

به نام خدا



دانشکده مهندسی برق و کامپیوتر

دانشگاه دولتی سمنان

سمینار کارشناسی ارشد

رشته ی هوش مصنوعی

آنالیز خودکار ویدیهای فوتبال

استاد مربوطه:

جناب آقای دکتر رحمانی منش

توسط:

گل افروز داودی فر

زمستان سال 1396

## فهرست مطالب

5	چکیده .....
6	مقدمه .....
6	تعاریف و روش های اولیه .....
6	فریم : .....
6	شات: .....
7	ریخت شناسی : .....
7	فرسایش : .....
7	گسترش : .....
8	باز کردن : .....
8	بستن : .....
9	فیلتر کالمن : .....
10	انتقال هاف: .....
10	انتقال هاف خطی : .....
11	انتقال هاف دایره ای : .....
12	شناسایی لبه ها: .....
12	فیلتر سوپل: .....
13	فیلتر کنی: .....
13	فاصله اقلیدسی: .....
13	فاصله باتاچاریا: .....
14	تطبیق بلوکی : .....
14	مدل مارکوف: .....
15	فاصله ی انحراف زمانی پویا : .....
15	تشخیص توپ در ویدیوهای فوتبال منتشر شده .....
23	تشخیص و ردیابی بازیکنان در ویدیوهای فوتبال منتشر شده .....

27	.....شناسایی وقایع و رویدادها در ویدیو های فوتبال
30	.....نتیجه گیری:
30	.....منابع:

7	شکل 1) نمونه ای از عملگر فرسایش
8	شکل 2) نمونه ای از عملگر گسترش
8	شکل 3) نمونه ای از عملگر باز کردن
9	شکل 4) نمونه ای از عملگر بستن
10	شکل 5) مراحل فیلتر کالمن
11	شکل 6) مثالی از انتقال هاف خطی
12	شکل 7) مثالی از انتقال هاف دایره ای
12	شکل 8) عملگرهای فیلتر سوبل
13	شکل 9) تصویر خروجی فیلتر سوبل
13	شکل 10) تصویر خروجی فیلتر کنی
14	شکل 11) تطبیق بلوکی
15	شکل 12) یک نمونه از مدل مارکوف با سه حالت 1-بارانی 2-خورشیدی و 3-ابری
16	شکل 13) منطقه ی مورد بررسی در میدان فریم
18	شکل 14) الگوریتم تشخیص توپ در دو نمای متفاوت
20	شکل 15) الگوریتم تشخیص توپ در فریم های متوالی
21	شکل 16) کنترل سرعت در فیلتر کالمن
21	شکل 17) خطوط میدان شناسایی شده
23	شکل 18) مراحل تشخیص توپ و بازیکنان در زمین بازی
24	شکل 19) نتایج بخش بندی بر اساس مقادیر برجستگی
25	شکل 20) شناسایی بازیکنان فوتبال
27	شکل 21) الگوریتم به روز رسانی مناطق مرجع
28	شکل 22) معماری چند فیلتره برای تشخیص شات
29	شکل 23) صحنه ی مربوط به میانه ی زمین
29	شکل 24) صحنه ی مربوط به نمای نزدیک
29	شکل 25) صحنه ی مربوط به مردم

## چکیده

فوتبال و علم مربوط به آن به صورت روزافزون در حال گسترش است. به همین ترتیب با افزایش ویدیوهای فوتبال به تدریج آنالیز و تحلیل خودکار ویدیوهای فوتبال جایگزین روش های دستی می شوند. مهم ترین و عمده ترین مراحل برای آنالیز و تحلیل خودکار ویدیوها فوتبال، شناسایی و ردیابی توپ و بازیکن و همچنین استخراج رویدادهای موجود می باشد. در این مطالعه به روش های مختلف شناسایی و ردیابی توپ، شناسایی و ردیابی بازیکن و استخراج رویدادها از ویدیو های مربوط به فوتبال پرداخته می شود. همچنین در این مقاله به دلیل در دسترس بودن و عمومیت بیشتر از مطالعات انجام شده بر روی ویدیوهای منتشر شده ی مسابقات فوتبال ، به منظور تحلیل و آنالیز خودکار استفاده می شود.

فوتبال یک بازی متشکل از دو تیم یازده نفره می باشد که در آن بازیکنان می توانند از تمام اعضای بدنشان به جز دست ها و بازوها برای وارد کردن توپ درون دروازه حریف استفاده کنند. بازیکنان براساس وظایفی که بر عهده دارند پست های مختلفی را به عهده می گیرند مانند مدافع، مهاجم، دروازه بان و غیره. تنها دروازه بان مجاز است در محوطه ای که ناحیه پنالتی 1 نامیده می شود از دست هایش استفاده کند. در نهایت تیمی که توپ های بیشتری را درون دروازه وارد کرده باشد به عنوان برنده اعلام می شود. [1]

فوتبال اکنون محبوب ترین بازی در جهان می باشد که میلیون ها نفر آن را بازی کرده و صدها میلیون نفر دیگر از آن لذت می برند [2]. تخمین زده می شود تا اواخر قرن بیست و یکم میلادی بیش از 250 میلیون بازیکن فوتبال و 1.3 میلیارد علاقمند به فوتبال وجود داشته باشد. در سال های اخیر فوتبال به عنوان یک صنعت و یک بخش از تجارت و کسب و کار شناخته می شود [3]. این محبوبیت، مجذوبیت تحلیل و آنالیز [2] ویدیوهای فوتبال در بین محققان را در پی داشته است [4]. تجزیه و تحلیل خودکار ویدیو های فوتبال یک پاسخ الزامی به رشد فزاینده تقاضا توسط متخصصین و هواداران فوتبال، برای استخراج اطلاعات مورد نیاز می باشد [5] و با توجه به رشد انفجاری ویدیوهای فوتبال پردازش خودکار تصاویر فوتبال یک جایگزین مناسب برای راه حل های دستی می باشد [7]. همچنین از داده های عددی برای ارزیابی کیفیت بازی استفاده می شود. [6] مهم ترین مسائلی که در پردازش خودکار ویدیو مطرح می باشد شناسایی و ردیابی توپ و بازیکنان است که باعث به وجود آمدن رویدادهای مختلف در فوتبال می شوند. بنابراین در این مطالعه به روش های تشخیص و ردیابی توپ و بازیکنان و همچنین استخراج رویداد از ویدیوهای منتشر شده فوتبال پرداخته می شود.

## تعاریف و روش های اولیه

### فریم<sup>1</sup>:

به طور کلی فریم ها کوچک ترین عضو یک ویدیو هستند که یک تصویر را در برمی گیرند که برای تحلیل سطح پایین مورد استفاده قرار می گیرند.

### شات<sup>2</sup>:

شات مجموعه ای از فریم های متوالی است که به صورت پیوسته از یک دوربین گرفته شده اند، به طوری که مشخصات دوربین مانند وضعیت زوم، نوع نما و... ثابت است یا تغییرات کندی دارد. با آشکارسازی مرز بین شاتها، بخش بندی ویدیو در سطح پایین انجام می شود. بنابراین آشکارسازی مرز بین شاتها به عنوان یک پردازش سطح پایین می تواند در پردازشهای سطح بالاتر مفید واقع شود. تغییر شات معمولاً به صورت ناگهانی و در یک قاب انجام می شود که به آن گذار ناگهانی یا برش 1 می گویند، اما ممکن است

به لحاظ زیبایی شناسی، تغییر شات به صورت تدریجی و در طی چند قاب انجام شود.

<sup>1</sup> frame

<sup>2</sup> shot

شات ها از لحاظ نما و چشم انداز عمدتا به سه دسته تقسیم می شود: نمای دور، نمای معمولی یا متوسط یا نمای نزدیک.

ریخت شناسی<sup>۱</sup>:

ریخت شناسی بسیار مانند اعمال فیلترهای زمانی-مکانی است که ساختار موردنظر همه ی پیکسل های تصویر را پوشش کرده و در نهایت باتوجه به اوپراتور اعمال شده تصویر جدیدی به وجود می آید. عملگرهای فرسایش، کاهش، باز کردن و بستن از جمله عملگرهای ریخت شناسی می باشند.

فرسایش<sup>۲</sup>:

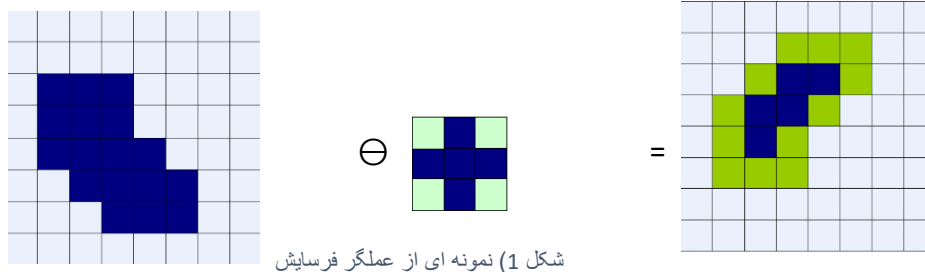
اگر  $A$  و  $B$  مجموعه هایی در  $Z$  باشند، فرسایش  $A$  و  $B$  که با  $A \ominus B$  نمایش داده می شود، به صورت زیر تعریف می شود، به صورت زیر تعریف می گردد:

$$A \ominus B = \{z / (B)_z \subseteq A\}$$

این عبارت بیان می کند که فرسایش  $A$  و  $B$ ، مجموعه تمام نقاط  $z$  به نحوی است که  $B$ ، که با  $z$  انتقال یافته، در  $A$  واقع باشد.  $B$  را یک عضو ساختاری می نامیم. می توانیم فرسایش را به شکل معادله زیر نمایش دهیم:

$$A \ominus B = \{z / (B)_z \cap A^c = \emptyset\}$$

که در این فرمول  $A^c$  متمم مجموعه  $A$  و  $\emptyset$  مجموعه تهی می باشد. شکل زیر یک نمونه از فرسایش را نشان می دهد.



گسترش<sup>۳</sup>:

اگر  $A$  و  $B$  مجموعه هایی در  $Z$  باشند، فرسایش  $A$  و  $B$  که با  $A \oplus B$  نمایش داده می شود، به صورت زیر تعریف می شود، به صورت زیر تعریف می گردد:

$$A \oplus B = \{z / (\hat{B})_z \cap A \neq \emptyset\}$$

گسترش  $A$  با  $B$ ، مجموعه تمام جابه جایی های  $z$  است، به نحوی که  $B$  و  $A$  حداقل یک جزء مشترک داشته باشند. می توانیم گسترش را به شکل معادله زیر نمایش دهیم:

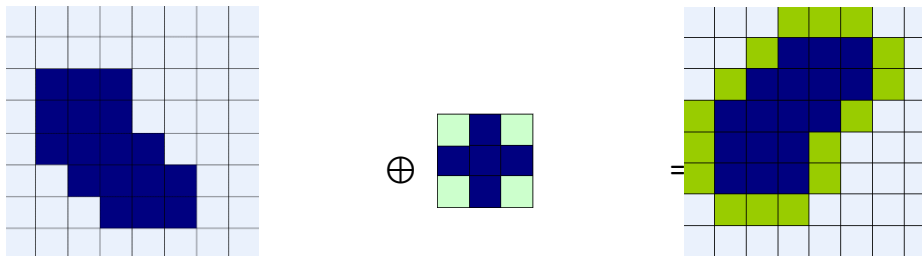
<sup>1</sup> morphology

<sup>2</sup> erosion

<sup>3</sup> dilation

$$A \oplus B = \{z / [(\hat{B})_z \cap A] \subseteq A\}$$

که B جزو ساختاری و A مجموعه ای است که باید گسترده شود. شکل زیر یک نمونه از فرسایش را نشان می دهد.



شکل 2) نمونه ای از عملگر گسترش

باز کردن<sup>۱</sup>:

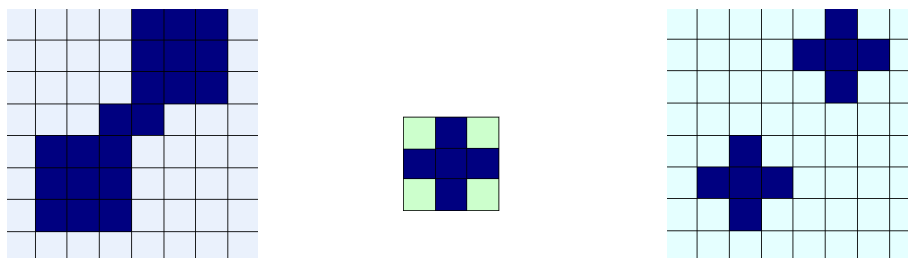
باز کردن محتوای یک شیء را هموارتر کرده، مسیرهای باریک را شکسته و برآمدگی های باریک را حذف می کند. باز کردن مجموعه A با جزء ساختاری B با  $A \circ B$  نمایش داده شده و به صورت زیر تعریف می شود:

$$A \circ B = (A \ominus B) \oplus B$$

که نشان می دهد باز کردن A با جزء ساختاری B معادل است با فرسایش A با B و سپس گسترش نتیجه با B. باز کردن را می توان به صورت زیر نیز توصیف کرد:

$$A \circ B = \bigcup \{ (B)_z \mid \subseteq A \}$$

شکل زیر نمونه ای از بازکردن را نشان می دهد.



شکل 3) نمونه ای از عملگر بازکردن

بستن<sup>۲</sup>:

بستن بخشی از منحنی را هموار می کند اما بر خلاف باز کردن شکستگی های باریک و فاصله های طولانی و باریک را به هم متصل می کند. بستن مجموعه A با جزء ساختاری B با  $A \bullet B$  نمایش داده شده و به صورت زیر تعریف می شود:

$$A \bullet B = A \circ B = (A \oplus B) \ominus B$$

<sup>1</sup> opening

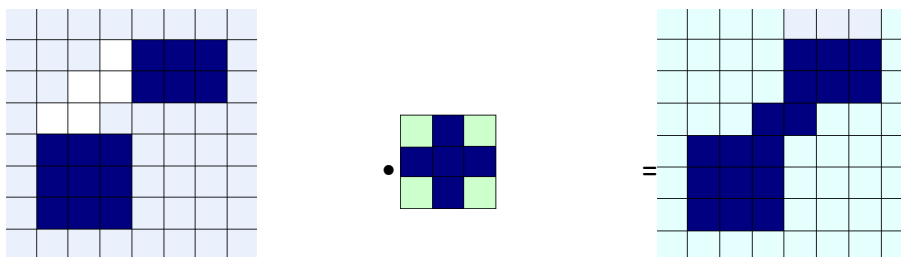
<sup>2</sup> closing



این معادله نشان می دهد باز کردن A با جزء ساختاری B معادل است با گسترش A با B و سپس فرسایش نتیجه با B. باز کردن را می توان به صورت زیر نیز توصیف کرد:

$$A \bullet B = \{(B)_z \mid (B)_z \cap A \neq \emptyset\}$$

شکل زیر نمونه ای از بستن را نشان می دهد.



شکل 4) نمونه ای از عملگر بستن

### فیلتر کالمن<sup>1</sup>:

فیلتر کالمن یک تخمین زننده بهینه می باشد که پارامترهای موردنیاز را از داده ها و اطلاعات غیر مستقیم و غیرقابل اطمینان حدس میزند. این الگوریتم تکرار شونده می باشد و پارامترهای به دست آمده دوباره مورد ارزیابی قرار می گیرند. فیلتر کالمن که بطور گسترده ای در کاربرد های ردگیری بکار می رود فرض میکند که سیستم مورد بررسی یک سیستم خطی است. این فیلتر به طور کلی سه مرحله دارد:

- 1- مرحله پیش بینی
- 2- مرحله مشاهده
- 3- تصحیح مدل پیشبینی براساس مشاهده

به طور کلی فیلتر کالمن ابتدا با استفاده از مدل تخمین موقعیت که در فرمول زیر نشان داده شده است موقعیت جسم را در فریم بعدی پیش بینی می کند:

$$y_k = Ay_{k-1} + Bu_k + w_{k-1}$$

که در این فرمول A مدل حرکت  $Bu_k$  فشارخارجی (در صورت وجود) و  $w_{k-1}$  نشان دهنده ی خطای تخمین زده شده برای مدل با کوواریانس Q می باشد. سپس موقعیت جسم با استفاده از معادله ی مشاهده و اندازه گیری می شود:

$$z_k = Hy_k + v_k$$

در مرحله بعد باتوجه به مشاهده موقعیت فعلی مدل تخمین با استفاده از معادله های زیر تصحیح می شود:

$$\hat{y}_k = \hat{y}_k + K(z_k - H \hat{y}_k)$$

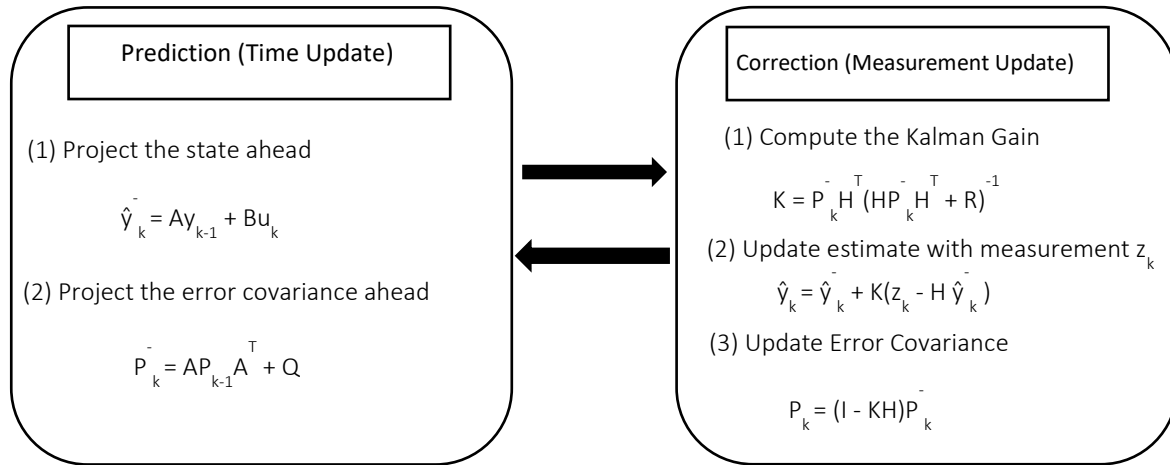
<sup>1</sup> Kalman filter

$$\bar{P}_k = A P_{k-1} A^T + Q$$

$$P_k = (I - KH) \bar{P}_k$$

$$K = \bar{P}_k H^T (H \bar{P}_k H^T + R)^{-1}$$

به طور کلی فیلتر کالمن به صورت زیر نشان داده می شود:



شکل 5) مراحل فیلتر کالمن

#### انتقال هاف<sup>۱</sup>:

یکی از چالش های موجود در زمینه ی تجزیه و تحلیل خودکار تصاویر دیجیتالی، تشخیص اشکال در تصاویر می باشد که در حوزه هایی مانند تشخیص و شناسایی اشکال اهمیت فراوانی دارند. انتقال هاف روشی است که مساله ی تشخیص منحنی ها و اشکال را به مساله ی تشخیص یک قله<sup>۲</sup> در فضای پارامتری تبدیل می کند.

#### انتقال هاف خطی<sup>۳</sup>:

این روش برای تشخیص خطوط در تصاویر به کار می رود. که اغلب روی تصاویر باینری اعمال می شود. فرض می کنیم  $(x, y)$  یک نقطه در صفحه  $xy$  باشد. خط به معادله ی  $y = k \cdot x + m$  از این نقطه می گذرد که مقادیر  $m$  و  $k$  ثابت و  $x$  و  $y$  متغیر می باشند بنابراین باتوجه به مقادیر مختلف، بی نهایت خط از نقطه ی  $(x, y)$  عبور می کند. حال فرض می کنیم مقادیر  $x$  و  $y$  ثابت و مقادیر  $k$  و  $m$  متغیر باشند در این صورت معادله ی  $y = k \cdot x + m$  را داریم که در صفحه ی  $km$  می باشد. این معادله یک خط را در صفحه  $km$  نشان می دهد که نقطه ی تلاقی خطوط در این صفحه مشخص کننده ی یک خط در صفحه ی  $xy$  می باشد. به طور مثال اگر نقطه ی تلاقی چندین خط در صفحه ی  $km$  نقطه ی  $(3, 1)$  باشد نشانگر وجود یک خط با شیب 3- و عرض از مبدا 1

<sup>1</sup> Hough transform

<sup>2</sup> peak

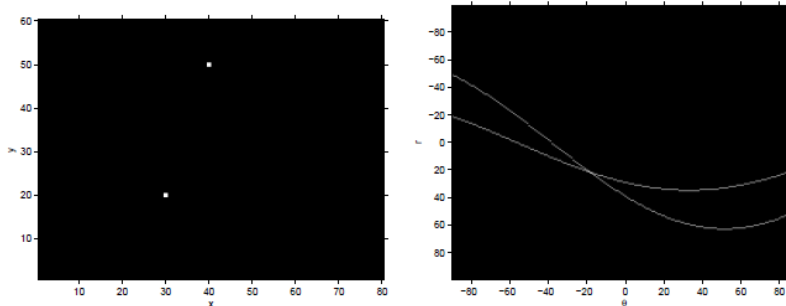
<sup>3</sup> Linear Hough transform

در صفحه  $xy$  است. اما در این روش خطوط افقی به دلیل دارا بودن شیب بی نهایت نمی توانند مورد استفاده قرار گیرند به همین دلیل در این روش از مختصات قطبی استفاده می شود.

$$r = x \cdot \cos\theta + y \cdot \sin\theta \Leftrightarrow$$

$$y = -\frac{\cos\theta}{\sin\theta} \cdot x + \frac{r}{\sin\theta}$$

سپس نقطه ی تلاقی خطوط در مختصات قطبی به عنوان خط در صفحه  $xy$  تلقی می شوند.



شکل 6) مثالی از انتقال هاف خطی

انتقال هاف دایره ای<sup>1</sup>:

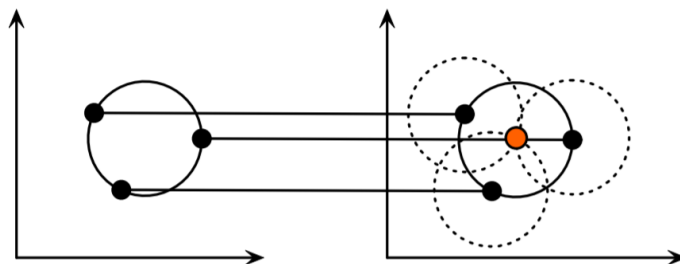
انتقال هاف دایره ای برای تشخیص دایره در تصاویر به کار می رود. یک دایره با شعاع  $R$  و مرکز  $(a, b)$  با معادله ی زیر نشان داده می شود:

$$x = a + R \cos(\theta)$$

$$y = b + R \sin(\theta)$$

زمانی که زاویه  $\theta$  از درجه ی صفر تا 360 را بپیماید نقاط  $x$  و  $y$  محیط یک دایره را رسم می کنند. در انتقال هاف نقاط از صفحه  $xy$  به صفحه ی  $ab$  منتقل می شوند به صورتی که هر نقطه در صفحه ی  $xy$  یک دایره در صفحه  $ab$  می باشد. سپس در صفحه  $ab$  نقطه ای که محل تلاقی محیط چندین دایره می باشد به عنوان مرکز یک دایره در صفحه  $xy$  مشخص می شود. در شکل زیر یک نمونه انتقال هاف دایره ای نشان داده شده است.

<sup>1</sup> circular Hough transform



شکل 7) مثالی از انتقال هاف دایره ای

#### شناسایی لبه ها:

در پردازش تصویر تشخیص لبه ها از اهمیت زیادی برخوردار هستند که تغییرات ناگهانی پیکسل ها در تصویر را نمایان می کند و درواقع مرزهای اشیاء موجود در تصویر را استخراج می کند. برای شناسایی لبه های تصویر روش های متفاوتی هم در حوزه ی زمان و هم در حوزه ی فرکانس موجود می باشد. دو مورد از فیلترهای مکانی که به طور گسترده برای استخراج لبه ها مورد استفاده قرار می گیرند فیلتر های سو بل<sup>1</sup> و کنی<sup>2</sup> می باشند.

#### فیلتر سو بل:

فیلتر سو بل یک فیلتر مکانی سه در سه می باشد که برای استخراج لبه ها از تصویر به کار می رود. شکل زیر فیلتر سو بل را نشان می دهد که کانولوشن شکل الف تصویر را در جهت x و کانولوشن شکل ب تصویر را در جهت y مشتق می گیرد. سپس جمع تصاویر به دست آمده از این دو فیلتر در نهایت لبه های تصویر را نشان می دهد.

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

شکل 8) عملگرهای فیلتر سو بل

یک تصویر به دست آمده از فیلتر سو بل در شکل زیر نشان داده شده است:

<sup>1</sup> sobel



فیلتر سوئل



شکل 9) تصویر خروجی فیلتر سوئل

فیلتر کنی<sup>۱</sup>:

فیلتر کنی نیز یکی دیگر از روش های استخراج لبه ها از تصاویر می باشد. در این روش ابتدا یک فیلتر گوسین برای از بین بردن نویزها و هموارسازی تصاویر به کار می رود. سپس یک فیلتر لبه یابی مانند سوئل یا لاپلاسین برای استخراج لبه ها استفاده می شود. در مرحله ی بعد لبه ها با استفاده از یک مقدار آستانه نازک می شوند.



فیلتر کنی



شکل 10) تصویر خروجی فیلتر کنی

فاصله اقلیدسی<sup>۲</sup>:

اگر  $p=(p_1,p_2,...,p_n)$  و  $q=(q_1,q_2,...,q_n)$  دو بردار  $n$  بعدی باشند، آنگاه فاصله ی اقلیدسی بین آن ها از طریق فرمول زیر محاسبه می شود:

$$D(p,q)=\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

فاصله باتاچاریا<sup>۳</sup>:

اگر توزیع های احتمالی  $p$  و  $q$  در دامنه  $x$  را داشته باشیم آنگاه، فاصله ی باتاچاریا بین این دو تئذیع به صورت زیر تعریف می شود:

<sup>1</sup> canny

<sup>2</sup> Euclidean distance

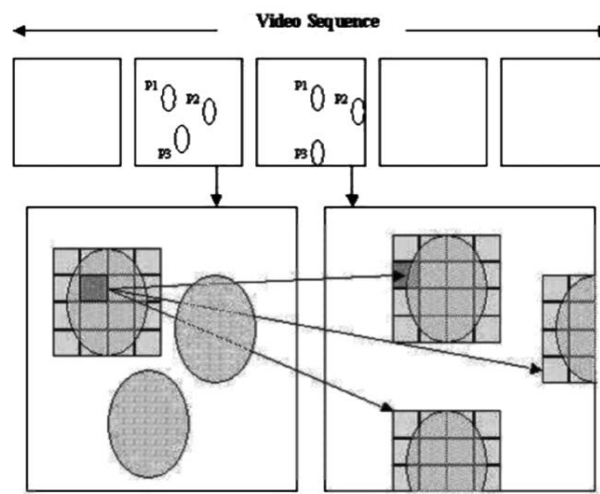
<sup>3</sup> Bhattacharyya distance

$$D_B(p, q) = -\ln(BC(p, q))$$

که در آن  $BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)}$  ضریب باتاچاربا برای توزیع های احتمالی گسسته می باشد.

#### تطبیق بلوکی<sup>۱</sup>:

تطبیق بلوکی یک روش برای تخمین حرکت اشیا در دنباله ای از فریم ها می باشد. تطبیق بلوک به معنی انتخاب یک بلوک به عنوان مرجع در یک فریم و سپس پیدا کردن مشابه ترین بلوک به بلوک مرجع در فریم های بعدی می باشد. فرضیه ی اساسی در این روش این است که الگوی اشیا و پس زمینه در فریم های متوالی حرکت می کنند. در این روش فریم به بلوک هایی تقسیم شده و با بلوک های فریم بعدی مقایسه می شوند. شکل زیر روش یک تطبیق بلوکی را نشان می دهد:



شکل 11 تطبیق بلوکی

یکی از مهم ترین روش های مقایسه ی بلوک ها، مقایسه بر اساس شدت روشنایی می باشد.

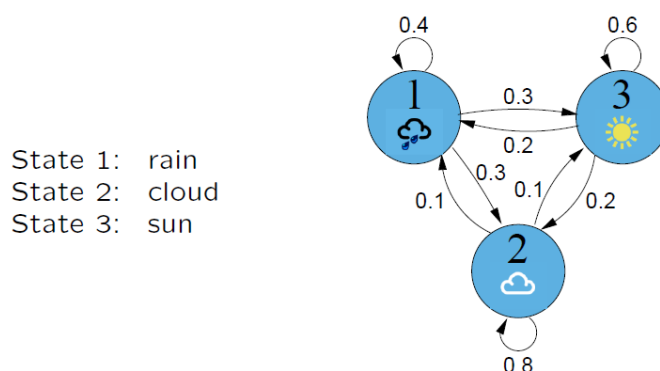
#### مدل مارکوف<sup>۲</sup>:

مدل مارکوف یک مدل تصادفی می باشد که داده های مرتب و متوالی را مدل می کند و در آن انتقال از یک حالت به حالت دیگر صورت می گیرد که تعداد این حالات قابل شمارش است. به طور کلی یک مدل مارکوف به صورت  $\lambda(A, B, \pi)$  نشان داده می شود که A نشان دهنده ی توزیع احتمال انتقال، B نشان دهنده ی توزیع احتمال مشاهدات و  $\pi$  نشان دهنده ی توزیع حالات اولیه می

<sup>1</sup> Block matching

<sup>2</sup> Markov model

باشد. در صورتی که وضعیت ها در مدل مارکوف مخفی باشند به آن مدل مخفی مارکوف<sup>۱</sup> می گویند. یک نمونه مدل مارکوف در شکل زیر نشان داده شده است:



شکل 12) یک نمونه از مدل مارکوف با سه حالت 1-بارانی 2-خورشیدی و 3-بری

فاصله ی انحراف زمانی پویا<sup>۲</sup>:

فاصله انحراف زمانی پویا یا به اختصار  $dtw$ ،  $d_{DTW}(x, y)$  بین دو دنباله  $X = \{x_1, x_2, \dots, x_n\}$  و  $Y = \{y_1, y_2, \dots, y_n\}$  به صورت زیر محاسبه می شود:

$$d_{DTW}(x, y) = D(n, m)$$

$$D(i, j) = d_{Eu}(x_i - y_j) + \min \begin{cases} D(i-1, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{cases}$$

که در این معادله  $x_i \in X$  و  $y_j \in Y$  و  $n$  و  $m$  اندازه ی  $X$  و  $Y$  می باشند. همچنین  $d_{Eu}(x_i - y_j)$  فاصله ی اقلیدسی بین دو دنباله می باشد.

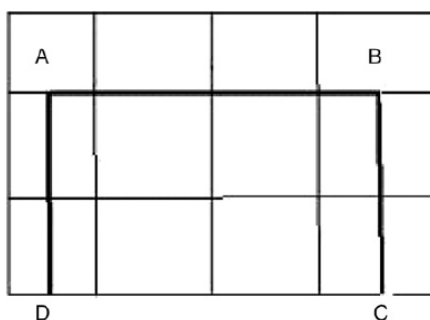
#### تشخیص توپ در ویدیوهای فوتبال منتشر شده

در بازی های فوتبال ، توپ بیشترین توجه را از سوی تماشاگران به خود جلب می کند زیرا مهم ترین اتفاقات و رویدادهای فوتبال را توپ رقم می زند . بنابراین تشخیص توپ در ویدیوهای فوتبال از اهمیت ویژه ای برخوردار است. از این رو تحقیقات زیادی پیرامون تشخیص توپ در ویدیوهای فوتبال انجام شده است. عمده مطالعات انجام شده در زمینه تشخیص و ردیابی توپ بر اساس فیلتر کالمن و انتقال هاف می باشد.

<sup>1</sup> Hidden markov model

<sup>2</sup>Dynamic Time Warping

در [14] ابتدا نمای فریم ها به سه دسته تقسیم می شود: نمای دور<sup>۱</sup> نمای معمولی یا متوسط<sup>۲</sup> و نمای نزدیک<sup>۳</sup>. باتوجه به این که در نماهای دور بیشتر مساحت فریم را چمن زمین بازی در برمی گیرد، می توان با استفاده از خصوصیت رنگ به راحتی نوع نما را مشخص کرد. برای این منظور ابتدا فریم مورد نظر به دوازده فریم تقسیم شده و قسمت میانی فریم که در شکل زیر نشان داده شده است، مورد بررسی قرار می گیرد.



نیکل 13) منطقه ی مورد بررسی در میدان فریم

این بخش از فریم به این دلیل انتخاب شده است که در اغلب موارد قسمت میانی فریم دربرگیرنده ی زمین بازی می باشد در صورتی که ممکن است قسمت های دیگر فریم حاوی تماشاگرها و یا تابلوهای تبلیغاتی باشند. سپس از فضای رنگ YIQ استفاده شده و مقادیر مینیمم و ماکزیمم مؤلفه های رنگ، به صورت تجربی و با بررسی رنگ چمن ویدیوهای مربوط به بازی های مختلف به دست می آید.

در مرحله ی بعد تمامی پیکسل های درون منطقه مورد نظر بررسی شده، در صورتی که حداقل 75 درصد از این پیکسل ها در محدوده رنگ مشخص شده برای چمن واقع شده باشند، فریم، مربوط به نمای دور، اگر این مقادیر بین 0.5 تا 0.75 باشد فریم مربوط به نمای معمولی و در غیر این صورت مربوط به نمای نزدیک می باشد. در نهایت این الگوریتم 96.68 درصد از فریم های نمای دور، 83.76 درصد از نماهای معمولی 87.63 درصد از نمای نزدیک را به درستی دسته بندی می کند و در تشخیص نماهای دور دقت بالاتری را نشان می دهد.

سپس موقعیت توپ در هر فریم مشخص می شود. برای این منظور، در ابتدا جهت کاهش نویز، فریم هموارسازی شده و از فیلتر سوپل برای یافتن لبه ها استفاده می شود. سپس با استفاده از انتقال هاف و فیلتر کلاه مکزیکی اشکال دایره ای در تصویر شناسایی شده و با بررسی شعاع توپ در فریم های مختلف، اندازه ی تقریبی شعاع، برای فریم های نمای دور و نمای معمولی به دست می آید. این اندازه برای نماهای دور بین 1 تا 4 پیکسل و برای نمای معمولی بین 10 تا 15 پیکسل می باشد.

<sup>1</sup> Long shot

<sup>2</sup> Medium shot

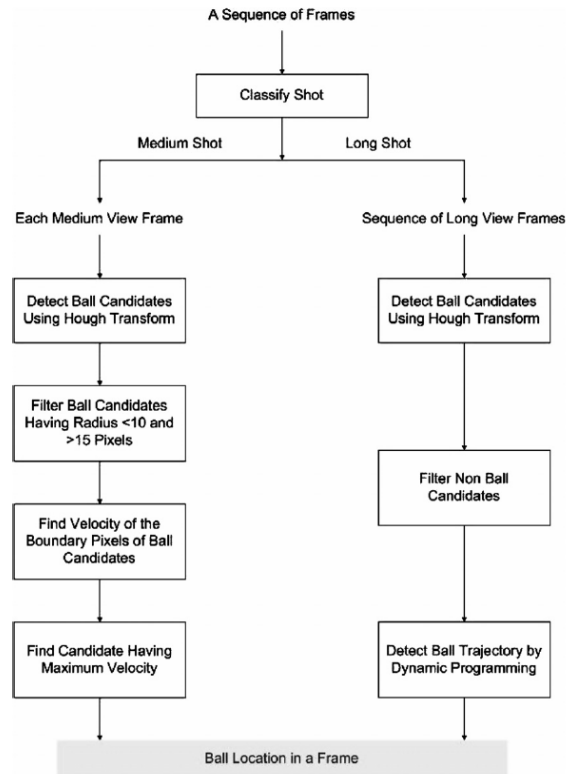
<sup>3</sup> Close shot



در مرحله ی بعدی با استفاده از روش هورن و شانک<sup>1</sup> سرعت اشیای مختلف محاسبه شده و باتوجه به این که معمولاً سرعت توپ بیشتر از سرعت بقیه اشیاء درون فریم است؛ موقعیت توپ در زمین مشخص می شود. اما این روش در نماهای دور دارای دقت کمی می باشد. به همین دلیل در نماهای دور از روش دیگری استفاده می شود. در نماهای دور ابتدا لوگوی تلویزیونی، که معمولاً در گوشه بالا سمت چپ فریم قرار دارد، حذف می شود. در مرحله بعد خط افقی که میدان بازی را از تابلوها ی تبلیغاتی و تماشاگرها جدا می کند شناسایی شده و تمام مناطق بالای این خط نادیده گرفته می شوند. همچنین برخی مناطق واقع روی خط عمودی وسط میدان نیز ممکن است به اشتباه به عنوان کاندیدایی برای توپ در نظر گرفته شود؛ بنابراین اگر تعداد کاندیداهایی که مختصات افقی آنها تقریباً مشابه است، بیشتر از چهار باشد تمام آن کاندیداها حذف می شوند. سپس باتوجه به این که دوربین در بیشتر مواقع توپ را دنبال کرده و در جهت توپ حرکت می کند، اشیایی که مسیر حرکتشان بر خلاف مسیر دوربین می باشند، شناسایی شده و حذف می شوند. برای محاسبه جهت توپ و جهت دوربین از الگوریتم هورن و شانک استفاده می شود. در قدم بعدی برای تشخیص موقعیت دقیق توپ در فریم، مسیرهای به دست آمده از کاندیداها بررسی می شوند. برای به دست آوردن مسیرها، از میان کاندیداها ی شناسایی شده، فقط آنهایی در فریم بعدی مورد بررسی قرار می گیرند که در محدوده ی یک پنجره به اندازه  $w/5$  و  $h/5$  واقع شده باشند.  $h$  طول فریم و  $w$  عرض فریم می باشد. با بررسی کاندیداها گرافی تشکیل شده که هر راس آن نشان دهنده ی یک کاندیدا و هر یال نشان دهنده ی نرخ وابستگی بین کاندیداها می باشد. نرخ وابستگی بین یک کاندیدا در فریم فعلی و کاندیداها ی موجود در دو فریم بعدی با استفاده از یک تطبیق بلوکی انجام می گیرد. برای انتخاب بهترین مسیر، مسیری که بیشترین طول را دارد انتخاب می شود. الگوریتم استفاده شده در شکل زیر نشان داده شده است:

---

<sup>1</sup> Horn and Schunck



شکل 14) الگوریتم تشخیص توپ در دو نمای متفاوت

این مقاله در تشخیص زمین بازی در شرایط مختلف به خوبی عمل می کند؛ اما هنگامی که توپ روی خط عمودی میانی زمین قرار می گیرد، قابل تشخیص نمی باشد. با استفاده از الگوریتم بهترین مسیر می توان تا میزانی این مشکل را برطرف نمود. این الگوریتم در نماهای دور نتایج بهتری را نسبت به نماهای متوسط به دست می آورد و دقت بالاتری را نشان می دهد.

در [15] ابتدا شیءهایی که با توجه به شکل، اندازه و رنگشان نمی توانند توپ باشند حذف شده و سپس بقیه اجسام به عنوان کاندیدای توپ معرفی می شوند. در این روش ابتدا رنگ چمن با توجه به رنگ غالب در فریم به دست می آید. سپس خطوط حذف شده و بازیکنان، داور، توپ و مقداری نویز باقی می ماند. بازیکنان و داور با توجه به اندازه و ابعادشان قابل تشخیص هستند. نویزها را نیز می توان با استفاده از اندازه و رنگ شناسایی و حذف کرد. اجسام باقی مانده را به عنوان کاندیداهایی برای توپ در نظر گرفته و سپس آنها را به سه دسته تقسیم می کنیم به صورتی که احتمال توپ بودن کاندیداهایی که در گروه اول جای می گیرند بیشتر از گروه دوم و به همین صورت احتمال توپ بودن کاندیداهایی که در گروه دوم جای می گیرند بیشتر از گروه سوم می باشد. در مرحله ی بعد با استفاده از فیلتر کالمن یک برآورد از موقعیت توپ در فریم به دست می آید.

در مرحله ی بعد با توجه به اینکه شناسایی دقیق موقعیت توپ در هر فریم مشکل می باشد، به جای بررسی موقعیت توپ، یک تصویر مشخصه کاندیدا تولید می شود. سپس مسیرهای حاصل از کاندیداهای توپ در فریم های متوالی با استفاده از فیلتر کالمن بررسی شده و مسیرهای غیرمعقول حذف می شوند. یک تصویر مشخصه کاندیدا، تصویری است که یک ترکیبی از ویژگی های

کاندیدای در چند فریم متوالی را رسم می کند. تصویر مشخصه کاندیدای XY، یک تصویر سه بعدی می باشد که طول آن تعداد فریم، عرض و ارتفاع آن همان عرض و ارتفاع فریم می باشد. برای حذف مسیرهای غیر معقول دو نکته را در نظر گرفته میشود:

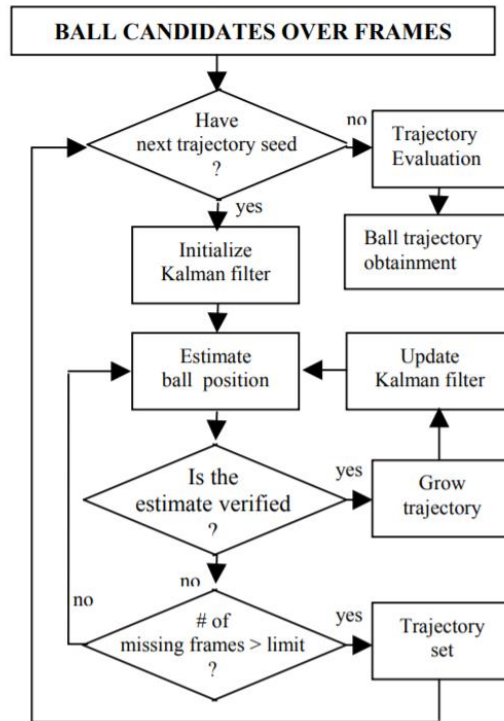
1- معمولا توپ بیشتر از 13 فریم متوالی به بقیه اجسام متصل نمی ماند.

2- معمولا مسیر توپ کمتر از 4 فریم نمی باشد.

بر اساس این دو مشاهده، کاندیداهایی که دارای این شرایط نمی باشند نادیده گرفته میشوند. به این معنی که اگر مسیر موردنظر تنها شامل کاندیداهایی از سومین دسته و یا دارای مسیرهایی به طول کمتر از 4 فریم باشند، حذف می شوند. در نهایت مسیر کاندیداها را با استفاده از فیلتر کالمن بررسی می شوند. به این صورت که ابتدا در تصویر مشخصه کاندیدا موقعیت توپ را در نظر گرفته و سپس با استفاده از فیلتر کالمن برآوردی از موقعیت توپ در فریم بعدی حاصل می شود و در صورتی که کاندیدایی نزدیک به موقعیت تخمین زده شده باشد، آن کاندیدا به مسیر افزوده شده و فیلتر کالمن به روز می شود. یک فریم، گم شده می باشد اگر در آن فریم کاندیدایی نزدیک به موقعیت تخمین زده شده موجود نباشد. در صورتی که تعداد فریم های گم شده ی متوالی از تعداد معینی بیشتر شود، بررسی آن مسیر متوقف شده و مسیر به دست آمده به مجموعه مسیرها اضافه می شود. سپس اعتبار هر مسیر با توجه به طول آن و همچنین تعداد کاندیداهای متعلق به گروه یک تا سه موجود در آن مسیر، از طریق فرمول زیر سنجیده می شود:

$$I(T) = a * L + b * N1 + c * N2 + d * N3.$$

که در این فرمول I نشان دهنده ی اعتبار مسیر، T یک مسیر، L طول مسیر و Ni تعداد کاندیداهای موجود در هر کدام از دسته های 1 تا 3 است. a، b و d ضرایب ثابت می باشند. در نهایت مسیرهایی که دارای هم پوشانی می باشند حذف می شوند. حذف شدن مسیرهای هم پوشان باعث به وجود آمدن خلایه های در مسیر می شود. در صورتی که این خلا کمتر از یازده فریم باشد، آن هارا میتوان به عنوان یک مسیر واحد در نظر گرفت. این الگوریتم در شکل زیر نشان داده شده است:



شکل 15) الگوریتم تشخیص توپ در فریم های متوالی

این روش با وجود دارا بودن دقت بالا، در فریم های مربوط به نمای نزدیک که در آن ها اندازه و ابعاد توپ بیشتر است، ناکام می ماند. به علاوه در این مقاله از فرمول پایه ای فیلتر کالمن استفاده شده است که در صورت تغییر ناگهانی سرعت توپ، قادر به ردیابی آن نمی باشد.

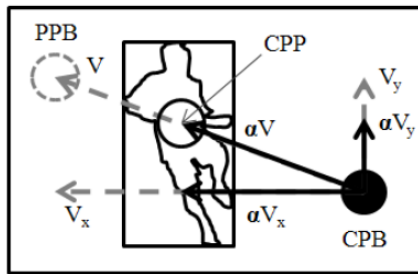
در [16] ابتدا برای تشخیص از زمین از شاخص رنگ و همچنین لبه ها استفاده می شود معادله ی زیر چگونگی تشخیص پیکسل های مربوط به زمین بازی را نشان می دهد:

$$Ground(x,y)= \begin{cases} 1 & (g(x,y) > r(x,y) > b(x,y)) \text{ and } (edge(x,y)=false) \\ 0 & else \end{cases}$$

در این مقاله از فیلتر کنی برای تشخیص لبه ها در فریم استفاده می شود.

سپس جایگاه تماشاگران و اسکوربرد از فریم مورد نظر حذف شده و در مرحله بعدی بازیکنان با توجه به چگالی و سائز پیکسل ها قابل تشخیص و جداسازی می باشند. در این مقاله فرض شده طول یک بازیکن در نمای کلی بین  $1/16$  تا  $1/4$  یک فریم و قطر توپ تقریباً نه اینچ می باشد. برای یافتن موقعیت توپ در هر فریم از فیلتر کالمن دینامیکی استفاده شده که در صورت تغییر سرعت و مسیر هم قادر به شناسایی توپ می باشد. پارامترهای مورد نیاز باتوجه به شرایط زیر تغییر می کنند:

- 1- اگر توپ و کاندیدایی در فضای جست و جوی مورد نظر یافت شود، مقادیر  $r$  و  $q$  در فیلتر کالمن را صفر قرار می دهیم.
- 2- زمانی که در فضای جست و جو (SA) توپی پیدا نشود اما در آن منطقه یک بازیکن شناسایی شود، فرض می کنیم توپ در آن منطقه موجود می باشد و تا زمان پیدا شدن توپ، به جای دنبال کردن توپ، بازیکن ردیابی و دنبال می شود. در این حالت مقدار  $r$  برابر با صفر و مقدار  $q$  در فیلتر کالمن برابر با یک قرار می گیرد.
- 3- اگر در ناحیه ی جست و جو هیچ توپ یا بازیکنی وجود نداشت، فرض می کنیم به دلیل نویز موجود موقتا قادر به شناسایی توپ نمی باشیم. بنابراین مقدار  $r$  را برابر با بی نهایت و مقدار  $q$  را برابر با صفر قرار می دهیم. در این حالت فضای جست و جو بسیار بزرگ بوده و کل فریم برای پیدا شدن توپ جست و جو می شود.
- 4- در صورتی که بازیکنی توپی را کنترل کند، سرعت توپ کم شده و فیلتر کالمن به دلیل در نظر نگرفتن این کاهش سرعت دچار اشتباه می شود. در این حالت باید سرعت در فیلتر کالمن با استفاده از معادله ی زیر کاهش پیدا کند.



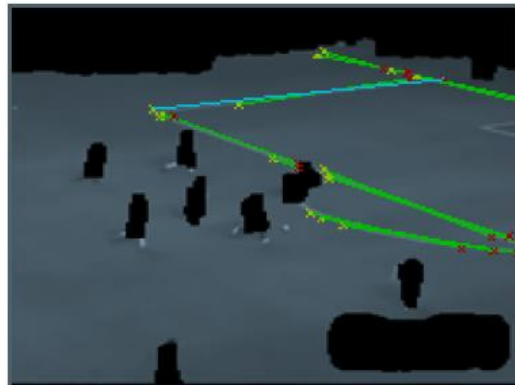
$$\left\{ \begin{array}{ll} \text{If PPB is out of player area} & (V_x, V_y) = \alpha(V_x, V_y) \\ \text{Otherwise} & \text{no change} \end{array} \right.$$

$$\alpha = \frac{CPP - CPB}{PPB - CPB}$$

شکل 16) کنترل سرعت در فیلتر کالمن

استفاده از روش کالمن دینامیکی دقت بالایی در شرایط مختلف از جمله تغییر سرعت توپ دارد. اما زمانی که در محوطه ای از زمین با جمعیت زیاد بازیکنان روبرو می شویم فیلتر کالمن دینامیکی با مشکل روبرو می شود.

در [17] با استفاده از روش توضیح داده شده در بخش قبلی پیکسل های مربوط به زمین بازی استخراج می شوند سپس با استفاده از روش کنی لبه ها شناسایی شده و در مرحله ی بعد با استفاده از عملگرهای مورفولوژی مانند بستن، خطوط به هم متصل می شوند. در نهایت خطوط با استفاده از انتقال هاف خطی شناسایی می شوند. نتیجه در شکل زیر نشان داده است:



شکل 17) خطوط میدان شناسایی شده

برای تشخیص توپ، ابتدا اشیاء از پس زمینه جدا شده و تصویر خاکستری می شود. سپس برای متصل کردن اشیاء از یک عملگر متصل کننده استفاده می کنیم. در مرحله بعد با توجه به نسبت های زیر مناطقی که نمی توانند توپ باشند حذف می شوند:

1- نسبت طول جعبه محدوده به عرض آن

2- نسبت محور بزرگ تر به محور کوچک

این نسبت ها باید تقریباً یک باشند، زیرا طول و عرض منطقه ای که دایره را در بر میگیرد تقریباً مساوی می باشند. در این مقاله با تغییر دادن این نسبت ها می توان توپ هایی که به دلیل سرعت زیاد در طول فریم تغییر شکل می دهند را نیز شناسایی کرد. در مرحله بعد مساحت کاندیدها محاسبه شده و کاندیدهایی که مساحت آنها خیلی بزرگ تر یا کوچک تر از حد معینی باشند حذف می شوند. از آنجایی که ممکن است بعضی از کاندیدها در تابلوی تبلیغات یا بازیکنان شناسایی شده در نظر گرفته شده باشند، کاندیدایی به عنوان توپ در نظر گرفته می شود که رنگ اطراف آن به رنگ زمین چمن باشد. سپس برای ردیابی توپ از فیلتر کالمن استفاده می شود. در این مقاله تعداد بسیاری از کاندیدها حذف شده و فقط یک کاندیدا به عنوان توپ تشخیص داده می شود که همین موضوع باعث کمتر شدن بار محاسباتی می شود.

در [18] ابتدا مانند بخش قبلی پیکسل های مربوط به میدان بازی با استفاده از مشخصه رنگ و همچنین لبه یابی مشخص و حذف می شوند. سپس تصویر باینری می شود. به این صورت که پیکسل های میدان سیاه و اشیاء موجود در زمین به رنگ سفید نشان داده می شوند. در مرحله ی بعد خطوط با استفاده از انتقال هاف پیدا شده و از تصویر حذف می شوند. سپس مساحت اجسام باقیمانده را محاسبه کرده و اشکالی که مساحت کمی دارند از تصویر حذف می شوند برای متصل کردن خطوط جدا شده از روش گسترش و فرسایش استفاده می شود، در نهایت دوباره از فیلتر سوبل برای تشخیص بازیکنان و توپ در فریم استفاده می شود. این مراحل در شکل زیر نشان داده شده اند.



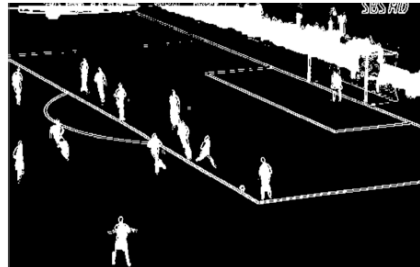
a



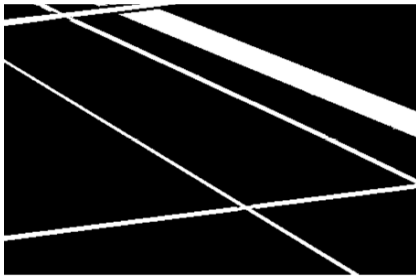
b



c



d



e



f



g



h

شکل 18) مراحل تشخیص توپ و بازیکنان در زمین بازی

#### تشخیص و ردیابی بازیکنان در ویدیوهای فوتبال منتشر شده

بعد از تشخیص توپ، تشخیص و ردیابی بازیکن مهم ترین بخش از تجزیه و تحلیل ویدیوهای فوتبال می باشد. مطالعات بسیاری در این زمینه صورت گرفته و با چالش های زیادی روبرو می باشد. در ادامه برخی روش های پیشنهاد شده برای تشخیص و ردیابی بازیکنان در ویدیوهای فوتبال منتشر شده مورد بررسی قرار می گیرند.

در [19] به منظور شناسایی بازیکنان ابتدا زمین بازی با استفاده از فضای رنگ YIQ شناسایی می شوند. برای این کار پیکسل هایی که مقادیر  $I$  آنها بین 25 تا 55 و مقادیر  $Q$  آنها بین 0 تا 12 باشند به عنوان پیکسل های زمین شناسایی شده و از فریم حذف می شوند. در مرحله ی بعد لوگوی مربوط به کانال در صورت وجود و باتوجه به موقعیت آن در فریم که به صورت دستی مشخص می شود، حذف می شود. سپس با استفاده از انتقال هاف خطوط پیدا شده و خط افقی که در بالاترین مکان فریم قرار دارد به عنوان خط جداساز زمین از تابلوهای تبلیغاتی و جایگاه تماشاگران در نظر گرفته می شود و مناطقی که در بالای این خط قرار گرفتند نادیده گرفته می شوند. در مرحله بعد خط میان زمین نیز حذف شده، سپس نواحی که مساحتی کمتر از 25 دارند نادیده گرفته می شوند و در نتیجه بازیکنان در فریم شناسایی می شوند. نقطه میانی هرکدام از بازیکنان شناسایی شده و یک مربع 32 در

32 اطراف این نقطه به عنوان بازیکن در نظر گرفته می شود. این اعداد به صورت تجربی و با استفاده از بررسی ویدیوها و فریم های مختلف به دست آمده است.

در مرحله ی بعد، برای تشخیص دقیق تر محل بازیکنان از مسیریابی و ردیابی آن ها استفاده می کنیم. در این مرحله یک گراف ساخته می شود که راس های آن مناطق بازیکنان در فریم های مختلف و یال های آن میزان همبستگی بین این مناطق را مشخص می کند. منطقه یک بازیکن در یک فریم همبستگی زیادی نسبت به منطقه همان بازیکن در فریم بعدی دارد. بنابراین برای پیدا کردن محتمل ترین مسیر باید بلندترین مسیر در گراف را پیدا کنیم.

در [20] ابتدا فریم هایی که تماشاگران یا نویز می باشند حذف می شوند؛ به این صورت که اگر تعداد پیکسل هایی که در محدوده رنگ سبز می باشند زیاد نباشد آن فریم نادیده گرفته می شود. سپس برای جداسازی زمین بازی از جایگاه تماشاگران، مقدار برجستگی برای هر پیکسل با استفاده از فرمول زیر محاسبه می شود:

$$D = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2}$$

که در آن مقادیر  $L$ ،  $a$  و  $b$  مقادیر مربوط به فضای رنگ  $L^*a^*b^*$  می باشند و مقادیر  $\Delta L$ ،  $\Delta a$ ،  $\Delta b$  تفاوت مقادیر  $a$ ،  $b$  و  $L$  بین دورنگ را مشخص می کنند. اگر رنگ های موجود در عکس را با  $c_1, c_2, \dots, c_n$  نشان بدهیم، مقدار برجستگی رنگ  $c_k$  را با  $S(c_k)$  نشان داده و آن را با استفاده از فرمول زیر محاسبه می کنیم:

$$S(c_k) = D(c_k, c_{k1}) + D(c_k, c_{k2}) + \dots + D(c_k, c_n)$$

سپس مقدار برجستگی هر پیکسل با استفاده از فرمول زیر به دست می آید:

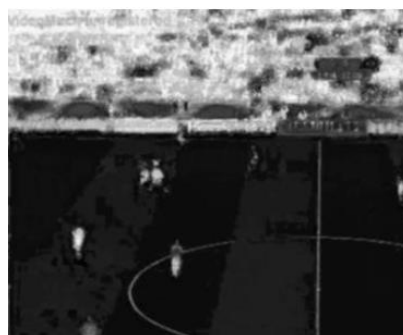
$$S(p1) = S(c_k) = \sum_{i=1}^n f_i D(c_k, c_n)$$

در این فرمول  $f_i = n_i/N$  که  $n_i$  تعداد پیکسلهایی که به رنگ  $i$  هستند و  $n$  تعداد کل پیکسل های تصویر می باشد.

سپس با استفاده از بخش بندی مناطق مربوط به تماشاگران حذف می شود. نتیجه به صورت شکل زیر می باشد.



الف) تصویر اصلی



ب) نگاشت برجستگی



ج) نتیجه بخش بندی

شکل 19) نتایج بخش بندی بر اساس مقادیر برجستگی



در مرحله بعدی با استفاده از الگوریتم اتسو<sup>1</sup> تصویر به گونه ای باینری میشود که بیشترین تفاوت بین پس زمینه و اشیای موجود در زمین ایجاد شود. سپس با استفاده از انتقال هاف و عملگر باز کردن خطوط حذف شده و بازیکنان شناسایی می شوند. در صورتی که سطوح خاکستری بازیکنان با زمین بازی تفاوت چندانی نداشته باشد، الگوریتم اتسو با مشکل روبرو می شود. به همین علت برای دستیابی به دقت بیشتر، از لبه یابی استفاده شده و سپس با نتیجه به دست آمده از روش اتسو ترکیب می شود. در مرحله بعدی برای ردیابی بازیکنان از فیلتر ذره ای استفاده شده و برای بررسی شباهت بین کاندیداهای بازیکن در فریم های متوالی، از خصوصیت  $h$  در فضای  $hsv$  و همچنین لبه ها استفاده می شود. با استفاده از فیلتر ذره ای زیر مسیر بازیکنان در فریم بعدی پیش بینی می شود:

$$x_{t+1} = 2x_t - x_{t-1} + N_t$$

$$Z_t = h(x_t) + v_t$$

که در این معادله  $Z_t$  مقدار مشاهده شده در زمان  $t$ ،  $h(x)$  تابع مشاهده،  $x_t$  تخمین حالت شیء در زمان  $t$  و  $N_t$  نشان دهنده ی نویز می باشد. سپس برای دقت بیشتر محل بازیکنان در چند فریم بعدی هم شناسایی شده و با استفاده از مقایسه هیستوگرام مربوط به مقدار  $hue$  و همچنین مقایسه ی لبه های موجود در مستطیل محیطی اطراف بازیکن، شبیه ترین شیء به بازیکن فعلی مشخص شده و به عنوان موقعیت بعدی بازیکن مشخص می شود. همچنین اگر یک شیء جدید شناسایی شده در فریم هیچ هم پوشانی با پیش بینی های انجام شده نداشته باشد به عنوان یک بازیکن جدید مشخص می شود. این الگوریتم دارای استحکام و دقت زیادی است اما همچنین در تشخیص بازیکنانی که نزدیک به تابلوهای تبلیغاتی و یا جایگاه تماشاگران هستند ناکام می ماند.



پ) موقعیت بازیکنان برچسب زده شده



ب) نتیجه شناسایی بازیکن فوتبال



الف) یک فریم از ویدیوی فوتبال منتشر شده

شکل 20) شناسایی بازیکنان فوتبال

در [21] از فیلتر کالمن برای ردیابی بازیکن استفاده شده است. یک بازیکن به صورت مستقیم و بازیکنان دیگر به صورت غیر مستقیم ردیابی می شوند. به جای اینکه بردار حالت، نشان دهنده ی موقعیت آنها در زمین باشد، نشان دهنده ی موقعیت نسبی آنها نسبت به بازیکن اصلی می باشد. بردار حالت به صورت زیر معرفی می شود:

$$X_k = [w \ h \ x \ y \ v_x \ v_y \ x_1 \ y_1 \ \dots \ x_n \ y_n].$$

<sup>1</sup> Otsu

که در این بردار  $w$  و  $h$  نشان دهنده ی عرض و ارتفاع مستطیل محیطی اطراف بازیکن،  $x$  و  $y$  نشان دهنده ی موقعیت مکانی،  $v_x$  و  $v_y$  نشان دهنده ی سرعت در جهت  $x$  و سرعت در جهت  $y$  و بقیه مربوط به موقعیت نسبی سایر بازیکنان نسبت به شیء اصلی و بازیکن اصلیدمی باشند. در مرحله ی بعد یک مدل حرکتی برای بازیکن مشخص می شود. مدل حرکتی ماتریسی است که نحوه ی حرکت بازیکن را نشان می دهد. نمونه ای از یک ماتریس حرکتی به شکل زیر می باشد:

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

باتوجه به ماتریس حرکتی بالا می توان نوشت:

$$x_k = x_{k-1} + v_x$$

$$y_k = y_{k-1} + v_y$$

$$v_{xk} = v_{xk-1}$$

$$v_{yk} = v_{yk-1}$$

که در این معادلات  $x_k$  و  $y_k$  نشان دهنده ی موقعیت شیء در فریم  $K$  و  $v_{xk}$  و  $v_{yk}$  نشان دهنده ی سرعت در جهت  $x$  و سرعت در جهت بردار  $y$  می باشد.

پس از پیش بینی موقعیت بعدی یک بازیکن، فریم بعدی مشاهده و مدل پیش بینی با توجه به آن تصحیح می شود. به این صورت که شیء کاندیدا در فریم های بعدی با بازیکن اصلی در فریم مقایسه شده و بازیکنی به عنوان بهترین کاندیدا انتخاب می شوند که کمترین تفاوت را با بازیکن اصلی داشته باشد. برای مقایسه ی اشیاء درون فریم ها ابتدا جعبه محیطی اطراف هر بازیکن را به سه قیمت مساوی تقسیم کرده سپس تفاوت هیستوگرام هر قسمت را با استفاده از فاصله باتاچاریا به صورت جداگانه محاسبه می کند و برای به دست آوردن تفاوت کلی از میانگین وزن دار استفاده می شود. وزن قسمت و سر یا تنه بازیکن بیشتر است و از اهمیت بیشتری برخوردار است. سپس باید مرجع هر کدام از این بازیکنان مشخص شود که برای این منظور از الگوریتمی که در شکل زیر نشان داده شده، استفاده می شود. در این الگوریتم برای تشخیص مرجع، از تمام کاندیداهای فریم های قبلی استفاده شده با این تفاوت که داده هایی که به فریم فعلی نزدیک تر هستند وزن بیشتری دارند.

---

$D$ =Histogram distance (reference patch , best candidate patch)

```
If  $D < \text{threshold1}$ 
    reference patch= best candidate patch
Else if ( $D > \text{threshold1} \&\& D < \text{threshold2}$ )
    Increase alpha until  $\alpha < 0.5$ 
    & use moving average formula for updating
    Reference patch
Else if ( $D > \text{threshold3} \&\& D < \text{threshold3}$ )
    Decrease alpha by rate 0.05 until  $\alpha > 0.2$ 
    & use moving average formula for updating
    Reference patch
Else
    Decrease alpha by rate 0.1 until  $\alpha > 0.2$ 
    & Don't update reference patch
End
```

---

شکل 21) الگوریتم به روز رسانی مناطق مرجع

میانگین متحرک<sup>1</sup> معادله ای می باشد که میانگین جدید را برای داده های وارد شده جدید محاسبه می کند؛ به صورتی که تاثیر داده های قدیمی را کمتر کند. در این الگوریتم هرچه آلفا بزرگ تر باشد تاثیر داده های قبلی با نرخ بیشتری کاهش می یابد. همچنین در این الگوریتم زمانی که سرعت یا جهت بازیکنان تغییر می کند مقدار  $Q$  که نشان دهنده ی نویز مدل تخمین فیلتر کالمن می باشد، افزایش می یابد، مقدار بهترین کاندیدا با مقدار منطقه مرجع مقایسه می شود و در صورتی که این اختلاف کم باشد به این معنی است که کاندیدای مناسبی پیدا شده و در نتیجه مقدار  $R$  کاهش می یابد. اما در صورت زیاد بودن اختلاف مقدار  $R$  که بیانگر نویز و خطای اندازه گیری در فیلتر کالمن است، افزایش می یابد.

در مبحث تشخیص و ردیابی بازیکنان روش های بسیاری به کار گرفته شده از جمله فیلتر های مکانی، الگوریتم های بهینه سازی و غیره. اما عمده این روش ها براساس فیلتر کالمن و فیلتر های مکانی می باشد. از جمله چالش های موجود در زمینه ی تشخیص بازیکن ادغام و هم پوشانی بازیکنان در ویدیو فوتبال می باشد که باعث ناکامی در ردیابی بازیکنان می شود.

#### شناسایی وقایع و رویدادها در ویدیو های فوتبال

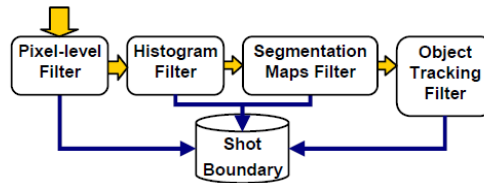
در [22] یک روش برای استخراج رویداد کرنر در ویدیوهای فوتبال ارائه شده. در این مقاله ابتدا فریم هایی که دارای نرخ زیادی از رنگ چمن می باشد، جدا شده و سپس از مدل مارکوف برای تشخیص رویداد کرنر استفاده می شود. در این مقاله فرض شده رویدادهای خاص فوتبال دارای ترتیب خاصی از صحنه ها و فریم ها می باشند. مثلاً در یک رویداد کرنر ابتدا یک فریم با یک بازیکن در نقطه ی کرنر نشان داده می شود سپس بعد از ضربه خوردن توپ فریم یک نمای دور از محوطه ی جریمه نشان می

---

<sup>1</sup> Moving average

دهد در نتیجه از حرکت لنز برای تشخیص رویداد کرنر می توان استفاده کرد. در ابتدا صحنه های مربوط به نقطه ی کرنر با استفاده از الگوریتم های آموزش داده می شوند.

در [23] ابتدا مرزهای شات ها شناسایی می شود. برای این منظور از مقایسه پیکسل ها، مقایسه هیستوگرام و همچنین بخش بندی و دنبال کردن اشیا در زمین استفاده می شود. این مراحل در شکل زیر نشان داده شده است:



شکل 22) معماری چند فیلتره برای تشخیص شات

در این شکل فیلتر سطح پیکسل<sup>1</sup>، میانگین درصد پیکسل های تغییر یافته در فریم های متوالی در یک شات را نشان می دهد. فیلتر هیستوگرام<sup>2</sup> میانگین تغییرات هیستوگرام در فریم های یک شات را مشخص می کند و باتوجه به اینکه در منطقه ی مربوط به پس زمینه تغییرات زیادی وجود ندارد، زمین چمن مشخص شده و نماهای دور و نزدیک استخراج می شوند. سپس مشخصات صوتی مانند حجم، انرژی و طیف با استفاده از نمونه برداری استخراج شده سپس با استفاده از این مشخصات یک درخت آموزش دیده و برای استخراج شات های مربوط به گل مورد استفاده قرار می گیرند. قبل از اینکه شات های ویدیو برای استخراج گل مورد بررسی قرار بگیرند توسط سه قانون زیر فیلتر می شوند:

- 1- شات کاندیدای مورد نظر سه ثانیه ابتدایی و سه ثانیه انتهایی باید دارای حداقل یک نقطه هیجانی باشد.
- 2- شات مربوط به گل باید حداقل دارای نرخ چمن چهل درصد باشد.
- 3- در دو شات بعد از گل حداقل باید یک شات در بردارنده ی نمای نزدیک باشد.

در [24] یک الگوریتم براساس مدل مخفی مارکوف به منظور استخراج رویدادها در ویدیو فوتبال ارایه شده است. مشخصاتی مانند پلی بک، زوم، وجود یا عدم وجود دروازه در فریم برای رویدادهای مختلف در ویدیو های فوتبال استخراج شده و برای آموزش مدل مخفی مارکوف توسط الگوریتم بام ولس مورد استفاده قرار می گیرند. برای هر رویداد یک مدل مارکوف جداگانه مورد استفاده قرار می گیرد سپس برای تشخیص یک رویداد چند فریم متوالی به عنوان ورودی به مدل مخفی مارکوف  $\lambda_i$  داده شده و مدلی که فرمول زیر را ارضا کند به عنوان خروجی مشخص می شود:

$$J = \arg \max P(Q, O | \lambda_i)$$

در [25] یک روش برای استخراج رویدادها براساس مجموعه های فازی پیشنهاد شده است. در این مقاله ابتدا دو گروه داده از ویدیوهای فوتبال استخراج می شوند: گروه اول شامل داده های آموزشی و تست، مربوط به میانه ی زمین، نماهای نزدیک و صحنه

<sup>1</sup> Pixel level filter

<sup>2</sup> Histogram filter

های مربوط به مردم (نمونه ای از هر کدام در شکل زیر نشان داده شده است) ، و گروه دوم شامل داده ها و صحنه های متعلق به رویدادهای مختلف می باشد.



شکل 25 صحنه ی مربوط به مردم



شکل 24 صحنه ی مربوط به نمای نزدیک



شکل 23 صحنه ی مربوط به میانه ی زمین

سپس سه صحنه استاندارد مربوط به نمای میانه ی زمین، نمای نزدیک و نمای شامل مردم انتخاب شده و برای هر کدام از داده های استخراج شده مقادیر زیر را محاسبه می کنیم:

$$dsc(H_a, H_b) = \sum_{i=0}^I \frac{(hai - hbi)^2}{hai} \quad \text{Chi-Square distance}$$

$$dco(H_a, H_b) = \frac{\sum_{i=0}^I (hai - \bar{H}a) - (hbi - \bar{H}b)}{\sqrt{\sum_{i=0}^I (hai - \bar{H}a)^2 \sum_{i=0}^I (hbi - \bar{H}b)^2}} \quad \text{correlation}$$

$$dbd(H_a, H_b) = \sqrt{1 - \frac{1}{\sqrt{\bar{H}a \bar{H}b N^2}} \sum_{i=0}^I hai hbi} \quad \text{Bhattacharyya distance}$$

$$din(H_a, H_b) = \sum_{i=0}^I \min(hai, hbi) \quad \text{intersection}$$

در این معادلات  $H_a$  و  $H_b$  دو هیستوگرام متفاوت از دو فریم مرتبط با یک کانال رنگی می باشند. همچنین  $H_a$  متعلق به یک صحنه ی استاندارد و  $H_b$  مربوط به یک فریم از داده های آموزشی می باشد و مقدار  $i$  سطوح مختلف هیستوگرام را نشان می دهد.  $\bar{H}x$  متوسط سطح رنگ هیستوگرام یک تصویر را نشان می دهد. بنابراین برای هر کدام از فریم ها و داده های آموزشی و آزمایشی دوازده ویژگی استخراج می شود:

$$V_T^i = \left\{ \begin{array}{l} dco_B^i, dco_G^i, dco_R^i, dcs_B^i, dcs_G^i, dcs_R^i, \\ din_B^i, din_G^i, din_R^i, dbd_B^i, dbd_G^i, dbd_R^i \end{array} O^i \right\}$$

در این بردار  $O^i$  نشان دهنده ی کلاس مربوط به صحنه ی فریم  $i$  می باشد .

سپس از الگوریتم خوشه بندی فازی سی مینز<sup>1</sup> برای تولید مجموعه های فازی نوع یک استفاده می شود. هر گروه از داده ها به سه مجموعه ی فازی گوسین کم، متوسط و زیاد تقسیم شده سپس با استفاده از این مجموعه ها قوانین فازی مناسب توسط الگوریتم بهینه سازی تولید می شود. در نهایت این قوانین فازی برای تشخیص صحنه ی مربوط به یک فریم مورد استفاده قرار می گیرند.

<sup>1</sup> Fuzzy C-Means clustering

سپس ترتیبی از فریم ها که نشان گر یک رویداد می باشند را جدا کرده و به عنوان داده های آموزشی استاندارد استفاده می کنیم. این داده های آموزشی ماتریسی به شکل زیر را تشکیل می دهند:

$$V_E = \begin{bmatrix} O_{class1}(V_U^1) & \cdots & O_{class3}(V_U^1) \\ \vdots & \ddots & \vdots \\ O_{class1}(V_U^l) & \cdots & O_{class3}(V_U^l) \end{bmatrix}$$

که در این ماتریس  $O_{class1}(V_U^l)$  نتیجه ی ارزیابی صحنه ی فریم مورد نظر، و  $l$  تعداد فریم های متوالی یک رویداد می باشد.

سپس فاصله **dtw** هر کدام از داده های آموزشی با داده های استاندارد را محاسبه کرده و در نهایت برای هر کلاس از رویدادها بر داری مانند زیر به دست می آید:

$$D_{Train}^h = \{d_{E1}^h, d_{E2}^h, \dots, d_{EN}^h\}$$

که  $h$  نشان دهنده ی کلاس رویداد و  $N$  نشان دهنده ی تعداد داده های آموزشی و  $d_{EN}^h$  نشان دهنده ی فاصله ی **dtw** هر کدام از داده های آموزشی با داده های استاندارد می باشد.

سپس برای ارزیابی یک صحنه یا رویداد، رویدادی به عنوان نتیجه انتخاب می شود که فرمول زیر را ارضا کند:

$$O_{DTW}^h = \min \{|d_o^h - d_{En}^h|\}$$

که در این معادله  $O_{DTW}^h$  نشان دهنده ی شباهت توالی فریم ورودی با توالی فریم های هر کدام از کلاس های پیش بینی شده می باشد.

#### نتیجه گیری:

در این مطالعه روش های مختلفی برای شناسایی و ردیابی توپ و بازیکنان در هر فریم و همچنین استخراج رویدادها و وقایع مختلف فوتبال معرفی شد. عمده روش به کار برده شده برای ردیابی توپ با استفاده از روش های مسیریابی مانند کالمن و همچنین روش های پیدا کردن اشکال دایره ای شکل مانند انتقال هاف می باشد. با این وجود عمده روش های به کار گرفته شده در مواقعی که توپ دارای سرعت بالا یا تغییر ناگهانی سرعت و جهت می باشند، دارای دقت کمتری هستند. ردیابی بازیکنان نیز با استفاده از روش های مختلف مانند فیلترهای مکانی و مسیریابی و الگوریتم های بهینه سازی انجام میگیرد. این الگوریتم ها با وجود دقت بالا در زمان ازدحام بازیکنان دچار اشکال می شوند. استخراج رویدادها و وقایع نیز اغلب با استفاده از استخراج فریم های مربوطه و استفاده از آنها در الگوریتم های یادگیری صورت می گیرد. این الگوریتم ها اغلب با توجه به فیلم های مربوط به لیگ های کشورهای مختلف نتایج مختلفی را نشان می دهند.

منابع:

[1] Football: WRITTEN BY: Eric Weil, Richard C. Giulianotti, Peter Christopher Alegi, Bernard Joy, Jack Rollin, <https://www.britannica.com/sports/football-soccer#toc550852main>

[2] John Wesson-The Science of Soccer-Institute of Physics Publishing (2002)

- [3] Barrie Pierpoint (2000) 'Heads above water': Business strategies for a new football economy, Soccer & Society
- [4] T. D'Orazio, M. Leo, "A review of vision-based systems for soccer video analysis", Pattern Recognition 43 (2010) 2911–2926
- [5] M. Manafifard, H. Ebadi, H. Abrishami Moghaddam, "A survey on player tracking in soccer videos", Computer Vision and Image Understanding 159 (2017) 19–46
- [6] Jean-Francis Grehaigne, Daniel Bouthier & Bernard David (1997): Dynamic-system analysis of opponent relationships in collective actions in soccer, Journal of Sports Sciences, 15:2, 137-149
- [7] M. Manafifard & H. Ebadi, H. Abrishami Moghaddam, "Multi-player detection in soccer broadcast videos using a blob-guided particle swarm optimization method", Springer Science+Business Media New York 2016
- [8] Upendra Rao Moyyila, "Detection and Recognition of Soccer Ball and Players", Department of Electronics & Communication Engineering National Institute of Technology Rourkela Rourkela, Odisha-769008 May 2015
- [9] Priyanka Mukhopadhyay, Bidyut B. Chaudhuri, A Survey of Hough Transform, Pattern Recognition, <http://dx.doi.org/10.1016/j.patcog.2014.08.027>
- [10] Berthold K.P. Horn and Brian G. Rhunck. Determining Optical Flow, *Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*
- [12] Jaroslav Borovička, "Circle Detection Using Hough Transforms Documentation", Image Processing and Computer Vision
- [13] Greg Welch & Gary Bishop, An Introduction to the Kalman Filter
- [14] V. Pallavi & Jayanta Mukherjee & Arun K. Majumdar & Shamik Sural, Ball detection from broadcast soccer videos using static and dynamic features, J. Vis. Commun. Image R. 19 (2008) 426–436
- [15] X. Yu, C. Xu, Q. Tian, H. W. Leong, "A ball tracking framework for broadcast soccer video", *Proc. ICME*, vol. II, pp. 273-276, 2003..

- [16] Jong-Yun Kim, Tae-Yong Kim, "Soccer Ball Tracking using Dynamic Kalman Filter with Velocity Control", 2009 Sixth International Conference on Computer Graphics, Imaging and Visualization
- [17] Jila Hosseinkhani, Soltanian-Zadeh Hamid, Mahmoud Kamarei, Oliver Staadt, "Ball Detection with the Aim of Corner Event Detection in Soccer Video", *International Symposium on Parallel and Distributed Processing with Applications Workshops*, pp. 147-152, May, 2011
- [18] M. M. Naushad Ali, M. Abdullah-AI-Wadud, Seok-Lyong Lee, "An Efficient Algorithm for Detection of Soccer Ball and Players", *Proceedings of Conference on Signal and Image Processing (SIP)*, pp. 1-8, 2012.
- [19] V. Pallavi, J. Mukherjee, A. K. Majumdar, S. Sural, "Graph-based multiplayer detection and tracking in broadcast soccer videos", *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 794-805, Aug. 2008.
- [20] Y. Yang, D. Li, Robust player detection and tracking in broadcast soccer video based on enhanced particle filter, *J. Vis. Commun. Image R.* (2017), doi: <http://dx.doi.org/10.1016/j.jvcir.2017.03.008>
- [21] Nima Najafzadeh, Mehran Fotouhi, Shohreh Kasaei, *Multiple Soccer Players Tracking*", Najafzadeh N, Fotouhi M, Kasaei S (2015) Multiple soccer players tracking", 2015 International Symposium on Artificial Intelligence and Signal Processing (AISP).
- [22] Haitao Yang, Jia Wang and Jingmeng Sun, "Detection of corner event in soccer video", *Advanced Science and Technology Letters Vol.111 (NGCIT 2015)*, pp.147-150 <http://dx.doi.org/10.14257/astl.2015.111.28>
- [23] S.-C. Chen, M.-L. Shyu, M. Chen, C. Zhang, "A decision tree-based multi-modal data mining framework for soccer goal detection", *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, pp. 265-268, 2004.
- [24] G. Jin, L. Tao, G. Xu, "Hidden Markov Model Based Events Detection in Soccer Video", *Proc. Int'l Conf. Image Analysis and Recognition*, pp. 605-612, 2004.
- [25] Wei Song and Hani Hagra, "A Type-2 Fuzzy Logic System for Event Detection in Soccer Videos", 978-1-5090-6034-4/17/\$31.00
- [26] B.K.P. Horn, B. Schunk, "Determining Optical Flow", *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.



- [27] G. Xuan, P. Chai and M. Wu, "Bhattacharyya distance feature selection," Proceedings of the 13th International Conference on Pattern Recognition, volume 2, pp. 195-199. IEEE, 25-29 Aug. 1996
- [28] D. H. Ballard, Generalizing the Hough Transform to Detect Arbitrary Shapes ,Pattern Recognition, 13(2):111-122, 1981
- [29] John Canny. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679–698, Nov. 1986.
- [30] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. Commun. ACM, 15(1):11–15, January 1972.
- [31] R. G. Brown and P. Y. C. Hwang. 1992. *Introduction to Random Signals and Applied Kalman Filtering, 2nd Edition*, John Wiley & Sons, Inc.
- [32] Greg Welch and Gary Bishop. 1995. "An Introduction to the Kalman Filter," University of North Carolina, Department of Computer Science, TR 95-041.