Zbornik radova 49. Konferencije za ETRAN, Budva, 5-10. juna 2005, tom I
Proc. 49th ETRAN Conference, Budva, June 5-10, 2005, Vol. I

DETEKCIJA I ESTIMACIJA POZICIJE POKRETNIH OBJEKATA IZ VIDEO SEKVENCI

Igor Jovandić, Branko Kovačević, *Elektrotehnički fakultet u Beogradu*

Sadržaj – *Jedan metod detekcije i estimacije pozicije pokretnih objekata na osnovu video sekvenci izložen je u ovom radu. Dat je kratak pregled teorije transformacija koordinatnih sistema, sa naglaskom na inverznu transformaciju perspektive. Opisan je problem kalibracije kamere i predloženo je rešenje bazirano na metodu najmanjih kvadrata. Dat je opis izvršenog eksperimenta, prikazani su i komentarisani rezultati. Istaknuta je jednostavnost predstavljenog metoda, što ga čini primenljivim u real-time aplikacijama.*

1. UVOD

U ovom radu predstavljen je metod za detekciju i estimaciju pozicije pokretnih objekata na osnovu video sekvenci. Primenjen je numerički jednostavan način detekcije objekta, baziran na analizi sadržaja boje digitalne slike. U cilju estimacije pozicije objekta u realnom svetu, na osnovu procene pozicije objekta na digitalnoj slici, primenjen je geometrijski pristup [1,2]. Pokazano je da se afine prostorne transformacije i transformacija perspektive mogu upotrebiti za rešavanje problema estimacije pozicije objekta. Usvojeni geometrijski pristup rezultuje rešenjem u analitičkoj formi, ali je rešenje teško praktično primenljivo zbog velikog broja nepoznatih parametara. Procena nepoznatih parametara opisanog problema je u literaturi poznata pod nazivom kalibracija kamere [1,2,3].

Metod predstavljen u ovom radu, primenjen je za detekciju i estimaciju pozicije mobilnog robota koji se kreće u ravni. Prikazani su i diskutovani rezultati i dat je komentar o mogućnostima dalje primene predstavljenog metoda, pre svega u cilju zatvaranja povratne sprege na osnovu video sekvenci.

2. TRANSFORMACIJE KOORDINATNIH SISTEMA

Koordinatni sistemi su neophodni da bi se kvantitativno odredila pozicija tačke u prostoru. Cilj transformacije koordinatnog sistema je uspostavljanje veze između mera pozicije tačke u različitim koordinatnim sistemima.

U cilju digitalne obrade slike, a naročito ako se koriste afine transformacije, korisno je tačku u prostoru predstaviti homogenim koordinatama [2]. Homogene koordinate tačke $p = \begin{bmatrix} x & y & z \end{bmatrix}^T$ definisane su sa:

$$\begin{bmatrix} sx & sy & sz & s \end{bmatrix}^T \tag{1}$$

gde je s - skalirajući faktor. Tačka predstavljena notacijom u jednačini (1) zove se homogena tačka. Široka klasa prostornih transformacija može se predstaviti kao proizvod konstantne matrice i homogene tačke.

Afina transformacija je ona koja ne menja osobinu kolinearnosti tačaka (tačke koje leže na pravoj pre transformacije, ostaju na pravoj i nakon transformacije) i odnos distance tačaka (srednja tačka duži ostaje srednja tačka duži i nakon transformacije). Posledica ove definicije je da originalno paralelne prave, ostaju paralelne i nakon afine transformacije. Osnovne afine transformacije su: translacija, rotacija, skaliranje, smicanje i refleksija. Svaka afina transformacija može se predstaviti kompozicijom navedenih osnovnih transformacija [1,2,3].

Ako se posmatraju dva koordinatna sistema koji su translirani jedan u odnosu na drugi za x_0, y_0 i z_0 po x, y i z osi, respektivno, i ako su homogene koordinate tačke p u koordinatnom sistemu $\{x, y, z\}$ date sa $\begin{bmatrix} x & y & z & 1 \end{bmatrix}^T$, tada se homogene koordinate iste tačke u transliranom koordinatnom sistemu $\{x_t, y_t, z_t\}$ mogu dobiti prostim matričnim množenjem:

$$\begin{bmatrix} sx_t \\ sy_t \\ sz_t \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{M_t} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2}$$

Matrica M_t u jednačini (2) predstavlja matricu translacije koordinatnog sistema.

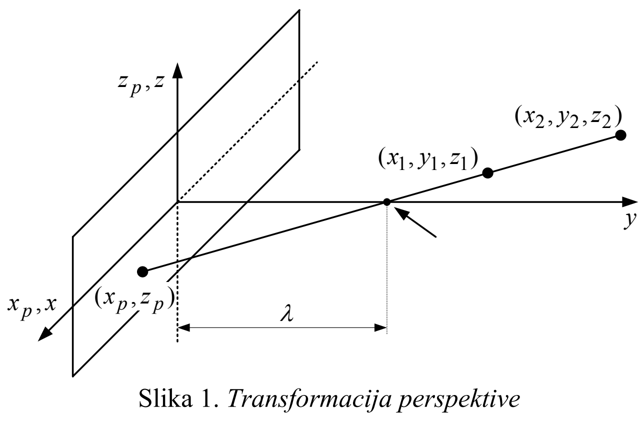
Slično jednačini (2), matrično množenje može se iskoristiti u cilju reprezantacije rotacije koordinatnog sistema. Rotacija oko proizvoljne ose u prostoru može se predstaviti opštom matricom rotacije, datom u jednačini (3):

$$M_r = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Matrica koeficijenta r_{ij} $(i, j \in \{1,2,3\})$ u jednačini (3) je ortonormalna: svaka kolona (vrsta) predstavlja jedan vektor jediničnog intenziteta, a vektori su uzajamno ortogonalni.

Treba primetiti da je poslednja (četvrta) vrsta matrice transformacije u jednačinama (2) i (3) uvek $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$, pa je skalirajući faktor s jednak jedinici. Ova napomena važi za svaku afinu transformaciju.

Bitna transformacija u digitalnoj obradi slike je **transformacija perspektive**. Slikom 1 prikazan je model perspektivne projekcije tačke u prostoru na ravan $\{x_p, z_p\}$. Ako se posmatra tačka u $\{x, y, z\}$ koordinatnom sistemu, od interesa je ustanoviti vezu između koordinata (x_1, y_1, z_1) posmatrane tačke i koordinata projekcije tačke (x_p, z_p).



Slika 1. *Transformacija perspektive*

Na osnovu sličnosti trouglova sa slike 1, dobija se:

$$\frac{x_p}{\lambda} = -\frac{x_1}{y_1 - \lambda} \tag{4}$$

$$\frac{z_p}{\lambda} = -\frac{z_1}{y_1 - \lambda} \tag{5}$$

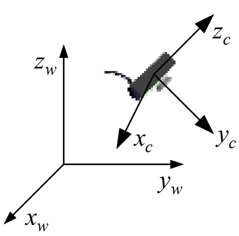
gde je λ - žižna daljina sočiva kamere. Uvodeći reprezentaciju tačke u homogenim koordinatama (1), jednačine (4) i (5) mogu se napisati u sledećoj matričnoj formi [2]:

$$\begin{bmatrix} sx_p \\ sy_p \\ sz_p \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1/\lambda & 0 & 1 \end{bmatrix}}_{M_p} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \tag{6}$$

Može se primetiti da je y koordinata tačke na $\{x_p, z_p\}$ ravni $y_p = 0$. Takođe, treba primetiti da se tačka sa koordinatama (x_2, y_2, z_2) sa slike 1, preslikava u istu tačku na $\{x_p, z_p\}$ ravni kao i tačka (x_1, y_1, z_1). Poslednje ima za posledicu nemogućnost inverzije transformacije perspektive: ako su poznate koordinate projekcije tačke (x_p, z_p), nemoguće je na osnovu njih jednoznačno odrediti koordinate originalne tačke (x_1, y_1, z_1). Međutim, ako je jedna od koordinata tačke (x_1, y_1, z_1) poznata, i ako su poznate koordinate (x_p, z_p), druge dve koordinate originalne tačke mogu se jednoznačno odrediti polazeći od jednačine (6). Transformacija perspektive nije afina transformacija, jer paralelne prave preslikava u prave koje se seku. To se može videti i inspekcijom forme matrice transformacije perspektive M_p u jednačini (6): poslednja (četvrta) vrsta nema oblik $[0 \quad 0 \quad 0 \quad 1]$, pa je skalirajući faktor $s \neq 1$ u opštem slučaju.

3. INVERZNA TRANSFORMACIJA PERSPEKTIVE

Osnovni problem u estimaciji pozicije objekata iz video sekvenci je nalaženje veze između koordinata objekta na digitalnoj slici i koordinata objekta u realnom svetu. U tom cilju, potrebno je odrediti vezu između koordinata tačke u koordinatnom sistemu prostora u kome se objekat kreće i koordinata iste tačke u koordinatnom sistemu vezanom za kameru. Relativna pozicija navedenih koordinatnih sistema prikazana je na slici 2.



Slika 2. *Koordinatni sistemi problema*

Kompozicija afinih transformacija može se iskoristiti u cilju nalaženja opisane veze koordinata. Prva transformacija je translacija koordinatnog sistema i opisana je matricom M_t u jednačini (2). Sledeće dve transformacije predstavljaju rotaciju transliranog sistema, prvo oko x ose (za ugao θ_x), a zatim oko z ose (za ugao θ_z). Pošto je tačka rotacije kamere oko x i z ose u opštem slučaju različita od tačke centra sočiva kamere, potrebno je izvršiti i translaciju koordinatnog sistema u pravcu z ose za iznos r koji predstavlja rastojanje tačke oko koje se kamera može rotirati i centra sočiva kamere (centriranje slike). Kompozicija četiri navedene transformacije data je jednačinom (7):

$$\begin{bmatrix} sx_c \\ sy_c \\ sz_c \\ s \end{bmatrix} = \underbrace{M_{tr} \cdot M_{rz} \cdot M_{rx} \cdot M_t}_{M} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{7}$$

Treba primetiti da je redosled množenja matrica u proizvodu datom u (7) inverzan u odnosu na redosled primene transformacija. Matrica transformacije M u jednačini (7) je oblika:

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

gde su koeficijenti m_{ij} poznate nelinearne funkcije parametara x_0, y_0, z_0, θ_x, θ_z i r. Lako je pokazati da proizvoljna kompozicija transformacija translacije i rotacije ima matricu transformacije iste forme kao matrica M u jednačini (8) [2].

Pošto se lik objekata projektuje na senzor slike digitalne kamere shodno modelu perspektivne projekcije (slika 1), sledi da nad koordinatama tačke objekta (x_c, y_c, z_c) u koordinatnom sistemu $\{x_c, y_c, z_c\}$ sa slike 2, treba izvršiti transformaciju perspektive. Pošto je y koordinata perspektivne projekcije tačke uvek jednaka nuli, i pošto množenje svakog elementa matrice transformacije konstantom c ne menja samu transformaciju, jednačina transformacije postaje:

$$\begin{bmatrix} sx_p \\ sz_p \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & 1 \end{bmatrix}}_{T_1} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{9}$$

Pošto se koordinate tačke na digitalnoj slici obično izražavaju u pikselima, a referentni smer vertikalne ose je suprotan usvojenom referentnom smeru ose z_p (drugim rečima, tačka sa koordinatama (0,0) nalazi se u gornjem levom uglu slike), potrebno je matricu T_1 komponovati sa matricom skaliranja [1,2], što opisuje jednačina (10):

$$\begin{bmatrix} sv \\ sk \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & -1/d_y & 0 \\ 1/d_x & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot T_1}_{T_2} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{10}$$

Forma matrice T_2 data je jednačinom (11). Veličine v (vrsta) i k (kolona) izražene su u pikselima i predstavljaju poziciju tačke (x_w, y_w, z_w) iz koordinatnog sistema $\{x_w, y_w, z_w\}$ na digitalnoj slici. Skalirajući faktori d_x i d_y predstavljaju horizontalnu i vertikalnu veličinu piksela u jedinicama mere koordinatnog sistema $\{x_w, y_w, z_w\}$ (npr. u metrima).

$$T_2 = \begin{bmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & 1 \end{bmatrix} \tag{11}$$

Očigledno je nemoguće jedinstveno odrediti vrednosti (x_w, y_w, z_w) na osnovu poznavanja vrednosti (v, k). Međutim, ako je jedna od koordinata objekta poznata (npr. $z_w = H$), jednačinu (11) moguće je napisati u sledećoj formi:

$$\begin{bmatrix} sv \\ sk \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} g_{11} & g_{12} & g_{13}H + g_{14} \\ g_{21} & g_{22} & g_{23}H + g_{24} \\ g_{31} & g_{32} & g_{33}H + 1 \end{bmatrix}}_{T_3} \cdot \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \tag{12}$$

Sada je koordinate tačke (x_w, y_w) moguće odrediti kao:

$$\begin{bmatrix} sx_w \\ sy_w \\ s \end{bmatrix} = T_3^{-1} \cdot \begin{bmatrix} v \\ k \\ 1 \end{bmatrix} \tag{13}$$

4. KALIBRACIJA KAMERE

Iako je analitički izraz za svaki od koeficijenata matrice T_2 u jednačini (11) poznat, u jednačinama koje određuju te koeficijente figuriše 9 nepoznatih parametara: x_0, y_0, z_0, θ_x, θ_z, r, λ, d_x i d_y. Te jednačine predstavljaju nelinearne funkcije nepoznatih parametara, pa bi rešenje za procenu tih parametara podrazmevalo minimizaciju pogodno definisanog kriterijuma. Ako bi se takav algoritam optimizacije i primenio, dobijene procene parametara bile bi upotrebljene za izračunavanje koeficijenata matrice T_2 u (11). Međutim, ako je krajnji cilj kalibracije kamere nalaženje koeficijenata matrice T_2, onda nema nikakve potrebe eksplicitno određivati navedene parametre.

Podaci na osnovu kojih se vrši kalibracija kamere predstavljaju mere pozicije određenih referentnih tačaka - markera, čije su koordinate u sistemu prostora u kome se objekat kreće poznate: (x_w^i, y_w^i, z_w^i), $i = 1,2,...,N$, gde je N - broj markera. Koordinate markera na digitalnoj slici (v^i, k^i), mogu se lako odrediti ako se za markere izaberu objekti karakterističnog oblika i/ili boje. Ako se sada jednačina (10) napiše u sledećoj formi:

$$v = (g_{11} - g_{13}v)x_w + (g_{12} - g_{32}v)y_w + (g_{13} - g_{33}v)z_w + g_{14} \tag{14}$$

$$k = (g_{21} - g_{31}k)x_w + (g_{22} - g_{32}k)y_w + (g_{23} - g_{33}k)z_w + g_{24} \tag{15}$$

jasno je da za bilo koji izbor parametara g_{ij}, jednačine (14) i (15) neće biti zadovoljene za svako (v^i, k^i) i (x_w^i, y_w^i, z_w^i). Razlog tome su greške u proceni pozicije markera na slici i greške merenja u koordinatnom sistemu prostora u kome se objekat kreće. Međutim, ako se jednačine (14) i (15) napišu u sledećoj matričnoj formi:

$$A_i^T \cdot q = B^i, \tag{16}$$

može se primeniti metod najmanjih kvadrata u cilju procene parametara matrice T_2 [2]. Pošto nepoznatih parametara ima 11, sledi da je za primenu metode najmanjih kvadrata nepohodno minimalno $N = 6$ markera. Za svaki od markera može se napisati matrična jednačina tipa (16), pa se tih N jednačina može napisati u sledećoj, integralnoj formi: $A \cdot q = B$. Matrice A_i, B_i, A i B, kao i vektor parametara q, dati su sledećim jednačinama:

$$A_i = \begin{bmatrix} x_w^i & 0 \\ y_w^i & 0 \\ z_w^i & 0 \\ 1 & 0 \\ 0 & x_w^i \\ 0 & y_w^i \\ 0 & z_w^i \\ 0 & 1 \\ -v^i x_w^i & -k^i x_w^i \\ -v^i y_w^i & -k^i y_w^i \\ -v^i z_w^i & -k^i z_w^i \end{bmatrix}, \quad q = \begin{bmatrix} g_{11} \\ g_{12} \\ g_{13} \\ g_{14} \\ g_{21} \\ g_{22} \\ g_{23} \\ g_{24} \\ g_{31} \\ g_{32} \\ g_{33} \end{bmatrix}, \quad A = \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_N^T \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_N \end{bmatrix}, \quad B_i = \begin{bmatrix} v^i \\ k^i \end{bmatrix}$$

Procena parametara matrice T_2 po metodu najmanjih kvadrata data je jednačinom (17):

$$\hat{q} = \left(A^T A\right)^{-1} A^T B \tag{17}$$

5. EKSPERIMENT I REZULTATI

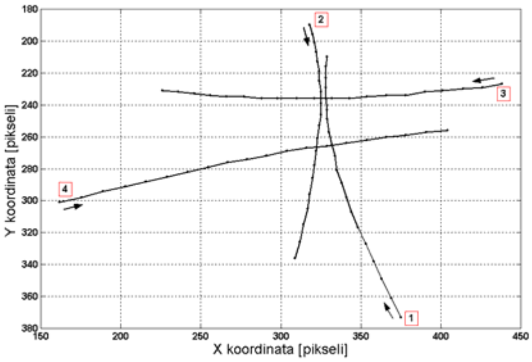
Primena opisanog metoda estimacije pozicije pokretnih objekata demonstrirana je na primeru mobilnog robota sa diferencijalnim pogonom (K-Team Hemisson) koji se kreće u ravni. U svrhu detekcije pozicije robota upotrebljena je klasična USB web kamera (Creative WebCam NX Ultra) sa mogućnošću generisanja RGB slike u rezoluciji 640×480 piksela.

Pozicija robota detektovana je na osnovu analize sadržaja boje digitalne slike, što je moguće uraditi samo u slučaju kada se boja robota (u ovom slučaju plava) značajno razlikuje od boje pozadine (žuta). Pošto je robot cilindričnog oblika, procena njegove pozicije dobijena je računanjem koordinata centra mase onih piksela koji zadovoljavaju sledeću nejednakost:

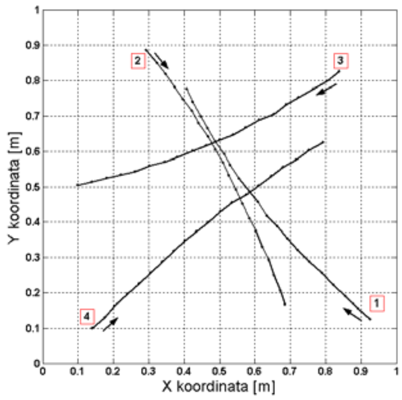
$$J(v,k) = b(v,k) - \frac{r(v,k) + g(v,k)}{2} > P, \tag{18}$$

gde su $r(v,k)$, $g(v,k)$ i $b(v,k)$ intenziteti crvene, zelene i plave komponente boje piksela sa koordinatama (v,k), respektivno. Vrednost praga P eksperimentalno je određena kako bi se dobilo zadovoljavajuće potiskivanje piksela pozadine. Kalibracija kamere izvršena je shodno jednačini (17), pri čemu je broj markera $N = 8$. Četiri markera nalaze se u ravni kretanja robota i predstavljaju temena kvadrata stranice $1\,[m]$, dok su preostala četiri markera, nenulte visine, postavljena na sredinu svake od stranica kvadrata. Markeri su karakterističnog oblika i boje, pa je njihova detekcija izvršena slično kao u jednačini (18). Veličina H u jednačini (12) predstavlja visinu robota i iznosi $H = 0.065\,[m]$.

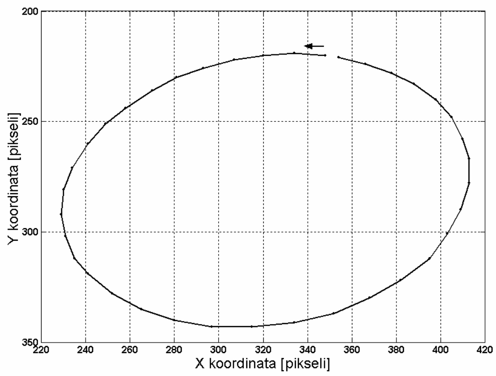
Pre izvođenja eksperimenta izvršena je procena pozicije statičkog objekta; dobijeni rezultati mogu se smatrati zadovoljavajućim, jer greška procene svake od koordinata x_w i y_w ne prelazi $0.02\,[m]$. Eksperiment je imao za cilj ispitivanje mogućnosti primene opisanog algoritma u estimaciji pozicije pokretnog objekta. Slikama 3 i 4 prikazani su rezultati procene pozicije robota koji se kreće približno pravolinijskom putanjom. Pozicija robota estimirana je svakih $0.5\,[s]$, što je sasvim zadovoljavajuće za manje zahtevne primene.



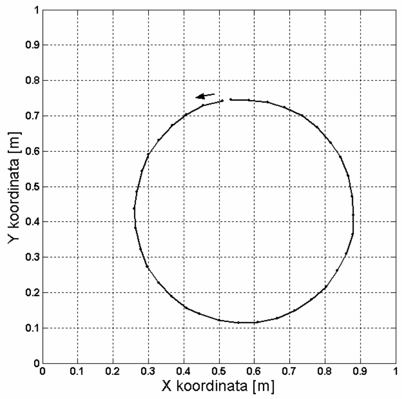
Slika 3. *Trajektorije u ravni kamere*



Slika 4. *Estimirane trajektorije sa slike 3*



Slika 5. *Kružna trajektorija u ravni kamere*



Slika 6. *Estimirana kružna trajektorija*

Iako je robotu zadata pravolinijska trajektorija, zbog slabe fiksacije osovine desnog točka (nepostojanje pravih kugličnih ležajeva), ista ispoljava ekscentrično rotaciono kretanje, što ima za posledicu skretanje robota u desnu stranu, a jasno se vidi sa slika 3 i 4. I pored opisanog konstrukcionog problema, ponavljanjem eksperimenta uz sporadične intervencije na fiksaciji ose rotacije desne osovine, dobijena je kružna trajektorija, što je provereno vizuelnom inspekcijom. Rezultati su prikazani slikama 5 i 6.

6. ZAKLJUČAK

Metod za detekciju i procenu pozicije pokretnih objekata na osnovu video sekvenci, predstavljen u ovom radu, dekupluje operacije detekcije pokretnog objekta i procene njegove pozicije. Iako je predložen jednostavan metod detekcije objekta (baziran na analizi sadržaja boje digitalne slike), moguće je upotrebiti i složenije algoritme (npr. detekcija prepoznavanjem oblika).

Algoritam estimacije pozicije objekta je numerički veoma jednostavan, što omogućava njegovu *real-time* primenu u zadacima praćenja i vođenja. Može se reći da su greške procene veoma male, imajući u vidu uslove pod kojim je eksperiment izvršen. Realizacija predloženog postupka kalibracije kamere veoma je jednostavna i numerički nezahtevna, što otvara mogućnost primene opisanog metoda u aplikacijama sa pokretnom kamerom.

Osnovni nedostatak algoritma estimacije je neophodnost poznavanja jedne od koordinata pokretnog objekta (čime je opseg problema sužen na ravanska kretanja) ili veze dveju njegovih koordinata (čime se uvodi restrikcija mogućih trajektorija objekta). Nedostatak se može prevazići upotrebom dve kamere (stereo vid), pri čemu se kalibracija svake od njih vrši postupkom koji je predložen u ovom radu.

Opisani metod detekcije i estimacije pozicije pokretnih objekata ima direktnu primenu u rešavanju problema pozicioniranja, vođenja i praćenja na osnovu video sekvenci (zatvaranje povratne sprege po kameri).

LITERATURA

[1] D.H. Ballard, C.M. Brown, *Computer Vision*, Prentice-Hall, 1982.

[2] L.G. Shapiro, G.C. Stockman, *Computer Vision*, Prentice-Hall Inc, New Jersey, 2001.

[3] R. Jain, R. Kasturi, B.G. Schunck, *Machine Vision*, McGraw-Hill, 1995.

Abstract – A method for detection and position estimation of moving objects from video sequences is presented in this paper. Short review of coordinate systems transformation theory is given, with the emphasis on perspective transformation. Camera calibration problem is defined and the solution based on the least squares method is presented. The performed experiment is described and the results are given and commented. The simplicity of the method is emphasized as a prerequisite for any real-time application.

DETECTION AND POSE ESTIMATION OF MOVING OBJECTS FROM VIDEO SEQUENCES

Igor Jovandić, Branko Kovačević