,4 až 2 metre. Ako sledovaný objekt bola použitá kniha s rozmerom 18x23 cm. Hodnoty nameraných vzdialeností v metroch sú uvedené v tab. 1.

Tabuľka 1
 Výsledky experimentu s meraním vzdialenosti.

Č. merania	Reálna vzdialenosť (m)	Nameraná vzdialenosť (m)	Rozdiel vzdialeností (m)
1.	0.5348	0.3873	0.1475
2.	0.6132	0.4553	0.1475
3.	0.6972	0.6403	0.0569
4.	0.7849	0.6778	0.1071
5.	0.8752	0.7223	0.1529
6.	0.9675	1.0049	-0.0374
7.	1.0142	0.8145	0.1997
8.	1.0611	1.0870	-0.0259
9.	1.1559	1.0040	0.1519
10.	1.2514	1.1738	0.0776
11.	1.2994	1.0040	0.2954
12.	1.3476	1.2771	0.0705
13.	1.4443	1.2771	0.1672
14.	1.4929	1.0851	0.4078
15.	1.5414	1.5327	0.0087
16.	1.6389	1.6927	-0.0538
17.	1.7367	1.3960	0.3407
18.	1.8346	1.5327	0.3019
19.	1.9329	1.6927	0.2402
20.	2.0313	1.6927	0.3386
Priemerná absolútna chyba merania:			0.1670

Hodnoty sú v metroch Každý riadok určuje jedno meranie, posledný stĺpec ukazuje chybu konkrétneho merania. Priemerná absolútna chyba všetkých meraní je 0.1670 m.

B. Dotyk objektu

Cieľom druhého experimentu bolo overiť schopnosti systému pri použití regulátora ruky založeného na neurónovej sieti pri pokuse o dotyk sledovaného objektu. Ako regulátor bola použitá dopredná neurónová sieť s tromi vstupmi aj výstupmi. Sieť mala dve skryté vrstvy, pričom každá z nich obsahovala štyri neuróny. V neurónoch bola použitá sigmoidálna aktivačná funkcia. Váhy siete boli inicializované v intervale $[-0.1; 0.1]$. Na učenie siete bolo použitých 151 tréningových vzoriek. Učiaci parameter α bol nastavený na hodnotu 0.2. Po 3824 cykloch učenia v čase 2.18 sekundy dosahovala sieť na tréningových dátach presnosť mapovania 95.0043 %. Počas experimentu sme sledovaný objekt 20 krát umiestnili do rôznych pozícií v dosahu ruky robota a zisťovali sme, koľkokrát došlo k dotyku ruky robota s objektom. Výsledky testovania¹ ukazuje tab. 2.

¹Video z priebehu časti experimentu je dostupné na <http://vimeo.com/66732997>

Tabuľka 2
 Výsledky experimentu s dotykom objektu.

Č. merania	Vstupy regulátora (súradnice objektu)			Výstupy regulátora (hodn. servomotorov)			OK?
	d (m)	v (rad)	h (rad)	s_p (rad)	s_r (rad)	e_r (rad)	
1.	0.134146	0.027178	0.290937	0.057461	0.357181	-0.788319	A
2.	0.190570	-0.068388	0.589369	-0.172792	0.477867	-0.169463	A
3.	0.198999	-0.073607	0.765424	-0.184782	0.740034	-0.160568	A
4.	0.164649	0.103553	0.424750	0.061034	0.431430	-0.588861	A
5.	0.221017	-0.173743	0.923930	-0.495389	0.778589	-0.039320	A
6.	0.223769	0.021826	0.735878	-0.166607	0.515333	-0.049318	A
7.	0.162134	-0.435731	0.842525	-0.614742	1.071152	-0.477502	N
8.	0.194244	-0.387002	1.168614	-0.553836	1.345710	-0.174646	N
9.	0.191741	-0.055399	0.831764	-0.122697	0.975315	-0.333835	A
10.	0.185708	-0.190491	0.733462	-0.329463	0.733202	-0.207705	A
11.	0.153799	-0.011105	0.150542	-0.038457	0.190594	-0.393123	A
12.	0.163303	0.320239	0.386022	0.225971	0.462984	-0.762712	A
13.	0.185293	0.017649	0.357128	-0.072228	0.261859	-0.185558	A
14.	0.217610	-0.034607	0.627409	-0.253171	0.383026	-0.042686	A
15.	0.201207	-0.011139	0.518828	-0.137416	0.347757	-0.096682	A
16.	0.209144	-0.268592	1.175758	-0.458409	1.302464	-0.116085	A
17.	0.129244	-0.330636	0.627409	-0.385313	0.940373	-1.098963	N
18.	0.141164	-0.175151	0.122507	-0.166845	0.178159	-0.397670	N
19.	0.193073	-0.401333	0.510929	-0.773710	0.276209	-0.030392	A
20.	0.193977	0.237270	0.403652	0.181841	0.320561	-0.257261	A
Počet úspešných dotykov: 16				Počet neúspešných dotykov: 4		Úspešnosť: 80 %	

Každý riadok určuje jedno meranie, hodnoty vzdialenosti d sú v metroch, všetky ostatné hodnoty sú v radiánoch. Vstupy neuroregulátora: d – vzdialenosť objektu, v – vertikálny smerový uhol objektu, h – horizontálny smerový uhol objektu. Výstupy neuroregulátora: s_p – plečný vertikálny servomotor (shoulder pitch), s_r – plečný horizontálny servomotor (shoulder roll), e_r – lakťový horizontálny servomotor (elbow roll). Posledný stĺpec určuje či došlo, alebo nedošlo k dotyku ruky robota s objektom.

VII. ZÁVER

Výsledky experimentu s meraním vzdialenosti ukázali, že systém nie je vhodný na presné meranie vzdialenosti a to najmä pri vzdialenejších objektoch. Značné nepresnosti systému, najmä pri väčších vzdialenostiach, spôsobuje rozdielna detekcia objektu oboma TLD systémami. Možným riešením je častejšia synchronizácia detektorov, ktorá ale prináša zníženie celkovej stability systému, keďže menej presný detektor sa musí neustále „reštartovať“.

Na druhej strane veľmi potešujúce výsledky priniesol v poradí druhý experiment. Ukázal, že pre niektoré aplikácie v oblasti robotiky úplne presné meranie vzdialenosti nie je potrebné, pretože z hľadiska robota je nameraná poloha vždy „presná“, v prípade že sa s inherentnou nepresnosťou počíta už v priebehu fázy učenia. Presnosť navrhnutého neuroregulátora je možné ďalej zlepšiť (viď. [1]), čo s vysokou pravdepodobnosťou umožní dosiahnuť presnosť dostatočnú nie len na dotyk, ale aj na uchopenie sledovaného objektu a ďalšiu manipuláciu s ním.

Problémom navrhnutého stereovízneho TLD systému je však fakt, že okrem polohy objektu v priestore nedokáže povedať nič o veľkosti, alebo tvare objektu. Tento nedostatok je pri snahe o presnú manipuláciu s objektmi rôznych tvarov a veľkostí veľmi závažný. Teoreticky je možné, že by sa dal prekonať využitím inej metódy sledovania obrazu (viď [7]). Avšak problémom sofistikovanejších sledovacích metód je ich nízka robustnosť (napr. oproti metóde TLD). Riešením by mohol byť hybridný systém využívajúci viaceré metódy.

REFERENCIE

- [1] M. Puheim. "Aplikácia metódy TLD na sledovanie objektov v stereovíznom obraze." Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2013. 68 s.
- [2] Z. Kalal, J. Matas, K. Mikolajczyk. "P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints." Conference on Computer Vision and Pattern Recognition. 2010.
- [3] Z. Kalal, J. Matas, K. Mikolajczyk. "Online learning of robust object detectors during unstable tracking." 3rd On-line Learning for Computer Vision Workshop 2009, Kyoto, Japan, IEEE CS, 2009.
- [4] Z. Kalal, J. Matas, K. Mikolajczyk. "Forward-Backward Error: Automatic Detection of Tracking Failures." International Conference on Pattern Recognition, Istanbul, Turkey, 23-26 August, 2010.
- [5] B. D. Lucas, T. Kanade. "An iterative image registration technique with an application to stereo vision." Proceedings of the International Joint Conference on Artificial Intelligence, pages 674–679, 1981.
- [6] L. Breiman. "Random forests." ML, 45(1):5–32, 2001.
- [7] A. Yilmaz, O. Javed, M. Shah. "Object Tracking: A Survey." ACM Comput. Surv. 38, 4, Article 13, 45 pages. Dec. 2006.