

# Analýza dynamiky vo videu pomocou lokálnych deskriptorov

Peter Benický, Ladislav Jurišica

## Abstrakt

V mobilnej robotike je nevyhnutné, aby robot obsahoval inteligenciu, na základe ktorej môže reagovať na prostredie, v ktorom sa nachádza. Prostredie je však natoľko rôznorodé, že je nevyhnutné spracovať informácie zo senzorickeho subsystému robota, ktorého súčasťou je aj kamerový systém. Video z kamerového systému obsahuje množstvo informácií vhodných pre riadenie mobilného robota ktoré môžeme získať zo spracovania statických obrazov alebo dynamiky objektov vo videu. Pre analyzovanie a klasifikáciu udalostí je nevyhnutné pracovať s lokálnymi deskriptormi, čo je nevyhnutné aj v mobilnej robotike, napríklad schopnosť robota zareagovať na pohyb iného objektu v okolí (robotický futbal). V tomto článku sú prezentované niektoré druhy lokálnych deskriptorov, pomocou ktorých je možné analyzovať a klasifikovať pohybujúci sa objekt vo videu a ich využitie je možné aj v mobilnej robotike.

**Kľúčové slová:** spracovanie video udalostí, lokálny deskriptor, klasifikácia pohybu

## Úvod

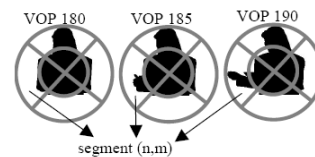
Pomocou zmien deskriptorov, ktoré opisujú tvar pohybujúceho sa objektu, môžeme klasifikovať nielen daný objekt, ale aj jeho akciu. Vzhľadom na skutočnosť, že daný objekt sa môže javiť ako pohybujúci sa objekt v dôsledku jeho skutočného pohybu, ale aj v dôsledku pohybu kamerového systému, môžeme deskriptory deliť do nasledovných troch skupín:

- globálny pohyb videa alebo pohyb kamery (priblíženie kamery, atď.)
- globálny pohyb video objektov (pohyb objektu v rámci scény)
- lokálny pohyb video objektu (pohyb ruky človeka)

Existuje niekoľko spôsobov pre výpočet lokálnych deskriptorov, pretože daný typ úlohy nebol zatiaľ dostatočne vyriešený a preto každý spôsob výpočtu lokálnych deskriptorov má svoje špecifické uplatnenie. V nasledovných kapitolách je uvedených niekoľko spôsobov výpočtu lokálnych deskriptorov.

## 1. Deskriptor založený na uhlovo kruhovej transformácii

Video objekt (VOP – z angličtiny „Video Object Plane“) je rozdelený do uhlových a kruhových segmentov (viď Obr. 1)



**Obr. 1 Rozdelenie video objektu na segmenty**  
**Fig. 1 Video object divided into segments**

Zmeny sú počítané zvlášť pre každý segment:

$$\sigma_{n,m}^2 = \frac{1}{A(n,m)K} \sum_{k=0}^{K-1} \sum_{\Theta=\Theta_m}^{\Theta_{m+1}} \sum_{\delta=\delta_n}^{\delta_{n+1}} (VOP_k(\delta, \Theta) - \mu_{n,m})^2 \quad (1)$$

$$\mu_{n,m} = \frac{1}{A(n,m)K} \sum_{k=0}^{K-1} \sum_{\Theta=\Theta_m}^{\Theta_{m+1}} \sum_{\delta=\delta_n}^{\delta_{n+1}} VOP_k(p, \Theta) \quad (2)$$

Pričom  $K$  je počet video objektov (VOP) a  $VOP_k$  je binárna maska video objektu v  $k$ -tom okamihu,  $VOP_k(\delta, \theta)$  je hodnota binárnej masky  $VOP_k$  s polohou  $(\delta, \theta)$  polárnych súradníc centralizovaných do stredu  $VOP_k$ ,  $A(n,m)$  je plocha,  $\theta_m$  je počiatočný uhol,  $\theta_m$  je počiatočný polomer v uhlovo – kruhovom segmente  $(n,m)$ . Plocha  $A(n,m)$  je definovaná nasledovne:

$$A(n,m) = \frac{\pi(\delta_{n+1}^2 - \delta_n^2)}{M}, \quad \Theta_m = m \times \frac{2\pi}{M}, \quad \delta_n = n \times \frac{\delta_{\max}}{N} \quad (3)$$

Kde  $M$  a  $N$  je počet uhlovo-kruhových sekcií,  $\delta_{\max}$  je definovaný nasledovne

$$\delta_{\max} = \max_{VOP_k \in VO} \{\delta VOP_k\} \quad (4)$$

Kde  $\delta VOP_k$  je polomer najmenšej kružnice schopnej pokryť  $VOP_k$ , pričom stred kružnice je v ťažisku  $\delta VOP_k$ .

## 2. Deskriptor založený na uhlovo polomerovej transformácii

Uhlovo-polomerová transformácia (UPT) je efektívna metóda pre získanie informácií o objekte. Pohybové deskriptory získané pomocou UPT môžu na rozdiel od deskriptorov získaných Fourierovou Transformáciou opísať zložitejšie a nespojité objekty. Deskriptory založené na UPT sú súčasťou MPEG-7. Koeficienty UPT sa počítajú pre každý video objekt (VOP) zvlášť. UPT je definovaná nasledovne:

$$F_{nm} = \int_0^{2\pi} \int_0^1 A_m(\Theta) R_n(\delta) f(\delta, \Theta) \delta d\delta d\Theta \quad (5)$$

Kde  $F_{nm}$  sú koeficienty UPT,  $f(\delta, \Theta)$  je binárny obraz objektu v polárnych súradniciach, funkcie  $A_m(\theta)$  a  $R_n(\delta)$  sú definované nasledovne:

$$A_m(\Theta) = \frac{1}{2\pi} e^{j.m.\Theta}, R_n(\delta) = \begin{cases} 1 & n = 0 \\ 2.\cos(\pi.n.\delta) & n \neq 0 \end{cases} \quad (6)$$

Pre výpočet diskretných koeficientov UPT musíme najprv normalizovať veľkosť binárneho obrazu objektu pomocou lineárnej interpolácie na preddefinovanú šírku  $S$  a výšku  $V$ , čím zaručíme invariantnosť voči veľkosti. Ťažisko binárneho objektu je zarovnané so stredom obrazu  $I(x,y)$ , t.j.  $I(S/2, V/2)$ . Diskretné koeficienty UPT  $k$ -teho video objektu ( $VOP_k$ ) potom vypočítame:

$$F_{nm}(VOP_k) = \sum_{x=-\frac{S}{2}}^{\frac{S}{2}} \sum_{y=-\frac{V}{2}}^{\frac{V}{2}} W_{nm} \left( \arctan \frac{y}{x}, \sqrt{x^2 + y^2} \right) I_{VOP_k} \left( x + \frac{S}{2}, y + \frac{V}{2} \right) \quad (7)$$

Kde,

$$W_{nm}(\delta, \Theta) = A_m(\Theta) R_n(\delta) \quad (8)$$

Následne z uvedených UPT koeficientov vypočítame zmeny:

$$O^2_{nm} = \sum_{k=0}^{K-1} (F_{nm}(VOP_k) - \mu_{nm})^2, \mu_{nm} = \frac{1}{K} \sum_{k=0}^{K-1} F_{nm}(VOP_k) \quad (9)$$

Pričom  $O^2_{nm}$  je pohybový deskriptor, ktorý je invariálny voči rotácií

## 3. Deskriptory založené na Fourierovej transformácii

Pre výpočet deskriptorov založených na Fourierovej transformácii sú potrebné hrany daného objektu [6]. Každý  $k$ -ty bod (pixel) je reprezentovaný súradnicami  $x(k), y(k)$ . Pre  $N$  bodov opisujúcich hrany daného objektu môžeme vyjadriť pomocou komplexných čísel (reálna časť komplexného čísla bude reprezentovať  $x$  os a imaginárna časť  $y$  os) nasledovne:

$$z(k) = x(k) + j.y(k), k = 0, 1, 2, 3, 4, 5, \dots, N-1 \quad (10)$$

V prípade, keď hrana daného objektu je uzatvorená, bude sa jednať o periodickú funkciu s periódou  $N$ . Fourierové deskriptory  $a(n)$  pre takúto hrany potom vypočítame nasledovne:

$$a(n) = \frac{1}{N} \sum_{k=0}^{N-1} z(k) e^{-j \frac{2\pi.n.k}{N}}, n = 0, 1, 2, 3, 4, 5, \dots, N-1 \quad (11)$$

Uvedené Fourierové deskriptory  $a(n)$  sú však závislé od rotácie, translácie, mierky ale aj od štartovacieho bodu pre opis hrany. Invariantnosť voči rotáciám a transláciám je však možné zaručiť ignorovaním nultej zložky (jednosmernej zložky)  $a(0)$  a imaginárnych častí ostatných zložiek (imaginárne časti v tomto prípade nesú informácie o rotácií). Invariantnosť voči mierke je možné dosiahnuť vydelením všetkých zložiek  $a(n)$  zložkou  $a(1)$ . Invariantnosť voči štartovaciemu bodu pre opis hrany môžeme dosiahnuť odčítaním fázy  $e^{j\theta_1}$  zložky  $a(1)$  hodnotou  $n$ . Aplikovaním uvedených pravidiel pre zaručenie invariantnosti, dostávame nasledovný vzorec pre výpočet Fourierových deskriptorov (invariantných od rotácie, translácie, mierky ale aj od štartovacieho bodu pre opis hrany)  $fd$ :

$$fd = \left[ \frac{a(n)}{a(1)} e^{-j\theta_1.n} \right], n = 2, 3, 4, 5, \dots, N-1 \quad (12)$$

Výhodou uvedeného deskriptora  $fd$  je riešenie problému 2D ( $D$  - dimenzia) v 1D.

Iný prístup výpočtu Fourierových deskriptorov je publikovaný v práci [7], kde sa počíta so vzdialenosťou medzi bodmi hrany  $x(k), y(k)$  a centrálného bodu daného objektu, resp. centroidu  $(x_c, y_c)$ .

$$r(k) = \sqrt{(x(k) - x_c)^2 + (y(k) - y_c)^2} \quad (13)$$

Následne vypočítame Fourierové koeficienty  $b(n)$ .

$$b(n) = \frac{1}{N} \sum_{k=0}^{N-1} r(k) e^{-j \frac{2\pi.n.k}{N}}, n = 0, 1, 2, 3, 4, 5, \dots, N-1 \quad (14)$$

Pretože  $r(k)$  nie sú komplexné čísla, ale len reálne, koeficientov  $b(n)$  bude len polovica. Taktiež vzhľadom na výpočet koeficientov s uvažovaním centroidu, daný deskriptor je invariálny voči transláciám. Invariantnosť voči rotáciám, mierky ale aj od štartovacieho bodu pre opis hrany, môžeme dosiahnuť podobným spôsobom ako v prvom prípade.

$$fd_c = \left[ \frac{b(n)}{b(1)} \right], n = 1, 2, 3, 4, 5, \dots, \frac{N}{2} \quad (15)$$

V prípade, že nepotrebuje pracovať s detailmi, môžeme použiť len prvých  $n$  koeficientov Fourierovho deskriptora. Empiricky bolo ukázané, že obyčajne stačí len prvých 10 koeficientov (viď Obr. 2).

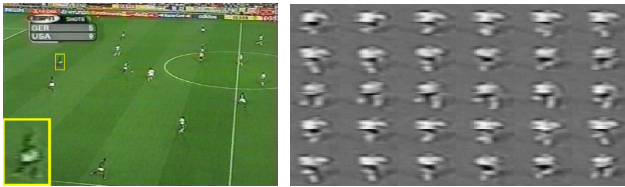


Obr. 2 Tvar človeka (vľavo) a rekonštrukcia hrán pomocou 10 koeficientov Fourierovho deskriptora

Fig. 2 Human body (left) and edge reconstruction by 10 coefficients of Fourier descriptor

## 4. Časovo-priestorový deskriptor

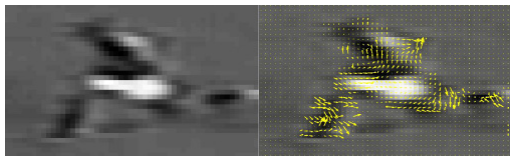
Princípom je spracovanie optického toku získaného nie zo spracovania zmien pixlov, ale pomocou časovo-priestorových deskriptorov pohybu, založených na priestorových šablonách. V práci [2] bola vytvorená metóda a program, ktorý môže klasifikovať určité pohyby v scéne ako napríklad beh, skok vo futbalovom zápase.



**Obr. 3** Typický obrázok (vľavo) z televízneho prenosu formátu NTSC. Vytvorený program klasifikoval kráčanie hráča smerom doprava. Sledovanie každého hráča a jeho stabilizácia v čase a v priestore (vpravo). Len takto upravené vstupné dáta sa ďalej používajú v algoritme (zvýšný vstupný obraz sa ignoruje)

**Fig. 3** Typical picture (left) of television broadcasting (NTSC format). Designed program classified the direction of walking soccer player. Tracing each soccer player can be done by time and space stabilization (right). Only after mentioned corrections, data are used in the algorithm (the rest of input data are ignored)

Vytvorila sa databáza obrázkov hráčov v daných akciách, tie boli následne popísané užívateľom. Táto databáza sa potom používala pre klasifikovanie vstupných dát. Pre popisanie pohybu sa využívajú deskriptory pohybu. Tieto deskriptory musia byť robustné vzhľadom na menej kvalitné vstupné dáta (nepresná poloha, nepresnosti v čase – priebeh pohybu). Pre odstránenie šumu z obrázku, je potrebné daný obrázok rozmazať, a až potom aplikovať algoritmus pre výpočet optického toku. V tomto prípade bol použitý algoritmus Lucas-Kanade [3].

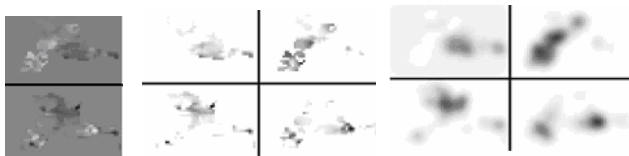


**Obr. 4** Originálny obraz (vľavo), obraz s vypočítaným optickým tokom (vpravo)

**Fig. 4** Original Picture (left), Picture with already calculated optical flow (right)

Optický tok  $F$  sa následne rozdelí na jeho horizontálnu zložku a vertikálnu ( $F_x, F_y$ ). Tieto komponenty optického toku sa ďalej rozdelia do 4 nezáporných kanálov nasledovne:

$$F_x = F_x^+ - F_x^- \quad F_y = F_y^+ - F_y^- \quad (16)$$



**Obr. 5** Rozdelený optický tok na horizontálnu a vertikálnu zložku  $F_x, F_y$  (vľavo), rozdelený optický tok do 4 nezáporných kanálov  $F_x^+, F_x^-, F_y^+, F_y^-$  (v strede), rozdelený a rozmazaný optický tok do 4 nezáporných kanálov  $Fb_x^+, Fb_x^-, Fb_y^+, Fb_y^-$  (vpravo)

**Fig. 5** The optical flow divided into horizontal and vertical component  $F_x, F_y$  (left), the optical flow divided into 4 non negative channels  $F_x^+, F_x^-, F_y^+, F_y^-$  (in the middle), the optical flow blurred and divided into 4 non negative channels  $Fb_x^+, Fb_x^-, Fb_y^+, Fb_y^-$  (right)

Nech uvedené štyri komponenty ( $F_x^+, F_x^-, F_y^+, F_y^-$ ) pre obraz  $i$  v sekvencii A sú  $a_1^i, a_2^i, a_3^i, a_4^i$ , potom podobnosť s obrazom  $j$  v sekvencii B sa vypočíta nasledovne:

$$S(i, j) = \sum_{t \in T} \sum_{c=1}^4 \sum_{x, y \in I} a_c^{i+t}(x, y) b_c^{j+t}(x, y) \quad (17)$$

Kde  $T$  a  $I$  sú časové a priestorové množiny pre jednotlivé pohybové deskriptory. Zjednodušenie tohto výpočtu môže byť realizované pomocou maticového počtu nasledovne: nech matica  $A_1$  obsahuje všetky  $a_1$  pre každý obraz v sekcii zoradené do stĺpcov, podobným spôsobom vytvoríme matice aj pre zvyšné komponenty. Podobnosť obrazov vo video sekvencii sa vypočíta nasledovne:

$$S_{ij} = A_1^T \cdot B_1 + A_2^T \cdot B_2 + A_3^T \cdot B_3 + A_4^T \cdot B_4 \quad (18)$$

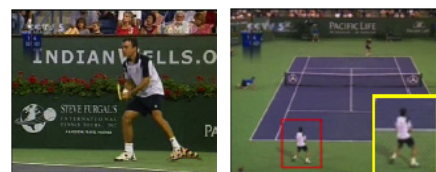
Ak sa hľadaná videosekvencia vyskytuje len zriedka, môžeme uplatniť na jej detekciu zjednodušený výpočet, ktorý je len hrubý odhad podobnosti. Spočíva na skutočnosti, že čím viac sa budú dáta v blízkom okolí zhodovať so vzorom, tým podobnosť so vzorom bude vyššia. Zjednodušenú podobnosť môžeme počítať nasledovne

$$K(i, j) = \sum_{r \in R} \omega(r) \chi(i, r, j) \quad (19)$$

Kde  $\omega(r)$  sú hodnoty váh

### 5. Deskriptor založený na optickom toku

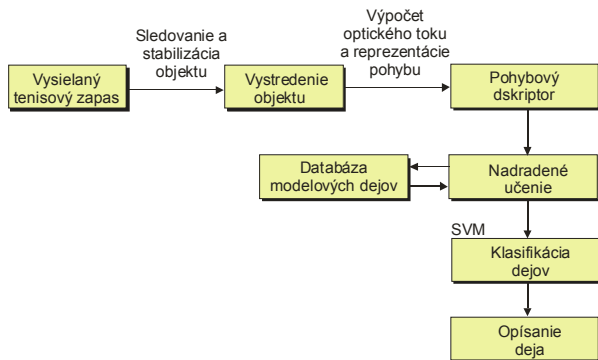
Metóda sa snaží klasifikovať úder hráča z videosekvencie (forhand, backhand) pomocou pohybových deskriptorov, ktoré sa vypočítali s optického toku. Uvedená metóda môže byť implementovaná aj v mobilnej robotike, kde je potrebné klasifikovať pohyb časti objektu naľavo od jeho ťažiska a naopak pohyb časti objektu napravo od svojho ťažiska. Optický tok je vyhladený Gaussovým filtrom, následne je rozdelený na obrazy, ktoré majú určité závislosti na samých seba v závislosti od pohybu rôznych častí tela. Takýto predspracovaný optický tok nazývame S-OFH (Slice Based on Optical Flow Histogram). S-OFH je následne spracovaný pomocou podporných vektorových počtov (SVM – Support Vector Machine) [4].



**Obr. 6** Klasifikácia úderov hráča z videosekvencie (for-hand, backhand)

**Fig. 6** Forehand/backhand classification of tennis player from video

V prvej fáze je potrebné hráča z obrazu sledovať a stabilizovať (Obr. 7) pomocou robustného algoritmu, pretože je potrebné minimalizovať vplyv šumu v obraze.



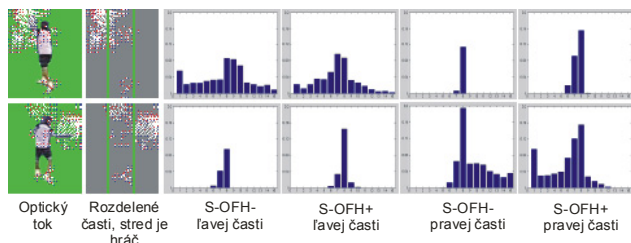
**Obr. 7 Princíp rozpoznávania akcií pomocou SVM a deskriptorov založených na optickom toku**

**Fig. 7 Block diagram of action classification by SVM (Support Vector Machine) and descriptors based on optical flow**

Existujúce metódy pre sledovanie tenisového hráča sú založené na porovnávaní vzorov, ktoré sú podobné metódam založených na korelácií. Tieto algoritmy nie sú však vhodné, pretože sú pomerne citlivé na šum, deformáciu objektu, dynamiku pozadia, nízke rozlíšenie obrazu. Z uvedeného dôvodu pomocou týchto metód je náročné sledovať objekt v dlhých videosekvenciách. Uvedený problém je možné eliminovať, ak budeme sledovať objekt v tzv. „dávках“ samostatne, t.j. v 30 za sebou idúcich obrazoch budeme sledovať objekt nezávisle. Je možné použiť sofistikovanú metódu nazývanú SVR (Support Vector Representation) „particle“ filter. Podrobnosti o tejto metóde je možné nájsť v práci [5]. Stabilizácia objektu sa realizovala pomocou vypočítania centroidu (stredu objektu, t.j.  $m_x, m_y$ ).

$$m_x = \frac{\sum_{x \in R} \sum_{y \in R} x \cdot f(x, y)}{\sum_{x \in R} \sum_{y \in R} f(x, y)}, \quad m_y = \frac{\sum_{x \in R} \sum_{y \in R} y \cdot f(x, y)}{\sum_{x \in R} \sum_{y \in R} f(x, y)} \quad (20)$$

Kde R je oblasť patriaca k objektu a  $f(x,y)$  je úroveň šedej v súradniciach  $x,y$ . Pomocou uvedeného sledovania a stabilizácie obrazu máme sledovaný obraz v strede pozorovacieho okna. Následne zo získaného pohybujúceho sa objektu je potrebné vypočítať optický tok, napríklad pomocou algoritmu Horn-Schunk. Pre získanie dôveryhodných pohybových deskriptorov je potrebné eliminovať vplyv šumu (napríklad potlačanie optického toku s nízkou amplitúdou, ktorý často vzniká v dôsledku šumu). Ten sa nachádza v každom obraze.



**Obr. 8 Optický tok rozdelený na časti okolo hráča a histogramy týchto častí, na základe ktorých sa nakoniec určuje typ úderu hráča (forhand/backhand)**

**Fig. 8 Optical flow is divided on components around tennis player. Histograms of these components are made as well, and finally based on these histograms we can classify forhand/backhand of the tennis player.**

## Záver

Spracovanie obrazu, videa a pohybu nie je triviálna úloha. V súčasnosti sa riešia len čiastkové úlohy. Je to oblasť, ktorá sa ešte vyvíja. Vývoj v oblasti hardvéru, kamerových systémov, ale aj štandardov MPEG, môže úlohám pre spracovanie obrazu, videa a pohybu pomôcť. Napríklad štandard MPEG7, ktorý bol vyvinutý v roku 2001, zohľadňoval potreby spracovania obsahu multimediálnych záznamov [9][10]. Práve tento štandard (MPEG7) poskytuje sústavu štandardizovaných nástrojov, pomocou ktorých je možné popísať a efektívne spravovať obsah multimediálnych záznamov. Uvedené štandardizované nástroje pre MPEG7 používajú niekoľko deskriptorov pre dosiahnutie svojich cieľov. Aj napriek veľkému úsiliu implementovať do tohto štandardu aj používanie deskriptorov v reálnom čase, sa v skutočnosti v tomto smere sa dosiahli len malé pokroky. V práci [11] je prezentované riešenie hardvérovej implementácie deskriptorov pre MPEG7, pričom toto riešenie je schopné pracovať v reálnom čase. Vývoj a optimalizácia deskriptorov pre spracovanie a klasifikovanie v reálnom čase pohybujúcich sa objektov tvorí dôležitú úlohu nielen v mobilnej robotike, ale aj pri vývoji multimediálnych štandardov. To dokazuje skutočnú potrebu po kvalitných deskriptoroch pri spracovaní obsahu multimediálnych záznamov.

Článok vznikol za podpory VEGA agentúry MŠ SR v projekte č. 1/3120/06.

## Literatúra

[1] Berna Erol a kol., Local Motion Descriptors, Department of Electrical and Computer Engineering, University of British Columbia

[2] Alexei A. Efros, Recognizing Action at a Distance, In IEEE International Conference on Computer Vision, Nice, France, October 2003

[3] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In DAP-RA Image Understanding Workshop, April 1981.

[4] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition, Kluwer Academic Publishers, Boston, USA, 1998

[5] Zhu G. a kol.. Improving Particle Filter with Support Vector Representation for Efficient Visual Tracking. IEEE International Conference on Image Processing. Volume 2. (2005) 422-425.

[6] Daniel Toth a Til Aach. Detection and recognition of moving objects using statistical motion detection and Fourier descriptors. Institute for Signal Processing, University of Luebeck, Germany

[7] D. S. Zhang a G. Lu. A comparative study on shape retrieval using Fourier descriptors with different shape signatures. International Conference on Intelligent Multimedia and Distance Education, pages 1-9, Fargo, ND, USA, June 1-3, 2001

[8] Jianzhong Fang a Guoping Qiu. Human Face Detection Using Angular Radial Transform and support vector machine. School of Computer Science, The University of Nottingham, United Kingdom

[9] S.F. Chang, A. Puri, T. Sikora, H. Zhang, Overview of the MPEG-7 Standard, IEEE Tran. Circ. Sys. Video Tech., vol. 11, pp. 688-695, June 2001.

[10] B. S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7 Multimedia Content Description Interface, J. Wiley, New York, 2002.

[11] Andreas Savakis, a kol., Real-time Video Annotation using MPEG-7 Motion Activity Descriptors, 1Department of Computer Engineering Rochester Institute of Technology Rochester, NY 14623, USA

### Abstract

It is necessary to have intelligence for mobile robot, by which robot is able to react in the environment where it exists. However, the environment can be so heterogeneous, that robot should have camera subsystem included. Video from the camera subsystem has much information useful for mobile robot controlling (not only information from static pictures, but also information from analyzing actions of

the object in the scene). In this paper we present some local descriptors that can be used for analyzing and for classification of moving objects.

### Ing. Peter Benický

Slovenská technická univerzita v Bratislave  
Fakulta elektrotechniky a informatiky  
Ústav riadenia a priemyselnej informatiky  
Ilkovičová 3  
81219 Bratislava  
e-mail: peter@benicky.info

### prof. Ing. Ladislav Jurišica, PhD.

Slovenská technická univerzita v Bratislave  
Fakulta elektrotechniky a informatiky  
Ústav riadenia a priemyselnej informatiky  
Ilkovičová 3  
81219 Bratislava  
e-mail: ladislav.jurisica@stuba.sk