

Tytuł: **System detekcji i śledzenia piłkarzy w transmisjach widowisk sportowych w cyfrowym sygnale wizyjnym**

Tytuł ang.: **A complex system for football player detection in broadcasted video**

**Autorzy:**

Sławomir Maćkowiak, tel. 61 6653890 e-mail: smack@multimedia.edu.pl

Maciej Kurc, e-mail: mkurc@multimedia.edu.pl

Jacek Konieczny, e-mail: jkonieczny@multimedia.edu.pl

Przemysław Maćkowiak, e-mail: przemyslaw.mackowiak@gmail.com

**Adres:** Katedra Telekomunikacji Multimedialnej i Mikroelektroniki, Politechnika Poznańska, ul.Polanka 3, 61-131 Poznań, Polska

**Streszczenie**— W artykule przedstawiono nowatorski system segmentacji służący do detekcji i śledzenia piłkarzy w zdekompresowanym cyfrowym sygnale wizyjnym. System oparty jest na połączeniu technik opisu za pomocą deskryptora HOG - Histogram of Oriented Gradients i liniowej maszyny wektorów nośnych (SVM - Support Vector Machine) dla celów klasyfikacji. Pomimo, że w ostatnim czasie metody HOG były z powodzeniem wykorzystywane przede wszystkim do detekcji ruchu pieszych, wyniki eksperymentalne przedstawione w niniejszym artykule pokazują, że połączenie HOG i SVM wydaje się być obiecującą techniką lokalizowania oraz wydzielania graczy w transmitowanym sygnale wizyjnym. Proponowany system detekcji to kompleksowe rozwiązanie wykorzystujące kolor dominujący do segmentacji boiska piłkarskiego, algorytm modelowania 3D boiska oparty na transformacji Hough i dedykowany algorytm do śledzenia graczy. Ocena systemu została przeprowadzana przy użyciu testowych sekwencji telewizyjnych o rozdzielczościach SD (720×576) oraz HD (1280×720). Dodatkowo, skuteczność proponowanego systemu została przetestowana w różnych warunkach oświetlenia (włączając w to niejednolite oświetlenie i wielokrotne cienie graczy) a także zakładając zmienną pozycję kamery.

**Streszczenie ang.:** In this paper a novel segmentation system for football player detection in broadcasted video is presented. The system is based on the combination of Histogram of Oriented Gradients (HOG) descriptors and linear Support Vector Machine (SVM) classification. Although recently HOG-based methods were successfully used for pedestrian detection, experimental results presented in this paper show that combination of HOG and SVM seems to be a promising technique for locating and segmenting players in broadcasted video. Proposed detection system is a complex solution incorporating a dominant color based segmentation technique of a football playfield, a 3D playfield modeling algorithm based on Hough transform and a dedicated algorithm for player tracking. Evaluation of the system is carried out using SD (720×576) and HD (1280×720) resolution test material. Additionally, performance of the proposed system is tested with different lighting conditions (including non-uniform pith lightning and multiple player shadows) and various camera positions.

## **I. Wprowadzenie**

Analiza sekwencji wizyjnych o treści sportowej stała się w ostatnich latach aktywnym tematem badań. Istnieje wiele prac o segmentacji sekwencji wizyjnych przedstawiających rozgrywki piłki nożnej w oparciu o różne techniki. Jednymi z najbardziej interesujących technik są podejścia oparte o analizę kształtu do identyfikacji graczy i piłki w zgrubnie wydzielonym obszarze pierwszego planu [1,2] lub techniki oparte o różne schematy klasyfikacji segmentów przedstawiających piłkarzy i przyporządkowania ich do drużyn [3,4,5]. Większość z istniejących podejść zakłada szczególne warunki, takie jak wiele kamer o nieziennej pozycji, jeden obiekt w ruchu, czy stosunkowo statyczne tło. W transmisjach bezpośrednich meczy piłkarskich te ściśle określone warunki nie są możliwe do spełnienia. Po pierwsze, kamery, które śledzą wydarzenia na boisku nie są umiejscowione na sztywno podążając za akcją. Po drugie, emitowany program jest zbiorem ujęć z wielu kamer zmienianych dynamicznie zgodnie z instrukcjami reżysera emisji. Po trzecie, na boisku znajduje wielu graczy poruszających się w różnych kierunkach. Ponadto, tło w transmisjach rozgrywek piłkarskich może zmieniać się bardzo gwałtownie. Te warunki powodują, że wykrywanie i śledzenie graczy w transmitowanym sygnale jest trudne. Dlatego przyszłe podejścia dla zastosowań dedykowanych do detekcji i śledzeniu obiektów w transmitowanych programach piłki nożnej powinny łączyć wiele różnych technik.

Na podstawie zaobserwowanych charakterystyk różnorodnych transmisji rozgrywek piłkarskich i analizie problemów istniejących algorytmów, proponujemy nowe podejście, które wykorzystuje segmentację w oparciu o kolor dominujący do detekcji boiska piłkarskiego, algorytm detekcji linii w oparciu o transformację Hough do modelowania boiska oraz połączenie technik opisu cech charakterystycznych segmentów za pomocą deskryptora HOG - Histogram of Oriented Gradient deskryptorów [15] wraz z algorytmem klasyfikacji SVM - Support Vector Machine [16] do wykrywania i śledzenia graczy. System przeznaczony jest zarówno do określania położenia i orientacji boiska, jak również wykrywania i śledzenia graczy na boisku.

Oryginalnym wkładem autorów pracy jest zastosowanie technik o niskiej złożoności i znacznym potencjale, wykorzystywanych dotąd głównie do detekcji ruchu pieszych. Dlatego też celem niniejszej pracy jest zbadanie możliwości wyżej wymienionych metod i sprawdzenie, czy proponowane podejście jest wystarczające dla celów segmentacji transmisji sportowych w nadawanym sygnale wizyjnym.

## II. Stan wiedzy

W celu stworzenia kompleksowego systemu segmentacji sekwencji wizyjnych prezentujących piłkę nożną istnieje potrzeba złożenia kilku rodzajów technik.

Jedną z technik stosowanych do wykrywania koloru dominującego jest zastosowanie deskryptora koloru dominującego (DCD) standardu MPEG-7. Deskryptor ten wymaga jednak działania na trójwymiarowej reprezentacji koloru a jego wyniki nie są niezależne od warunków oświetleniowych [6]. Inne podejście oparte jest na wykorzystaniu odległości euklidesowej w przestrzeni IHS do adaptacyjnej detekcji koloru dominującego w obrazie [7]. Praca pod redakcją Ren [8] przedstawia metodę klasyfikacji bloków obrazu na podstawie wariacji składowej odcienia po klasyfikacji wartości odcienia przez wyuczony model mikstur Gaussowskich.

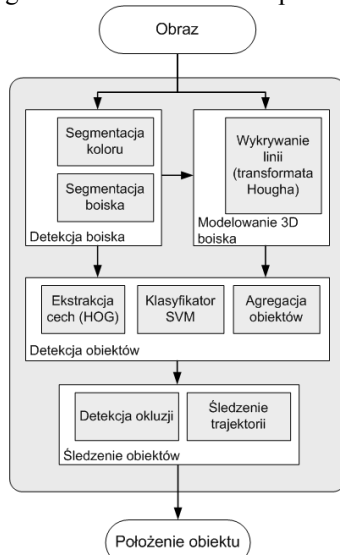
Większość algorytmów wykrywania linii oparta jest na transformacie Hougha obrazu binarnego linii [9], która może wykryć obecność prostej struktury linii i oszacować jej orientację oraz pozycję. Niektóre metody korzystają również ze zmodyfikowanej transformaty Hougha np. probabilistycznego przekształcenia Hougha [10] lub transformaty blokowej Hougha [11] w celu zmniejszenia złożoności obliczeniowej. Thuy [12] zaproponował modyfikację przekształcenia Hougha, która umożliwia wykrywanie odcinka linii zamiast prostej obecności linii. Innym podejściem jest użycie losowej metody poszukiwania. Metoda [13] ta wykorzystuje algorytm losowego wyszukiwania, który wybiera dwa punkty i sprawdza, czy istnieje linia między nimi. Odrębną kwestią jest generowanie obrazu linii. W tym przypadku najlepsze wyniki osiągają algorytmy wykrywania krawędzi i inne metody gradientowe [14].

Detekcja obiektów jest zawsze oparta na ekstrakcji kilku charakterystycznych cech obiektu. Dalal [15] wprowadził deskryptor HOG dla celów detekcji pieszych i osiągnął z jego wykorzystaniem dobre rezultaty.

Kolejnym ważnym problemem w wykrywaniu obiektów jest ich klasyfikacja, której zadaniem jest oddzielenie obiektów należących do różnych klas między sobą. Jednym z najczęściej stosowanych klasyfikatorów obiektów jest klasyfikator SVM [16,17].

## III. System

Szczególne warunki w segmentacji obrazów sekwencji wizyjnej meczy piłkarskich wymagają odpowiedniego podejścia. W związku z tym zaproponowano dedykowany system detekcji piłkarzy (rys. 1). Głównymi składnikami systemu są: detektor boiska, moduł modelowania boiska wraz z dopasowaniem do rzeczywistego jego obrazu oraz moduł rozpoznawania regionów przynależnych do obiektów i moduł śledzenia obiektów.



Rys. 1. Schemat blokowy system detekcji. (Fig. 1. Detection system overview.)

### A. Detekcja boiska

Zakłada się, że obszar boiska jest jednorodnym regionem o stosunkowo jednolitym odcieniu. Dlatego pierwszym etapem wykrywania boiska jest kwantyzacja wektorowa wartości składowych chrominancji. Po kwantyzacji, kolory są klasyfikowane w oparciu o a-priori definicję koloru zielonego. Wspólny obszar wszystkich odcieni zielonego traktowany jest jako wstępna maska boiska. Zakłada się, że obszar boiska ma być największym obszarem zieleni w całym obrazie.

### B. Moduł dopasowania modelu boiska

Dopasowanie modelu wymaga informacji o pozycji niektórych punktów odniesienia w każdej ramce. Punkty te są otrzymywane na podstawie punktów przecinania się linii. Linie boiska są wykrywane przy użyciu zmodyfikowanej transformacji Hougha [11] stosowanej do obrazu binarnego linii. Najpierw, obraz linii boiska jest generowany zgodnie z [14]. Następnie, obraz jest poddawany progowaniu i algorytmowi morfologicznego pocieniania. W celu odrzucenia wykrytych fałszywych nieistniejących linii, binarny obraz linii dzielony jest na bloki. Dla każdego bloku określany jest niezależnie zestaw parametrów linii a następnie są one używane w procedurze głosowania w zmodyfikowanej transformacji Hougha. Wstępni kandydaci linii są wybierani poprzez wyszukiwanie lokalnych maksimów w przestrzeni parametrów transformacji Hougha. Aby osiągnąć większą dokładność, parametry kandydatów linii są poprawiane z wykorzystaniem regresji liniowej. Kandydaci ze zbyt dużym błędem regresji są odrzucani. Ostatnim krokiem jest agregacja kandydatów linii i śledzenie.

W systemie wykorzystywany jest predefiniowany model pola karnego, o odpowiednich proporcjach rozmiarów, jako wzorzec. Szablon jest dopasowany do każdej możliwej pozycji boiska na podstawie orientacji wykrytych linii. Dla każdej iteracji obliczany jest koszt dopasowania modelu i pozycja o najmniejszym koszcie dopasowania jest uważana za ostateczne rozwiązanie.

### C. Rozpoznawanie obszarów obiektów

Moduł wykrywania graczy oparty jest na deskrypcji HOG [15]. Kształt gracza jest reprezentowany przez okno o rozmiarze  $16 \times 32$  punktów. W celu wykrycia graczy o różnych rozmiarach stosowany jest detektor o wielu skalach. Gdy określany jest deskryptor HOG dla analizowanego obszaru obrazu, liniowy klasyfikator SVM jest szkolony na specjalnej bazie danych szablonów piłkarzy do sklasyfikowania danego obszaru jako gracza lub klasy nie-gracza. Szablony zostały uzyskane przez manualną segmentację sekwencji obrazów na pozytywne i negatywne przykłady graczy. W celu wykrycia graczy w różnych pozach wykorzystano trzy klasyfikatory SVM. Pierwszy z nich był szkolony do wykrywania piłkarzy w pozycji pionowej od przodu, drugi dedykowano pionowemu profilowi a ostatni do łącznego wykrycia wszystkich pionowych póz. Wszystkie klasyfikatory SVM wykorzystują ten sam zestaw próbek negatywnych. Wszystkie trzy detektory SVM pracują równolegle, a ich wyniki są sumowane dla celów wygenerowania jednego zestawu wyników wykrywania dla naszego systemu detekcji.

Jak zauważyliśmy jeden gracz może powodować wiele wykryć w czasie i, w konsekwencji obszar wynikowy dla detekcji pojedynczego gracza może się nakładać w wielu przypadkach. W celu uzyskania pojedynczego prostokąta wokół wykrytego gracza zaproponowano dodatkową procedurę łączenia tych obszarów. Procedura składa się z następujących etapów: filtracja pola detekcji (pola o nieodpowiednim rozmiarze lub zawierające zbyt wiele punktów boiska są odrzucane), test pokrywania się obszarów dla każdej pary wykrytego obszaru (w tym ustalenie największych spójnych przestrzeni punktów nie należących do boiska w każdym prostokątnym obszarze i testowanie czy te obszary się pokrywają) i wreszcie etap agregacji pola (dwa pokrywające się pola zostają scalone w jeden wynikowy prostokąt).

### D. Śledzenie obiektów

Śledzenie obiektów jest stosowane w celu poprawy odporności systemu dla przypadków przesłaniania się graczy i zmian ostrości wywołanych gwałtownym ruchem kamery lub powiększaniem. W proponowanym podejściu każdy obiekt jest opisany przez położenie i rozmiar pola detekcji ograniczającym ten obiekt. W każdej chwili czasu, obiekty, które są obecnie śledzone przez algorytm są porównywane z polami znalezionymi przez detektor SVM w analizowanej ramce. Dla każdego obiektu określone jest podobieństwo prostokątnych pól pomiędzy polem skojarzonym z obiektem i polem kandydatem z wykorzystaniem specjalnej funkcji kosztu. Funkcja obejmuje zarówno pokrywający się obszar i rozmiar dwóch prostokątnych pól. Po dokonaniu oceny kosztu, pole kandydata z minimalnym kosztem jest wybierane do reprezentowania śledzonego obiektu. Dodatkowo, każdy obiekt ma przypisany wektor ruchu. Wektor ruch jest obliczany na podstawie położenia obiektu w poprzednich klatkach i w konsekwencji, jeśli obiekt nie jest dopasowany do żadnego wykrytego prostokątnego pola w bieżącej ramce, jego pozycja jest określana na podstawie tego wektora ruchu.

## IV. Ocena systemu detekcji graczy

Skuteczność systemu została oceniona przy użyciu zarówno sekwencji o rozdzielczości SD ( $720 \times 576$ ) jak i HD ( $1280 \times 720$ ). Ponadto, przygotowany zbiór sekwencji testowych składa się z materiału reprezentującego różne warunki oświetleniowe, takie jak np. nierównomiernie oświetlone boisko, wiele cieni graczy, jak również założono różne położenia kamer wykorzystanych do rejestracji widowiska sportowego. Biorąc powyższe warunki pod uwagę, 9 sekwencji testowych z imprez piłkarskich o długości od 25 do 50 klatek zostało wybranych, tworząc tym samym zestaw testowy do oceny systemu. Dla każdej sekwencji testowej ręcznie wyselekcjonowano bazę uwiarygodnionych obszarów wskazujących piłkarzy.

Prezentowany system jest oceniany z wykorzystaniem dwóch miar - parametrów *precision* oraz *recall*, definiowanych w następujący sposób:

$$precision = \frac{TP}{TP + FP}, \quad (1)$$

$$recall = \frac{TP}{TP + FN}, \quad (2)$$

gdzie  $TP$  jest zbiorem pozytywów prawdziwych (poprawne detekcje),  $FP$  – zbiór pozytywów fałszywych (fałszywe detekcje) oraz  $FN$  – zbiór negatywów fałszywych (zgubione obiekty) definiowane jako:

$$TP = \{r | r \in D: \exists g \in G: s_{\theta}(r, g) \geq T\}, \quad (3)$$

$$FP = \{r | r \in D: \forall g \in G: s_{\theta}(r, g) < T\}, \quad (4)$$

$$FN = \{r | r \in G: \forall g \in D: s_{\theta}(r, g) < T\}. \quad (5)$$

W powyższych równaniach:  $s_{\theta}(a, b)$  określa stopień pokrywania się pomiędzy dwoma obszarami  $a$  i  $b$  (tj. prostokątów ograniczających wykryte obiekty):

$$s_{\theta}(a, b) = (a \cap b) / (a \cup b) \quad (6)$$

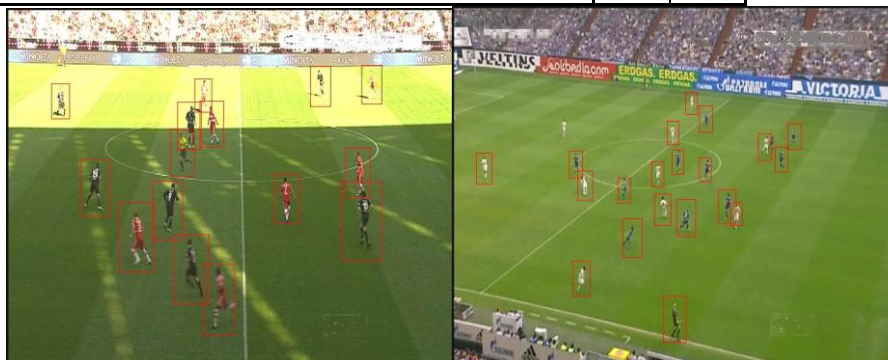
$T$  wyznacza próg, który określa stopień pokrywania wymagany aby założyć, że dwa obszary przesłaniają ten sam obszar analizowanego obrazu.  $G$  jest zbiorem uwiarygodnionych obszarów a  $D$  jest zbiorem wykrytych obszarów dla danej ramki, definiowanymi jako:  $G = \{g_1, \dots, g_n\}$  i  $D = \{d_1, \dots, d_m\}$ , z  $n$  wskazującym liczbę uwiarygodnionych obszarów i  $m$  – liczbę obszarów wykrytych w analizowanej ramce.

Do oceny systemu wykrywania piłkarzy prezentowanych w tej pracy użyto progu  $T$  o wartości równej 0,4. Wartość ta została wybrana na podstawie interpretacji parametru  $T$ :  $T = 0,4$  oznacza, że ponad połowa z obszarów prostokąta ograniczającego wykryty obszar oraz manualnie wskazanego uwiarygodnionego obszaru pokrywa się. Wartość ta jest wystarczająca, aby ocenić, jak skutecznie obiekty są wykrywane przez system. Jednakże w przypadku oceny systemu dedykowanego do segmentacji obiektów i jego wydobywania z tła powinny być stosowane wyższe wartości  $T$ .

W tabeli 1 przedstawiono szczegółowe wyniki oceny systemu detekcji w odniesieniu do każdej sekwencji zawartej w zbiorze testowym.

**Tabela 1.** Rezultaty oceny systemu detekcji dla progu  $T=0.4$  (P-precision, R-recall). Detection system evaluation results for threshold  $T=0.4$  (P-precision, R-recall).

Sekwencja testowa	P	R
1. szybki obrót kamery, jednolite oświetlenie	1.00	0.84
2. niejednolite oświetlenie, cienie, przesłanianie się obiektów	0.97	0.90
3. przesłanianie się obiektów, różne pozy graczy	1.00	0.88
4. niejednolite oświetlenie, przesłanianie się obiektów	0.97	0.94
5. niejednolite oświetlenie, przesłanianie się obiektów	1.00	0.87
6. przepłot, jednolite oświetlenie	1.00	1.00
7. rozmyty ruch, małe postacie, przesłanianie się obiektów	0.96	0.80
8. rozmyty ruch, przesłanianie się obiektów, jednolite oświetlenie	0.96	0.67
9. zielone stroje, przesłanianie się obiektów	0.96	0.96
<b>Średnia</b>	0.98	0.87



Rys. 2. Wizualizacja wyników detekcji piłkarzy. Fig. 2. Visualization of the player detection results

Wyniki oceny przedstawione powyżej pokazują wyraźnie, że liczba fałszywych wykryć w systemie jest stosunkowo niewielka, co przekłada się na wartość parametru *precision* przekraczającego 0,95 we wszystkich analizowanych przypadkach. Jednak liczba niewykrytych obiektów zmienia się znacząco wśród zestawu testowego - patrz wartości parametru *recall* w tabeli 1. Jest to głównie spowodowane przesłanianiem się obiektów w scenie.

Oprócz obiektywnych testów, wydajność proponowanego systemu wykrywania piłkarzy oceniano również subiektywnie – rysunek 2.

## V. Podsumowanie

W pracy zaproponowano system segmentacji do detekcji i śledzenia piłkarzy w zdekompresowanym sygnale wizyjnym. Wyniki pokazują, że proponowane rozwiązanie umożliwia osiągnięcie wysokiej obiektywnej i subiektywnej oceny w zakresie dokładnej lokalizacji wykrytych obiektów, jednak liczba niewykrytych obiektów powinna jeszcze zostać zredukowana.

## **Bibliografia**

- [1] Haiping, S., Lim, J. h., Tian, Q., Kankanhalli, M. S.: Semantic labeling of soccer video, Proceedings of IEEE Pacific-Rim Conference on Multimedia ICICS-PCM, 1787-1791 (2003)
- [2] Huang, Y., Llach, J., Bhagavathy, S.: Players and Ball Detection in Soccer Videos Based on Color Segmentation and Shape Analysis, Lecture Notes in Computer Science, Volume 4577/2007, 416-425 (2007)
- [3] Nuñez, J. R., Facon, J., Brito Junior, A. d. S.: Soccer Video Segmentation: referee and player detection, 15th International Conference on Systems, Signals and Image Processing, 2008. IWSSIP 2008, 279 – 282 (2008)
- [4] Vandembroucke, N., Ludovic, M., Postaire, J-G.: Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis, Computer Vision and Image Understanding 90, 190–216 (2003)
- [5] Guangyu, Z., Changsheng, X., Qingming, H., Wen, G., Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter, Int. Conf. Multimedia & Expo, 1629-1632 (2006)
- [6] Hong, S., Yueshu, W., Wencheng, C., Jinxia, Z.: Image Retrieval Based on MPEG-7 Dominant Color Descriptor, ICYCS, 753-757 (2008)
- [7] Ying, L., Guizhong, L., Xueming, Q.: Ball and Field Line Detection for Placed Kick Refinement, GCIS, vol 4, 404-407 (2009)
- [8] Ren, R., Jose, J.,M.: Football Video Segmentation Based on Video Production Strategy, Lecture Notes in Computer Science, 3408, 433-446
- [9] Candamo, J., Kasturi, R., Goldgof, D., Sarkar, S.: Detection of Thin Lines using Low-Quality Video from Low-Altitude Aircraft in Urban Settings, Aerospace and Electronic Systems, IEEE Transactions on, vol.45, no.3, pp.937-949, July (2009) Guo, S., Y., Kong, Y.,G., Tang, Q., Zhang, F.: Probabilistic Hough transform for line detection utilizing surround suppression, International Conference on Machine Learning and Cybernetics (2008)
- [10] Yu, X., Lai, H.C., Liu, S.X.F., Leong, H.W.: A gridding Hough transform for detecting the straight lines in sports video. ICME (2005)
- [11] Thuy, T., N., Xuan, D., P., Jae, W., J.: An improvement of the Standard Hough Transform to detect line segments, ICIT (2008)
- [12] Jiang, G., Ke, X., Du, S., Chen, J.: A straight line detection based on randomized method, ICSP (2008)
- [13] Li, Q., Zhang, L., You, J., Zhang, D., Bhattacharya, P.: Dark line detection with line width extraction, ICIP (2008)
- [14] Dalal N., Triggs B., Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition 1, 886-893 (2005)
- [15] Yu-Ting Ch., Chu-Song Ch.: Fast Human Detection Using a Novel Boosted Cascading Structure With Meta Stages, IEEE Transactions on Image Processing 17, 1452-1464 (2008)
- [16] Paisitkriangkrai S., Shen, C. Zhang, J.: Performance evaluation of local features in human classification and detection, IET Computer Vision 2, 236-246 (2008)