Riku Valleala

URHEILUVIDEOIDEN SISÄLLÖNTUNNISTUS JA LUOKITTELU

Tietojärjestelmätieteen
kandidaatintutkielma

21.12.2008

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Jyväskylä

TIIVISTELMÄ

Valleala, Riku Jouni Sakari
Tietojärjestelmätieteen kandidaatintutkielma / Riku Valleala
Jyväskylä: Jyväskylän yliopisto, 2008.
33 s.
Urheiluvideoiden sisällöntunnistus ja luokittelu

Urheiluvideot ovat yksi nopeimmin kasvavista ja kiinnostavimmista videosisältöjen alueista tänä päivänä. Videosisältöjen automaattisille ja tehokkaille luokittelumenetelmille on siten lisääntyvää kysyntää. Tämän kandidaatintutkielman tavoitteena on selvittää videoiden sisällöntunnistukseen ja luokitteluun kehitettyjä menetelmiä ja niiden soveltamista urheiluvideoiden sisällöntunnistuksessa. Aihealuetta käsitellään kirjallisuuskatsauksen tarjoamin menetelmin. Videoiden sisällöntunnistus ja luokittelu aloitetaan yleensä video-otosten tunnistamisella, josta edetään tunnistettujen otosten sisällöntunnistukseen ja niin sanottujen alemman tason muuttujien hakemiseen. Analysointimenetelminä voidaan käyttää erilaisia väri-informaatioon, liikkeisiin, valoisuuteen, tekstiin tai ääniin liittyviä menetelmiä. Sisällöntunnistuksen viimeisimpänä tavoitteena on yleensä korkeamman tason merkityksellisten kohteiden tai tapahtumien tunnistaminen videosisällöistä. Urheiluvideoiden sisällöntunnistusta vaikeuttavat lajien monimuotoisuus omine sääntöineen ja ympäristöineen. Kuvaa, ääntä ja tekstejä analysoivat multimodaaliset lähestymistavat tarjoavat kuitenkin monipuolisen mahdollisuuden urheiluvideoidenkin sisällöntunnistukseen. Videoiden sisällöntunnistuksen menetelmät ja käsitteet vaativat kuitenkin jatkossa yhdenmukaistamista ja standardointia entistä parempien menetelmien kehittämiseksi.

AVAINSANAT: video, sisällöntunnistus, luokittelu, urheilu

SISÄLLYSLUETTELO

1 JOHDANTO

Videomateriaalin määrä maailman tietoverkoissa kasvaa koko ajan kiihtyvällä vauhdilla. Urheiluvideot ovat yksi suosituimmista ja markkinalähtöisimmistä videomateriaaleista, joita erilaiset verkkopalvelut ihmisille tarjoavat. Kuluttajien vaatimukset videoiden hakemiseen ja selailuun omien mieltymystensä mukaisesti lisääntyvät jatkuvasti. Harva enää haluaa selata pikakelauksella lävitse koko jalkapallo-ottelua nähdäkseen vain oman suosikkijoukkueensa maalintekotilanteet. Erilaiset valmiit videosisältöjen tunnistamiseen ja luokitteluun pyrkivät järjestelmät tarjoavat sen sijaan valmiiksi luokiteltuja videosisältöjä, jolloin juuri haluttujen tapahtumien ja tilanteiden hakeminen on mahdollista.

Urheiluvideoiden sisällöntunnistus ja luokittelu on hyvin aktiivinen tutkimusalue, jossa yhdistyvät videotekniikan, signaalinkäsittelyn, tilastotieteen ja ohjelmoinnin osa-alueet. Manuaaliset menetelmät urheiluvideoiden luokitteluun ja tiedonhakuun ovat useimmiten aivan liian hitaita ja kalliita toteuttaa, jolloin vaihtoehdoksi nousevat erilaiset automaattiset menetelmät. Videosisältöjen luokittelua on tehty jo suhteellisen pitkään esimerkiksi videovalvontamateriaalin analysointiin liittyen. Uutena sovellusalueena mainittakoon myös tekijänoikeuksin suojatun videomateriaalin tunnistaminen verkkopalveluissa (MikroPC.net IT-uutiset 2008). Erilaisten videokirjastojen ja uutispalveluiden lisäksi automaattista sisällöntunnistusta ja luokittelua voidaan käyttää apuna muun muassa urheiluvalmennuksessa erilaisissa pelitaktiikka- ja tekniikka-analyyseissä sekä suoritusanimaatioissa jopa 3D-tekniikan keinoin (Yu & Farin 2005). Tämän tutkielman tekemiseen motivoi kirjoittajaa erityisesti myös hänen päivätyönsä soveltavan urheilututkimuksen ja videopohjaisten tekniikka-analyysien parissa.

Aihealueeseen liittyy olennaisena käsitteenä *video*, joka tarkoittaa teknologiaa, jossa sähköisiä signaaleja käsitellään niin, että niistä saadaan liikkuvia kuvia. Pakkaamaton videokuva koostuu yksittäisistä kuvista, joita toistetaan riittävän tiheästi liikkeen vaikutelman tuottamiseksi. Toinen tärkeä yleiskäsite on *sisällöntunnistus (feature extraction)*, jolla tarkoitetaan toimintaa, jossa erilaisilla signaalinkäsittelyn menetelmillä ja algoritmeilla haetaan tiettyjä muutoksia ja tapahtumia videokuvassa. Nämä

menetelmät perustuvat muun muassa värialueiden, valoisuuden ja liikevektoreiden analysointiin joko pikseli-pikseliltä tai laajemmissa kuvan osioissa. Videoiden *luokittelu (classification ja segmentation)* käsitetään tässä yhteydessä videon jakamisena osiin esimerkiksi sisällöntunnistuksella löydettyjen video-otosten tai muilla algoritmeilla analysoitujen tapahtumien mukaisesti.

Tutkielman tavoitteena on selvittää videoiden automaattiseen sisällöntunnistukseen ja luokitteluun kehitettyjä menetelmiä, sekä tarkastella näiden menetelmien hyödyntämistä urheiluvideoiden sisällöntunnistuksessa. Aiheen käsittelyssä pitäydytään sisällöntunnistuksessa ja luokittelussa käsittelemättä muuten kuin välttämättömiltä osiltaan videosisältöjen indeksointitapoihin tai tiedonhakuun liittyviä tekijöitä. Ensin kerrotaan digitaaliseen videoon ja sen luokitteluun liittyvistä standardeista. Sen jälkeen esitellään yleisimmät sisällöntunnistuksen menetelmät, josta siirrytään edelleen näiden menetelmien soveltamiseen urheiluvideoissa. Lopuksi arvioidaan pohdinnan kautta tutkielman menetelmänä käytetyn kirjallisuuskatsauksen antamaa kuvaa aihealueesta sekä tehdään yhteenveto koko työn sisällöstä.

Tutkielman tulokset antavat hyvän yleiskuvan videoiden sisällöntunnistuksessa ja luokittelussa käytettävistä menetelmistä ja niiden soveltamisesta urheiluvideoihin. Tulokset osoittavat kyseisen alueen olevan erittäin laajan, ja sen, että erilaisia menetelmiä videosisältöjen tunnistamiseen ja luokitteluun on valtaisa määrä. Yhtenäisiä malleja tai standardeja menetelmien yhtenäistämiseksi ei ole juurikaan esitetty. Urheiluvideoiden kohdalla oman erityistarpeensa antavat lajikohtaiset säännöt ja rajoitukset. Kirjallisuus ja tehdyt tutkimukset näyttävät selvästi sen, että urheiluvideoiden sisällöntunnistuksessa on keskitytty pitkälti joukkuelajeihin. Yksilölajien kohdalla olisi varmasti paljon potentiaalista käyttäjäkuntaa videoiden sisällöntunnistukselle, mutta alue on vielä selvästi hyödyntämättä.

2 DIGITAALISEEN VIDEOON JA SEN LUOKITTELUUN LIITTYVÄT STANDARDIT

Videoiden sisällöntunnistukseen liittyy olennaisena osana itse videotiedosto, jonka tiedostomuotojen määrittelyyn on olemassa hyväksyttyjä standardeja. Näiden mukaan määräytyvät videovirran tekniset ratkaisut ja käytettävissä olevat videon ja äänen pakkausmenetelmät. Yleisimmin käytetyt videon tiedostomuodot ovat AVI ja MPEG eri versioineen. Videoiden luokitteluun ja erityisesti niiden sisältöjen hakemiseen liittyy osaltaan myös ohjelmointi- ja sisällönkuvauskieliä, joista yleisimmin on käytetty XML-kieltä. Seuraavaksi perehdytään videon tiedostomuotoihin ja videosisältöjen luokitteluun liittyviin standardeihin.

2.1 AVI-tiedostomuoto

Digitaalisen videon käsittelyssä perinteisin tiedostomuoto on AVI, joka tulee sanoista Audio Video Interleave. Tiedostomuoto on Microsoftin vuonna 1992 kehittämä formaatti, jolla voidaan yhdistää kuva ja ääni, sekä toistaa niitä synkronoituna toisiinsa. Useimmat AVI-tiedostot tukevat lisäksi OpenDML-tiedostomuotoa, joka tarjoaa muun muassa mahdollisuuden yli yhden gigabitin tiedostokokoihin sekä muihin edistyksellisempiin toimintoihin (OpenDML AVI M-JPEG File Format Subcommittee 1996). Kyseinen AVI-tiedostomuoto tunnetaan myös nimillä AVI 2.0 ja AVI2.

Tiedostomuotona AVI on laajalle levinnyt ja paljon käytetty videotiedoston muoto sen helppokäyttöisyyden ja erittäin hyvän yhteensopivuuden johdosta. Lisäksi pakkaamattoman AVI on esimerkiksi videotuotannoissa edelleen suosituin videomuoto sen hyvän kuvanlaadun vuoksi. AVI ei itsessään sisällä varsinaista pakkausta ja purkua helpottavaa kodekkia, vaan siihen käytettävät kodekit ovat erillisiä. Näiden perinteisten kodekkien teho on nykyajan edistyksellisempiin pakkausmenetelmiin verrattuna varsin heikko. Tämän vuoksi digitaalisen videon tiedostomuotona AVI:n suosio on selvästi vähenemässä. Nykyaikaisten digitaalisten videokameroiden ominaisuuksien säilyttämiseksi on kehitetty lisäksi DV AVI -tiedostomuoto, joka perustuu pitkälti alkuperäiseen AVI-tiedostomuotoon. (Windows Hardware Developer Central 2001)

2.2 MPEG-tiedostomuoto

Sana MPEG tulee englanninkielisistä sanoista Motion Picture Experts Group. Kyseinen, vuonna 1988 perustettu työryhmä on merkittävä ISO/IEC:n jäsen vastaten digitaalisen videokuvan ja siihen liittyvän äänen pakkaus- ja purkumenetelmien ja käsittelytapojen kansainvälisestä standardoinnista. Tähän mennessä ryhmä on tuottanut MPEG-1, -2, 4-, -7 ja MPEG-21 videostandardit. Näillä kaikilla standardeilla on pyritty nykyaikaiseen ja korkealaatuiseen videokuvan ja äänen pakkaamiseen. Tiedostokoko pyritään toisin sanoen minimoimaan ja samalla pitämään kuva mahdollisimman hyvälaatuisena. Siten lähes kaikista MPEG-videostandardeista on tullut erittäin suosittuja videokuvan käsittelyssä ja jakamisessa niin television kuin internetinkin välityksellä. (Motion Picture Experts Group, 2008)

Videokuvan pakkaaminen MPEG-videoissa perustuu häviölliseen DCT-algoritmiin (Discrete Cosine Transformation), jonka avulla kuvasta karsitaan pois ylimääräistä tietoa. Kuva jaetaan 8x8 pikselin alueisiin, joille jokaiselle lasketaan DCT-kertoimet ja kuva esitetään varsinaisesti kertoimista koostuvana matriisina. Ihmissilmän epätarkkuutta voidaan täten hyödyntää koodaamalla eri kuvalohkoja samalla väriarvolla ja kertoimien keskiarvoilla. Lisäksi MPEG-videon sisältö perustuu kolmen erilaisen kuvan koodaamiseen. Niin sanotut I-kuvat ovat yksinkertaisia, täyden informaation sisältäviä, pysäytyskuvia. P-kuvat ovat sen sijaan ennustettuja kuvia perustuen lähimpään edeltävään I- tai P-kuvaan. Kolmantena muotona ovat kaksisuuntaisesti muodostetut B-kuvat, jotka ennustetaan ja interpoloidaan kahdesta lähimmästä I- tai P-kuvasta. Tyypillinen koodattujen kuvien sarja näyttää seuraavalta: IBBPBBPBBPBBIBBPBBPB. Edellä mainittua I- ja P-kuvien välissä tapahtuvaa kuvan muutosta kuvataan liikevektorilla sekä pysty- että vaakasuunnassa ja sitä täydennetään ennustuksen virhekertoimella, jonka koodauksessa käytetään hyväksi DCT-menetelmää. (Berkeley Multimedia Research Center 2008)

MPEG-1 on ensimmäinen työryhmän valmis standardi. Se on suunniteltu digitaalisen videokuvan ja sen äänen pakkaamiseen ja toistamiseen aina 1,15 megabittiä/sekunti datavirtaan saakka. Kyseinen datavirta riittää hyvin esimerkiksi VHS-videomateriaalin tasoisen kuvan pakkaamiseen ja purkamiseen. MPEG-1 saavutti aikanaan runsaan

suosion erilaisten videomateriaalien pakkausmuotona erityisesti CD-ROM levyille tallennetuissa videoissa. Tiedostomuoto on edelleen varsin laajalti käytetty, koska siihen perustuvat videotiedostot ovat edelleen toistettavissa käytännössä lähes jokaisella tietokoneelle. MPEG-1:n audiopakkauksen ”Kerros III” –pakkausmuoto on nykyisen, muun muassa erilaisissa kannettavissa musiikkilaitteissa suosittavan, mp3-pakkauksen alkuperäismuoto. (ISO/IEC 1992)

MPEG-2 standardin spesifikaatio valmistui kokonaisuudessaan vuonna 1996. Standardi sisältää jo selvästi suuremman määrän eri osioita kuin MPEG-1. DVD-elokuvissa sekä nykyisessä digi-TV:ssä käytetään edelleen MPEG-2 standardin mukaista pakkausmenetelmää. Tärkeänä MPEG-2:n etuna on muun muassa se, että formaatti pystyy käsittelemään värejä jopa studiotason väriresoluutiolla 4:2:2. Formaatti tukee lisäksi monikanavaääntä ja videovirtaa voidaan koodata maksimissaan nopeuteen 8 megabittiä/sekunti saakka. (ISO/IEC 1994)

MPEG-4 sisältää useita MPEG-1:n ja MPEG-2:n osia ja ominaisuuksia. Tämän lisäksi standardin toiminnallisuuksia on laajennettu huomattavasti verrattuna MPEG-2:een. MPEG-4 tiedostomuoto sisältää muun muassa VRML-tuen (Virtual Reality Modelling Language) 3-ulotteisten kohteiden laskentaan, objekti-pohjaisen yksittäisten kuvan osien koodauksen, useita interaktiivisia toimintoja sekä digitaalisen tekijänoikeuksien hallinnan. MPEG-4 standardin eri osat käyttävät hyväkseen uusia pakkaus- ja purkukodekkeja, joka on aiheuttanut ongelmia monille kyseisen formaatin käyttäjille MPEG-4:n kehityksen alkuvaiheissa. Standardin mukaisia videotiedostoja on alettu käyttää hyvin aktiivisesti erityisesti mobiilivideoissa sekä erilaisissa videopalveluissa. Suosio perustuu pitkälti sen monipuolisiin ominaisuuksiin sekä tehokkaaseen pakkausmenetelmään, jolla saavutetaan esimerkiksi MPEG-2:een verrattuna sama kuvanlaatu, mutta puolet matalammalla datavirralla. (ISO/IEC 1999)

MPEG-7 ei ole varsinainen videostandardi perinteisessä mielessä, vaan multimediasisältöjen kuvaustavan oma standardi. Se määrittelee, miten videoon liittyvä metatieto muodostetaan ja miten sitä hyödynnetään esimerkiksi videon sisältöön kohdistuvissa hakumenetelmissä. MPEG-7:ssä hyödynnetään erilaisia työkaluja, jotka ovat kuvaaja (descriptor), kuvausmalli (description scheme), XML-kieleen perustuva kuvauksen määrittelykieli (description definition language) sekä muutamia

järjestelmätyökaluja. MPEG-7 mahdollistaa videosisältöihin tehtäviä hakuja, jota voidaan hyödyntää esimerkiksi digitaalisissa multimediakirjastoissa, tv-lähetysten mediavalinnoissa, kodin viihdelaitteissa sekä verkko-opiskelussa. (ISO/IEC 2001)

MPEG-21 tarjoaa eräänlaisen toimintakehyksen (multimedia framework) erilaisten multimediasisältöjen tuottamiseksi, jakamiseksi ja käyttämiseksi eri toimijoiden kesken. MPEG-21:n mukaan tällainen ympäristö vaatii *käyttäjiä* ja *digitaalisia kohteita*, joihin käyttäjät kohdistavat *toimenpiteitä* tuottaen uusia *digitaalisia kohteita*, joista tulee edelleen toiminnan kohteita. Tämä kaikki onnistuu vain tarkasti määritellyillä teknologioilla ja järjestelmillä, jotka MPEG-21 pyrkii standardissaan kuvaamaan. (ISO/IEC 2002)

2.3 XML-kieli

XML (Extensible Markup Language) on SGML:n (Standard Generalized Markup Language) osajoukko, joka on luotu rakenteisen tiedon esittämiseen. Sitä alettiin kehittää jo vuonna 1996 ja tällä hetkellä on julkaistu XML 1.0 -spesifikaation 5. versio. Rakenteisissa dokumenteissa rakenne, sisältö ja ulkoasu ovat toisistaan eroteltavissa ja erikseen kuvattuina. XML:n avulla esitettävä tieto pystytään ryhmittelemään eräänlaisiksi sisältöyksiköiksi, joita kutsutaan XML-dokumenteiksi. Näin muodostettuja dokumentteja luetaan XML-prosessoriksi kutsutuilla ohjelmisto-osioilla, jolloin päästään käsiksi dokumenttien sisältöön ja rakenteeseen. Spesifikaatio kuvaa osaltaan myös sen, miten XML-prosessorin tulee lukea XML-data ja mitä tietoa sen tulee välittää varsinaiselle isäntäsovellukselle. (W3C 2006)

XML-dokumenttien fyysinen rakenne koostuu entiteeteistä, jotka voivat olla tiedostoja sisältäen tekstiä ja muuta dataa tai nimettyjä merkkijonoja. Entiteetit voivat olla jäsennettyjä tai jäsentämättömiä, sisäisiä tai ulkoisia sekä yleisentiteettejä tai parametrientiteettejä. Dokumenttien loogisen rakenteen osia ovat sen sijaan elementit, jotka merkitään HTML-syntaksin tapaisella merkkauksella. Näin yksittäinen elementti saa nimen ja sisällön sekä mahdollisesti attribuutteja ja lapsielementtejä. Tieto voidaan tallentaa elementtiin siten sekä sisältönä että mahdollisesti attribuutin arvona. (Harold 2001)

Luokiteltujen ja indeksoitujen videosisältöjen haussa voidaan hyödyntää muun muassa XML:n laajennosta XQuerya. Se on monipuolinen XML-pohjainen kyselykieli tiedon hakemiseen rakenteisista ja puolirakenteiseista dokumenteista, relaatiotietokannoista sekä oliotietokannoista. XQueryn avulla voidaan kyselyitä tehdä riippumatta siitä onko kohdeaineisto suoraan XML-muodossa vai esitettynä XML-muodossa väliohjelman avulla. Kieli on suunniteltu helposti toteutettavaksi kieleksi, jossa kyselyt ovat tiivistettyjä ja helposti ymmärrettäviä. (W3C 2007)

3 VIDEOIDEN SISÄLLÖNTUNNISTUKSEN MALLIT

Laajojen videotietokantojen sisällön luokittelu mielekkäästi haettavaan muotoon vaatii tehokkaita sisällöntunnistuksen ja luokittelun menetelmiä. Brunellin, Michin ja Modenan (1999) mukaan digitaaliset tekstidokumentit ovat periaatteessa itse itseään sisältönsä avulla kuvaavia, mutta videotiedostojen sisällönluokittelu on paljon työläämpää ja sen tekeminen manuaalisesti olisi luonnollisesti tarkinta, mutta erittäin hidasta ja kallista. Automaattiset menetelmät tarjoavat siten selvästi kustannustehokkaamman ja nopeamman vaihtoehdon videosisältöjen luokittelussa.

Videotiedostojen jakaminen osiin yleisellä tasolla perustuu Snoekin & Worringin (2005) mukaan usein erilaisiin muotojen ja kuvioiden analysointimenetelmiin. Usein käytettyjä menetelmiä ovat mallien vertailu, tilastollinen luokittelu, rakenteellinen vertailu ja neuroverkkotekniikat. Näistä tilastollinen lähestymistapa on useimmiten käytetty yksittäinen osa-alue, jonka tekniikoita ovat muun muassa bayeslainen luokittelija (Bayes classifier), päätöspuu (Decision tree), K:n lähin naapuri (k-Nearest neighbor) ja kätketty Markovin malli (Hidden Markov Model, HMM).

Videosisältö on luonteeltaan moniulotteista sisältäen kuvaa, videota, ääntä ja tekstiä. Debin & Zhangin (2005) mukaan yleinen videoiden sisällöntunnistus ja tiedonhaku vaatii seuraavat neljä vaihetta: 1. Yksittäisen video-otoksen alku- ja loppukohtien tunnistaminen, 2. Avainkuvien valinta näistä otoksissa, 3. Avainkuvien ominaisuuksien ja sisällön analysointi, sekä 4. Sisältöperusteinen tiedonhaku videoista. Video-otosten tunnistamisella video saadaan jaettu mielekkäisiin osiin, jonka jälkeen on helpompi aloittaa niin sanottujen alemman tason muuttujien (värit, muodot, tekstit) tunnistaminen (Brunelli ym. 1999). Toisaalta jo pelkästään videon jakaminen haluttuihin otoksiin voi tarjota eräänlaisen tiivistelmän koko videosta, joka jo sinänsä helpottaa videosisältöjen hakua ja selaamista. Otoksen tunnistamiseen on käytetty erilaisia menetelmiä riippuen siitä, onko käytetty video ollut pakattua vai pakkaamatonta videodataa (Deb & Zhang 2005).

Yleensä otoksiin jakamisvaihetta seuraa Debin & Zhangin (2005) mukaan kunkin otoksen avainkuvan valinta perustuen lähinnä liikkeen, värin ja muotojen analysointiin, jotta otosta parhaiten edustava avainkuva löydetään. Usein tämä avainkuva on otoksen

ensimmäinen kuva. Seuraavaksi sisällöntunnistuksella pyritään erilaisten laskentamallien avulla alemman tason muuttujien erotteluun ja siitä edelleen haluttujen ylemmän tason kohteiden semanttisen käyttäytymisen selvittämiseen.

Debin & Zhangin (2005) mukaan videon oikeiden semanttisten sisältöjen tunnistaminen ja luokittelu kaipaa edelleen selkeitä ja hyväksyttyjä lähestymistapoja. Tämä on kuitenkin hänen mukaansa edelleen haasteellista koska videosisältöjen tunnistettujen alemman tason muuttujien, kuten värit, muodot ja tekstit, sekä edistyksellisempien ylemmän tason muuttujien, kuten esimerkiksi pöytä, tuoli, auto ja talo, väliltä puuttuu selkeä yhteys. Voidaan kuitenkin todeta, että muutamia menetelmiä alemman ja ylemmän tason muuttujien automaattiseksi yhdistämiseksi on jo kirjallisuudessa raportoitu (mm. Zhou & Dao 2001, Zhang, Tan, Smoliar & Yihong 1995 sekä Tjondronegoro, Chen & Pham 2005).

4 YLEISET SISÄLLÖNTUNNISTUKSEN JA LUOKITTELUN MENETELMÄT

Videoiden sisällöntunnistusta voidaan tehdä monenlaisin menetelmin. Koska videosisältö on lähes aina multimodaalista sisältäen kuvaa, ääntä ja tekstiä, voidaan sisällöntunnistuksen ja luokittelun erilaiset menetelmät jakaa käytännössä kolmeen kategoriaan. Suurin osa erilaisista sisällöntunnistuksen menetelmistä keskittyy videon kuvasisältöön visuaalisin menetelmin, mutta niiden lisäksi on olemassa joukko videoiden ääni- sekä teksti-informaation analysointiin kehitettyjä menetelmiä.

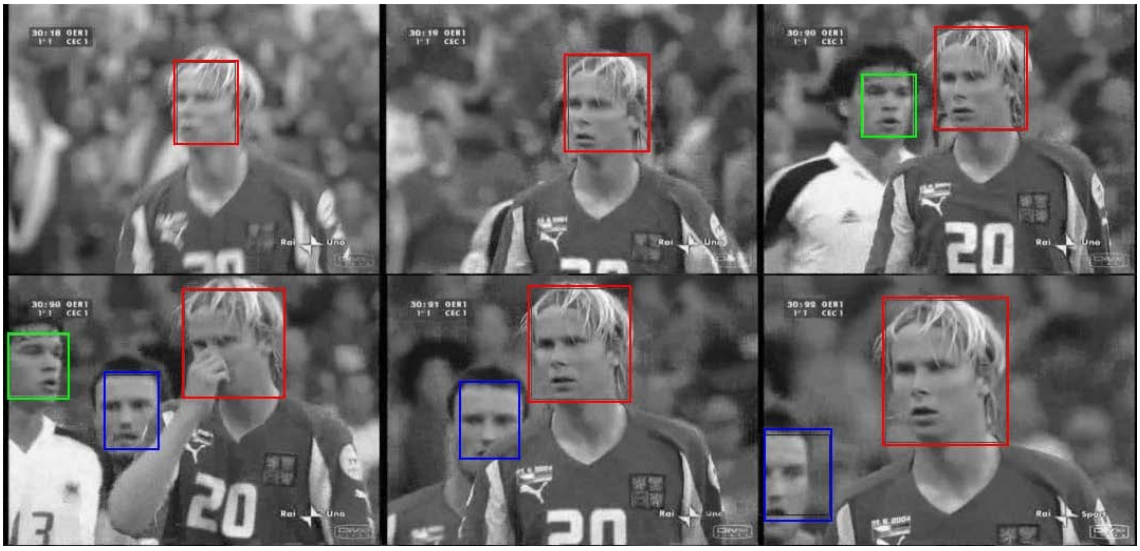
4.1 Visuaalinen analysointi

Videokuvan visuaalinen analysointi perustuu jossakin määrin samoihin menetelmiin kuin valokuvien visuaalisen tiedon analysointi. Videossa on kuitenkin läsnä lisäksi liike itse kuvassa sekä kameran liikkeet (Brunelli ym. 1999). Visuaalinen analysointi liittyy ensinnäkin video-otoksen alku- ja loppukohtien tunnistamiseen, johon on tekniikoita kuvattu useiden tutkijoiden toimesta (mm. Deb & Zhang 2005, Brunelli ym. 1999, Snoek & Worring 2005). Lisäksi visuaalisilla menetelmillä analysoidaan tunnistettujen otosten sisältöjä, jotta ylemmän tason tietoa ja tapahtumia voidaan muodostaa (Brunelli ym. 1999).

Video-otoksen tunnistamisessa käytetyt menetelmät voidaan jakaa Brunellin ym. (1999) mukaan erikseen pakkaamattomalle ja pakatulle videodatalle. Pakkaamattoman video-otoksen tunnistamisen laskenta-algoritmejä on kehitetty jo 1990-luvun alkupuolelta lähtien. Näihin kuuluvat muun muassa kuvapikselin tarkkuudella tapahtuva vertailu sekä kuvan osa-alueisiini perustuva samankaltaisuuden vertailu. Useimmin käytetään kuitenkin kuvan intensiteettiin tai väriin liittyvän histogrammin analysointia peräkkäisistä kuvista, koska histogrammit tarjoavat paremman kokonaistuloksen kuvasta verrattuna virhealttiimpiin pikselin tarkkuudella tehtäviin analyyseihin. Muita käytettyjä pakkaamattoman videon menetelmiä ovat Yakimovskyn samankaltaisuussuhteen testi, Kolmogorovin-Smirnovin testi sekä yksittäisten ei-päällekkäisten alueiden analysointi valoisuusasteen perusteella peräkkäisissä kuvissa.

Pakatun videodatan visuaalisessa analysoinnissa käytettyjä menetelmiä ovat DCT-kertoimen laskentaan perustuvat menetelmät, liikevektorimenetelmät sekä näiden kahden menetelmän yhdistelmät (Brunelli ym. 1999). DCT:hen perustuvilla menetelmillä hyödynnetään suoraan MPEG-videoiden koodauksessa käytettäviä algoritmejä, jolloin sisällöntunnistus on mahdollista jo ennen varsinaista videon purkamista ja läpikäyntiä. Esimerkiksi Arman, Hsu ja Chiu (1993) käyttivät jo vuonna 1993 menetelmää, joka perustui M-JPEG -videoiden peräkkäisten kuvien DCT-kertoimien korrelaatioihin. Hieman uudempaa menetelmää käyttivät Chum, Philbin, Isard ja Zisserman (2007) pakattujen MPEG-1 -tiedostojen otoksen tunnistamisessa. Heidän menetelmänsä perustui ”lähes identtisten kuvien” tunnistamiseen hierarkkisten värihistogrammien ja paikallisten ominaisuuskuvaajien hyödyntämiseen SIFT (Scale-invariant feature transform) -menetelmän avulla.

Kohtauksiin ja otoksiin jaetuista videosisällöistä halutaan yleensä seuraavaksi tunnistaa esimerkiksi ihmiset tai esineet ja niiden liikkeet. Snoekin & Worringin (2005) mukaan ihmishahmojen tunnistaminen yksinkertaistetaan useimmiten kohdistamalla toimenpiteet ainoastaan kasvojen tunnistukseen. Jo pelkällä kasvojen tunnistuksella (kuva 1) voidaan videosisältö jakaa osiin esimerkiksi uutisvideoissa ja toisaalta luokitella sisältöä lähi- ja yleiskuviin kasvojen koon perusteella (Brunelli ym. 1999). Kasvojentunnistuksen tekniikoita on esitelty Brunellin ym. (1999) mukaan kirjallisuudessa useita. Näitä ovat muun muassa oleellisen komponentin analyysit (principal component analysis), todennäköisyys-tiheys analyysit (probability density estimation), tieto-teoreettinen lähestymistapa (information-theoretic approach), MPEG-videoiden tunnistukseen kehitetty menetelmä hyödyntäen DC- ja DCT-kertoimia ja neuroverkkotekniikoihin perustuvat menetelmät. Snoekin ja Worringin (2005) mukaan näistä tekniikoista on osoitettu neuroverkkotekniikoihin perustuvien menetelmien tekevän vähiten virheitä.



Kuva 1. Esimerkki jalkapallo-ottelun valituista avainkuvista ja tunnistetuista kasvonalueista (Bertini, Del Bimbo ja Nunziati 2006).

Joissakin tapauksissa kasvojen ja pään tunnistusta voidaan laajentaa kaikkien kehon osien tunnistukseen, joka voisi erityisesti urheiluvideoiden kohdalla olla erittäin kiinnostavaa. Tällaisesta menetelmästä raportoivat Mohan, Papageorgiou ja Poggia (2001), joiden menetelmä perustui neljään erilaiseen tunnistimeen, jotka opetetaan havaitsemaan erikseen pää, jalat, vasen käsi ja oikea käsi. Seuraavaksi esimerkki-pohjainen luokittelija yhdistää havaitut komponentit ja päättää onko kyseessä henkilö vai ei. Tällaisen erillisiin komponentteihin perustuvan menetelmän on todettu antavan parempia tuloksia kuin koko ihmiskehon havaitseminen yhtenä kokonaisuutena.

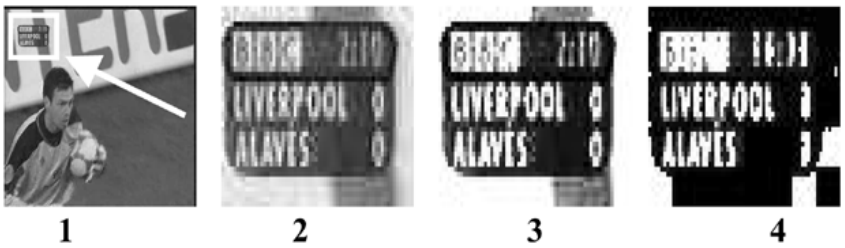
Erilaisilla liikkeen havaitsemismenetelmillä voidaan erottaa kuvassa olevien kiinnostavien objektien liikkeet sekä kameran liikkeet, kuten kuvan zoomaus ja kameran kääntäminen (Brunelli ym. 1999). Useimmiten liikkeen havaitsemisessa käytetään hyväksi videon aikaan ja paikkaan sidottuja ominaisuuksia niin sanotuilla spatio-temporaalisella analyysillä. Siinä videosisältö pyritään jakamaan erilaisiin kerroksiin liiketiedon mukaisesti, jolloin eri kerroksilta on havaittavissa kiinnostuksen kohteena olevat objektit. Debin & Zhangin (2005) mukaan yksi mahdollisuus liikkuvien objektien havaitsemiseen on niin sanottu optisen virtauksen laskeminen, joka perustuu kuvan valoisuusmuotojen liikenopeuksien jakauman analysointiin. Brunellin ym. (1999)

mukaan videot sisältävät jopa kuutta erilaista kameran liikettä, jotka ovat kameran kääntäminen sivu- tai pystysuunnassa, zoomaus, kohteen seuraaminen sekä kamerajalustan liikuttaminen sivu- tai pystysuunnassa. Kameraliikkeen tunnistamisessa käytetään yleensä hyväksi kuvan liikevektoreiden analysointia.

4.2 Teksti-informaation analysointi

Videoilla esitettävän tekstin analysointi antaa huomattavaa lisäarvoa pelkkään visuaaliseen analysointiin verrattuna. Videoilla esitettävät tekstit voivat Brunellin ym. (1999) mukaan yksinkertaisesti tukea tai toistaa puhuttua informaatiota, kertoa asian ajallisesta tiedosta tai antaa kokonaan uutta informaatiota. Teksti-informaation tunnistuksen menetelmät perustuvat muun muassa pikselitasolla tapahtuvaan kuvan kirkkauden analysointiin ja vertailuun sekä tekstitiedon oletetun paikan ja muodon perusteella tapahtuvaan tunnistukseen. Urheiluvideoissa tekstien avulla kerrotaan useimmiten pelin tärkeimmistä tapahtumista, kuten piste-/maalitilanne, pelaajan nimi ja numero ja pelitilastot esimerkiksi pelin ensimmäiseltä puoliajalta.

Ensimmäiseksi videolla oleva teksti tulee tunnistaa, jonka jälkeen se voidaan analysoida esimerkiksi tulkitsemalla sisältö OCR-menetelmällä (optical character recognition). Kuvassa 2 on esitelty Sadlierin & O'Connorin (2005) käyttämä malli jalkapallotekstien tunnistamiseen videokuvasta. Muutamat käytetyistä tekniikoista keskittyvät sen sijaan ainoastaan tunnistamaan tekstitystapahtumien ajankohdat videon otosten ja haluttujen osien tunnistamiseksi.



Kuva 2. Tulostaulun tekstien tunnistaminen jalkapallossa. 1: tulostaulun sijainti kuvassa, 2: suurennettu tulostaulun teksti, 3: sama teksti kontrastiparannuksen jälkeen, 4: analysoitava tekstisisältö lopullisessa muodossaan. (Sadlier & O'Connor 2005).

Tehokkaan teksti-informaation analysointimenetelmän esittelivät Li, Doerman ja Kia (2000), joiden menetelmä kykeni tekstien tunnistamiseen ja seurantaan digitaalisista videoista. Järjestelmä perustui keinoäly-pohjaiseen luokittelijaan sekä SSD-pohjaiseen (sum of squared difference) ja ääriviiva-pohjaiseen kahteen erilliseen moduuliin, joiden avulla myös liikkuvia tekstielementtejä (esimerkiksi rullaavat elokuvan lopputeksti) pystyttiin luotettavasti analysoimaan.

4.3 Ääni-informaation analysointi

Monissa tapauksissa myös ääni-informaation analysointi on tärkeä osa videon sisällöntunnistusta. Yksinkertaisimmillaan äänen analysointi on Snoekin ja Worringin (2005) mukaan äänivirran hiljaisten hetkien analysointia, joka voidaan tehdä esimerkiksi äänen keskimääräistä energiaa tutkimalla tai hyödyntämällä tämän lisäksi vielä ZCR-menetelmä (zero crossing rate), jonka mukaan äänen nollarajan ylittäminen tapahtuu kun peräkkäiset näytteet ovat erimerkkiset. Lisäksi esimerkiksi musiikin erottaminen muusta äänivirrasta voidaan tehdä hyödyntämällä äänen harmonisuusastetta, taajuuden keskittymistä tietyille alueille, ZCR:n vaihtelua sekä ZCR:n värähtelytaajuuden laajuutta.

Tarkempaan äänisisältöjen analysointiin kehittivät menetelmän muun muassa Li, Sethi, Dimitrova ja McGee (2001), jotka raportoivat jatkuvan ääni-informaation analysoimisesta. Heidän mallinsa kykeni tunnistamaan jatkuvasta äänivirrasta seitsemän erilaista kategoriaa, jotka olivat jatkuva hiljaisuus, yksi puhuja äänessä, musiikki, ympäristön melu, monta puhujaa äänessä, samanaikainen puhe ja musiikki sekä samanaikainen puhe ja taustamelu. Järjestelmä kykeni luokittelemaan äänen vaiheita 90 prosentin tarkkuudella ja olisi varsin hyvin sovellettavissa ominaisuuksiltaan myös digitaalisen videon ääniraidan analysoinnissa.

4.4 Multimodaalinen lähestymistapa

Snoekin & Worringin (2005) mukaan tehokas videoiden sisällöntunnistus ja luokittelu vaatii multimodaalista lähestymistapaa, jolloin sisällön analysointi tapahtuu vähintään kahden eri informaatiokanavan kautta. Tämä tarkoittaa käytännössä esimerkiksi visuaalisen ja teksti-informaation sisällöntunnistusmenetelmillä saatujen tietojen

yhdistämistä. Tämä on kuitenkin teknisesti selvästi haastavampaa kuin ainoastaan yhden informaatiokanavan analysointi.

Multimodaalinen ajattelutapa videosisältöön on käytännössä lähtöisin jo videon kuvaus- ja editointityöstä, joiden kautta ohjaaja luo videolle tavoitteensa mukaisen sisällön hyödyntäen visuaalisia, äänellisiä sekä tekstiin pohjautuvia tekijöitä. Täten sitä voidaan pitää hyvin luonnollisena lähestymistapana myös videoiden sisällöntunnistukselle. Usein tärkein osa erilaisia videosisältöjä on sen visuaalinen ilmentymä, jolle merkitys luodaan kuviin liittyvän ajan ja paikan, kohteiden sekä ihmisten kautta. Ihmisten puhe on luonnollinen osa sisältöä. Samoin musiikilla ja muilla äänillä luodaan tunnelmaa sekä muun muassa helpotetaan siirtymiä otoksesta toiseen. Teksti-informaatiolla sen sijaan välitetään joko kokonaan uutta informaatiota tai toistetaan jo kuvan tai äänen kautta välitettyä sisältöä. (Snoek & Worring 2005)

5 SISÄLLÖNTUNNISTUKSEN SOVELTAMINEN URHEILUVIDEOIHIN

Erilaisia sisällöntunnistuksen ja luokittelun menetelmiä on käytetty hyvin monenlaisissa videosisällöissä, joista suosituimpia ovat erilaiset uutis- ja urheiluvideot. Käytännössä urheiluvideoissa voidaan käyttää kaikkia yleisimpiä sisällöntunnistuksen menetelmiä, huomioiden kuitenkin urheiluvideoihin liittyvät erityispiirteet. Urheiluvideoiden sisällöntunnistuksen ja luokittelun menetelmiä on kehitetty jo suhteellisen useisiin lajeihin, joista joukkueurheilut näyttävät olevan selvässä suosiossa.

5.1 Urheiluvideoiden erityispiirteet

Yleisiä sisällöntunnistuksen ja luokittelun menetelmiä voidaan hyödyntää myös urheiluvideoissa muutamin reunaehdoin. Yun ja Farinin (2005) mukaan urheiluvideoiden sisällöntunnistukseen haasteen tuovat lajispesifiset säännöt ja toimintaympäristöt. Suoraan kaikkiin urheilulajeihin sovellettavaa sisällöntunnistuksen menetelmää on lähes mahdotonta luoda juuri eri lajien monimuotoisuuden vuoksi (Snoek & Worring 2005). Muutamissa tutkimuksissa (kuten Tjondronegoro, Chen & Joly 2008) on raportoitu menetelmiä, joilla pystytään samoin menetelmin useamman eri lajin sisällöntunnistukseen, mutta nämä menetelmät ovat vielä hyvin harvinaisia.

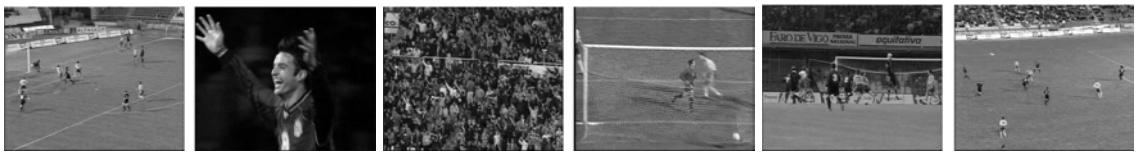
Urheiluvideoissa on multimodaalisesta lähestymistavasta niin ikään suuri hyöty. Usein esimerkiksi pelin tai kilpailun selostajan puheessa on paljon hyödynnettävää informaatiota sisällön luokittelemiseksi. Samoin tekstiä käytetään hyvin paljon muun muassa sen tärkeimmän asian eli piste- tai maalitilanteen esittämiseksi. Yun ja Farinin (2005) mukaan myös urheiluvideoissa tulee vastaan haaste alemman tason yksinkertaisten videosisältöä kuvaavien muuttujien ja semanttisempien kohteiden ja tapahtumien yhdistämisessä. Lajiin oleellisesti liittyvät tapahtumat, kuten esimerkiksi lyönnit, heitot, tietyt pelaajan liikkeet, vapaapotkut ja maalit ovat niitä toivottuja ja merkityksellisiä sisältöjä videon käyttäjän kannalta, kun esimerkiksi halutaan hakea tietty pelitapahtuma jostakin videotietokannasta.

Yun ja Farinin (2005) mukaan urheiluvideoiden sisällöntunnistuksesta ja luokittelusta voidaan saada monenlaista hyötyä. Käyttäjille voidaan tarjota automaattisia

videokoosteita esimerkiksi pelin kohokohdista tai vaihtoehtoisesti vapaammilla videosisältöön kohdistuvilla hakutoimenpiteillä toteutettuja videopalveluita. Toinen käyttökohde on lajin taktiset ja suoritukselliset analyysit, joilla voidaan saada tietoa esimerkiksi yksittäisen pelaajan tai koko joukkueen toiminnasta pelistrategian kannalta. Samoin monipuolisia tilastoja pelin kulusta on mahdollista muodostaa automaattisten analyysien perusteella. Urheiluvideoista on lisäksi mahdollista nykyään muodostaa eräänlaisia virtuaalisia tapahtuman uudelleen elämisen mahdollisuuksia muun muassa 3D-tekniikoita hyödyntäen (esimerkiksi Koyama, Kitahara & Ohta 2003). Lisäksi jopa reaaliaikaisiin TV-lähetyksiin on mahdollista lisätä sisällöntunnistuksen kautta tuotettua lisäinformaation pelistä tai suorituksesta. Urheilukoosteiden ja mukautettujen videon pakkausmenetelmien hyöty tulee esille myös koko ajan yleisimmiksi käyvissä mobiililaitteissa, joilla videoiden katsominen nykytekniikalla on vielä hieman rajoitettua ja kallista.

5.2 Joukkuelajit

Joukkuepeleistä yksittäisenä lajina on jalkapallossa tehty selvästi aktiivisimmin videon sisällöntunnistukseen ja luokitteluun liittyvää tutkimusta. Tämä selittynee pitkälti lajin suurella levinneisyydellä ja suosiolla ympäri maailmaa. Jalkapallovideoiden sisällöntunnistuksessa on useissa tapauksissa analysointi tehty multimodaalisesti sekä kuvaan että ääneen perustuvin menetelmin. Ekin, Tekalp ja Mehrotra (2003) yhdistivät jalkapallovideoiden analysoinnissa sekä alemman tason että ylemmän tason tunnistukseen liittyviä algoritmeja. Menetelmillään he kykenivät muodostamaan jalkapallovideoista haluttuja koosteita sisältäen 1. pelin kaikki hidastuskohtaukset, 2. pelin kaikki maalit ja 3. kohteen mukaan valikoidut hidastukset pelissä. Alemman tason menetelmät sisälsivät vallitsevan värialueen analysoinnin, otoksen tunnistamisen sekä otosten luokittelun. Ylemmän tason laskentamenetelmiä käytettiin maalien, tuomarin ja rangaistusalueen tunnistamiseksi. Esimerkiksi hyväksyttyyn maaliin liittyvät aina tietyssä järjestyksessä esitetyt tietyntyyppiset otokset (kuva 3). Mainittakoon myös, että pelin hidastuskohtausten tunnistaminen onnistui järjestelmältä keskimäärin noin 85 prosentin tarkkuudella.



a) b) c) d) e) f)

Kuva 3. TV-lähetyksessä näytetyt otokset maalin jälkeen: a) yleisnäkymä itse maalinteosta, b) lähikuva maalintekijästä, c) yleisö, d) ensimmäinen uusinta, e) toinen uusinta sekä f) yleisnäkymä pelin keskialoituksesta (Ekin, Tekalp & Mehrotra 2003).

Tekstintunnistusta osana jalkapallovideoiden sisällöntunnistusta hyödynsivät Bertini ym. (2006), Xu, Wang, Wan, Li ja Duan (2006) sekä Huayong (2004). Tekstintunnistusta videokuvasta käytettiin pääasiallisena menetelmänä pelaajan tunnistamiseksi Bertinin ym. (2006) tutkimuksessa. Jos tunnistaminen ei tunnistetun pelipaidan numeron tai muun videolla näkyvän tekstin avulla onnistunut, otettiin avuksi myös kasvontunnistus, joka perustui aikaisemmin mainittuun SIFT-tekniikkaan. Xun ym. (2006) menetelmä jalkapallovideoiden tapahtumien tunnistamisessa perustui varsin mielenkiintoiseen ja hieman erilaiseen lähestymistapaan. He yhdistivät reaaliaikaisesti sekä videoanalyysin että web-casting tekstien analysoinnin. Videoanalyysin kautta tunnistettiin pelin alkamisaika sekä juokseva peliaika, kun puolestaan internetin kautta lähetetyn web-cast –tekstin kautta saatiin analysoitua pelin tapahtumat tarkasti hetki hetkeltä.

Reaaliaikaisesti analysoi jalkapallo-otteluita myös Huayong (2004) hyödyntäen visuaalista kuvan analysointia, selostajan puheen tunnistusta, puheen signaalinkäsittelyä sekä tekstintunnistusta. Pelkkä visuaalinen analysointi tuotti noin 82 prosentin tarkkuuden, kun sen sijaan äänen analysointia myös hyödyntäen päästiin noin 91 prosentin tarkkuuteen. Jalkapallovideoiden kohokohtien tunnistaminen voidaan tehdä myös ajallisen päättelyn menetelmillä (temporal logic), kuten Assfalg, Bertini, Colombo, Del Bimbo ja Nunziati (2003) raportoivat. Heidän ratkaisunsa haki pelin ratkaisuhetket kulloisenkin pelikentän osan ja kameraliikkeiden analysoinnin avulla.

Koripallovideoiden yhdeksän erilaisen tapahtuman tunnistamiseen pystyvän järjestelmän esittivät Zhou, Vellaikai ja Kuo (2000). Heidän järjestelmänsä perustui

videon liikkeiden, väri-informaation sekä kuvavaihtojen analysointiin ja siten saatujen alemman tason muuttujien sääntöpohjaiseen luokittelijaan. Osana luokittelijan muodostusta hyödynnettiin myös päätöspuu-menetelmää ”jos-sitten” sääntöjen muodostamiseksi. Menetelmä sopii erityisesti reaaliaikaisesti katselu- tai lähetystilanteessa käsiteltävien videoiden analysointiin.

Baseballiin liittyviä sisällöntunnistuksen ja luokittelun tutkimuksia on myös tehty suhteellisen aktiivisesti. Esimerkiksi Fleischman & Roy (2007) esittelivät menetelmänsä videon visuaalisen sisällön, tekstintunnistuksen ja äänen perusteella tapahtuvaan automaattiseen sisällöntunnistukseen. Ensin visuaalisen sisällön, kameraliikkeiden ja ääni-informaation avulla video jaettiin osiin. Sen jälkeen tiedonlouhinnan keinoin muodostettiin ajallisten tapahtumien malli, joka lopulta lingvistisellä yhteensovituksella yhdistettiin tekstityksestä poimittujen sanojen kanssa. Näin muodostettu kieliperusteinen malli (grounded language model) testattiin noin 275 tunnin videomateriaalilla baseballista. Myös Akiri, Kumano ja Tsukada (2003) raportoivat automaattisesta baseball-pelin kohokohtien tunnistamisesta reaaliaikaisesti. Heidän menetelmänsä perustui sekä visuaaliseen pelitilanteiden tunnistukseen (syöttöhetkien tunnistaminen) että peliselostuksen analysointiin erillisen baseball-sanaston avulla. Sekä kuvan- että äänentunnistuksen tulokset vietiin metadatana tietokantapalvelimelle XML-muodossa, jonka kanssa yhdistettiin toisella palvelimella sijaitseva varsinainen videosisältö, muodostaen lopulta sisältöhakuja mahdollistavan www-sovelluksen käyttäjilleen.

Samanaikaisesti useamman eri joukkuelajin videosisällön analysointiin sopivia menetelmiä on raportoitu kirjallisuudessa muutamia (muun muassa Tjondronegoro ym. 2006, Sadlier & O’Connor 2005 sekä Duan, Xu, Chua, Tian & Xu 2003). Jalkapallon, koripallon ja australialaisen jalkapallon kohokohtien tunnistamismenetelmän esittelivät Tjondronegoro ym. (2006). Tavoitteena oli käyttää mahdollisimman vähän lajispesifisiä muuttujia ja manuaalista työtä, jotta menetelmä olisi sovellettavissa helpommin usean eri lajin kohokohtien tunnistamiseen. Pelitilanteiden tunnistaminen tehtiin useaan lajiin yhteneväisesti etsien esimerkiksi maalin/korin tekoon liittyviä tekijöitä, kuten maalin kuvaa, tulosinformaatiota numeroita, lähikuvaa ja uusintahidastuksia. Jonkin verran muodostettuja algoritmeja ja tilastoja jouduttiin kuitenkin muokkaamaan manuaalisesti

ennen seuraavaan lajiin siirtymistä. Hieman samantyylisellä lähestymistavalla Sadlier ja O'Connor (2005) kehittivät järjestelmän useiden lajien oleellisten tapahtumien tunnistamiseen audio-visuaalisen informaation sekä SVM:n (Support Vector Machine) avulla. He testasivat järjestelmän toimivuuden jalkapallossa, rugbyssä, maahockeyssa ja gaelilaisessa jalkapallossa.

Alemman ja ylemmän tason sisältömuuttujien väliin sijoittuvia, niin sanottuja keskitason muuttujia hyödynsivät tutkimuksessa Duan ym. (2003) luodessaan menetelmäkehystä urheiluvideoiden semanttisen sisällön esittämiseksi. Heidän mukaansa näillä keskitason muuttujilla menetelmä saadaan tehokkaammaksi ja paremmin erilaisiin lajeihin soveltuvaksi. Alemman tason muuttujien tunnistamisessa he hyödynsivät keskiarvon siirtymän menetelmää (Mean Shift Procedure) sekä väri- että liikevektorien analysoinnin tukena. Järjestelmä testattiin viidellä eri palloilulajilla, mukaan lukien myös tennis, joka luonnollisesti poikkeaa jonkin verran palloilun joukkuelajeista.

5.3 Yksilölajit

Urheiluvideoissa yksilölajien sisällön- ja tapahtumien tunnistus on keskittynyt pääasiassa edellä mainittuun tennikseen, josta on saatavilla muutamia tutkimuksia. Muiden yksilölajien kohdalla tutkimukset ovat vain yksittäistapauksia. Tjondronegoro, Chen ja Joly (2008) raportoivat suhteellisen laajasta tutkimuksestaan urheiluvideoiden sisällöntunnistuksessa ja tiedon haussa soveltuen jalkapalloon, tennikseen, uintiin ja uimahyppyihin. Järjestelmänsä kehittelyn helpottamiseksi he luokittelivat urheilulajit neljään eri kategoriaan sekä ajallisten muuttujien että tapahtumien perusteella: 1. Vakioidun ajallisen pelijakson lajit (jalkapallo, käsipallo, koripallo jne.), 2. Pisteiden perusteella erän pituuden määrittelevät lajit (tennis, sulkapallo, lentopallo jne.) 3. Suoritusaikaan perustuvat lajit (juoksu, uinti, pyöräily, triathlon jne.) sekä 4. Suoritustekniikkakeskeiset lajit (voimistelu, yleisurheilun hyppy- ja heittolajit), painonnosto jne.). Videoiden sisällöntunnistus ja luokittelu perustui puoli-skeema –pohjaiseen (semi-schema based) luokittelumenetelmään yhdistettynä OR-lähestymistapaan (Object-Realtionship). Järjestelmä mahdollistaa asteittaisen sisältöindeksin rakentamisen hyödyntäen älykkäästi sisällöntunnistuksen algoritmeja.

Lisäksi dynaamiset ja käyttäjälähtöiset haut indeksoituun videosisältöön toteutettiin XQuery 1.0 –kyselykielellä. Tutkijat totesivat myös, että luotu indeksointimalli on helposti sovellettavissa MPEG-7 –videoformaatin standardin mukaiseksi.

Tenniksen sisällöntunnistuksesta raportoivat Sudhir, Lee ja Jain jo vuonna 1998. He perustivat järjestelmänsä mallille tenniskentän rajaviivoista sekä kiinteän kameran kuvakulman geometriasta. Erikseen kehitettiin algoritmit tenniskentän rajojen tunnistamiseen ja pelaajan liikkeiden seurantaan. Lisäksi värialueiden analysoinnilla tunnistettiin pelikenttäkohtaukset raa'asta videomateriaalista, joka sisälsi käytännössä paljon muutakin kuvamateriaalia kuin pelkästään pelitilanteita. Yhdistämällä kentän rajojen tunnistaminen ja pelaajan liikkeet päästiin muodostamaan korkeamman tason muuttujia liittyen oleellisiin pelitapahtumiin, joita olivat muun muassa aloitussyöttö, pallorallit, ohituslyönti ja verkkopeli.

Multimodaalisen, videon kuvaa ja ääntä hyödyntävän, lähestymistavan mailapelien tapahtumien luokitteluun ja kohokohtien rekisteröimiseen esittelivät myös Zhao, Zhou ja Tang (2006). Vallitsevan värialueen ja liikemallien (kuva 4) analysoinnin avulla video-otokset jaettiin ensin yleisen kuvakulman otoksiin ja ei-yleisen kuvakulman otoksiin (lähikuva ja puolilähikuva). Tämän jälkeen etsittiin äänellisiä muuttujia, kuten mailan osuma palloon ja yleisön aplodit, joiden avulla tutkittiin tapahtumien merkityksellisyyttä. Näin voitiin erottaa esimerkiksi syöttövirheet, läpisyötöt, pallorallit ja uusinnat. Lisäksi lopuksi luokiteltiin tapahtumat vielä järjestykseen niiden intensiivisyytensä mukaan. Järjestelmää arvioitiin aineistolla, joka sisälsi videotiedostoja tenniksestä, pöytätenniksestä ja sulkapallosta. Tenniksessä tapahtumien tunnistuksen tarkkuudessa päästiin yli 90 prosenttiin.

Kuva 4. Liikkeen tunnistuksen kohteet tenniksen kolmessa tyypillisimmässä kuvassa: 1. yleiskuva, 2. lähikuva ja 3. puolikuva (Zhao, Zhou & Tang 2006).

6 POHDINTA

Urheiluvideoiden sisällöntunnistukseen ja luokitteluun liittyvä tutkimusala on varsin laaja ja kirjava jo pelkästään erilaisten käytettyjen menetelmiensä osalta. Selkeän yleiskuvan saamista tältä alueelta vaikeuttaa käsitteiden epäyhdenmukaisuus ja erilaiset lähestymistavat. Samoistakin asioista esimerkiksi video-otosten tunnistamisessa saatetaan puhua varsin erilaisin termein ja käsittein. Tunnetut tilastolliset menetelmät ja algoritmit ovat kuitenkin vakiintuneita ja helpommin tunnistettavissa.

Yhdenmukaisia menetelmiä tai toimintamalleja videosisältöjen tunnistamiseksi on luotu varsin vähän. Eri tutkijat ja tutkimusryhmät käyttävät lisäksi kukin tilanteen mukaan hieman erilaisia yhdistelmiä multimodaalisesta sisällöntunnistuksesta. Monimuotoisuutta alueeseen tuo myös urheilulajien erilaisuus lajin sääntöjen, urheilijoiden määrän, suoritusalueen ja välineiden suhteen. Onneksi muutamat tutkimukset ovat jo pyrkineet yleisempään lähestymistapaan, jota voitaisiin soveltaa sellaisenaan tai pienin muutoksin useammassakin eri lajissa.

Teknisellä tasolla alemman tason muuttujien ja ylemmän tason merkityksellisten kohteiden ja tapahtumien tunnistamisen välillä vallitsee edelleen havaittavissa oleva semanttinen ”kuilu”. Tähän on kuitenkin useissa tutkimuksissa kiinnitetty huomiota ja ratkaisumalleja siihen on esitetty. Siitä huolimatta asia nousee esille edelleen uusimmissakin tämän alan tutkimusartikkeleissa.

Automaattinen sisällöntunnistus ei jatkuvasta kehityksestään huolimatta ole vielä kuitenkaan 100 prosentilla tarkkuudella toimivaa tekniikkaa. Se tuo hyvin toimiessaan edullisen ja nopean tavan sisällöntunnistukseen ja luokitteluun, mutta tarkkuudessa ja toistettavuudessa on edelleen parannettavaa. Manuaaliset menetelmät ovat vielä toistaiseksi ainoita keinoja, jotka pystyvät aukottomaan sisällöntunnistuksen ja luokitteluun.

Urheiluvideoiden sisällöntunnistuksen tutkimukset ja kehitetyt menetelmät ovat keskittyneet pääasiassa palloilulajeihin ja niissäkin erityisesti joukkuelajeihin. Joukkuelajeissa on kuluttajiltakin tuleva kysyntä varmasti ollut suurempaa erilaisten videopalveluiden suhteen ja lisäksi näihin lajeihin liittyy usein suuremmat mainosarvot

ja yleisöt. Toisaalta, esimerkiksi yleisurheilukilpailuissa voisi hyödyllistä lisäarvoa tuoda jo pelkästään sisällön pilkkominen video-otoksen tunnistustekniikoilla lajikohtaisesti. Perinteisesti yleisurheilukilpailuiden videot ovat nimittäin hyvin ”levottomia” kuvan vaihtuessa jatkuvasti lajista toiseen. Jos tähän pystyttäisiin vielä yhdistämään kuvassa useimmiten välitettävä teksti-informaatio suoritusten tuloksista, niin lopputuloksena olisi varmasti hyvinkin mielekäs ja käyttökelpoinen sisällöntunnistuksen ja luokittelun sovellustapa.

Videoiden sisällöntunnistukseen ja luokitteluun sopivia valmiita järjestelmiä on suhteellisen vähän tarjolla tai niiden löytäminen on yksinkertaisesti hankalaa. Tampereen teknillisen yliopiston kuva- ja videoanalysoinnin ryhmän jäsenet ovat vuosien työn tuloksena kehittäneet MUVIS-nimisen järjestelmän (http://muvis.cs.tut.fi) kuva- ja videosisältöjen semanttiseen luokitteluun, tietokantapohjaiseen tallentamiseen ja monipuolisiin sisältöhakuihin. Ohjelmisto on kenen tahansa ladattavissa ja asennettavissa omaan käyttöön. Kehitystyön taustalla on useita tutkimusjulkaisuja ja Serkan Kiranyazin väitoskirjatyö. Toinen esimerkki valmiista järjestelmästä on yhdysvaltalaisen Virage Technologyn kehittämä kaupallinen Virage VideoLogger (http://www.virage.com/home), joka on käytännössä täysin automaattinen ja reaaliaikainen erilaisten videosisältöjen indeksointijärjestelmä. Lisäosilla sen toimintaa voidaan laajentaa kattamaan myös tekstintunnistus ja äänianalyysit. Jalkapallossa paljon käytetty ja tunnettu sisällöntunnistukseen liittyvä järjestelmä on ProZone, jota käytetään hyvin aktiivisesti muun muassa Englannin Valioliigassa. Järjestelmän periaatteena on animoitu pelin uudelleenmuodostaminen ja analysointi. Toisin sanoen, jokaisen pelaajan ja pallon liikkeet analysoidaan tallennettavasta videosta, jonka perusteella muodostetaan pelianimaatio, jota on mahdollista tarkastella hyvin käyttäjälähtöisellä tavalla. Samassa yhteydessä on mahdollista myös monipuoliset suoritus- ja taktiikkamuuttujien laskennat niin yksilö- kuin joukkueen tasolla tarkastellen.

7 YHTEENVETO

Tässä tutkielmassa tarkasteltiin videoiden sisällöntunnistuksen ja luokittelun erilaisia menetelmiä. Lisäksi selvitettiin kyseisten menetelmien käyttöä urheiluvideoiden analysoinnissa ja luokittelussa. Kirjallisuudessa on raportoitu useita erilaisia menetelmiä niin pakatun kuin pakkaamattoman videon sisällöntunnistukseen. Saatujen tulosten mukaan yleisimpänä menetelmänä videosisältö jaetaan ensin video-otoksiin perustuen valikoituihin sisällöntunnistuksen tekniikoihin, hyödyntäen visuaalista sekä teksti- ja ääni-informaation analysointia. Tämän jälkeen joko valitaan parhaiten otosta edustava avainkuva tai aloitetaan suoraan niin sanotun alemman tason sisällön tunnistaminen, johon on niin ikään olemassa useita erilaisia menetelmiä. Paras ratkaisu olisi monessa tapauksessa suoraan merkityksellisempien ylemmän tason muuttujien tapahtumien tunnistaminen, mutta yleensä se on hyvin vaikeaa. Siten siihen päästään vasta alemman tason muuttujien yhdistämisellä toisiinsa suhteellisen monimutkaisin menetelmin.

Useat sisällöntunnistuksen tutkimukset hyödyntävät nykyään jo varsin edistyksellisesti monen eri informaatiokanavan kautta saatavaa tietoa multimodaalisesti. Näin esimerkiksi selostajan puheen tunnistuksella tai kuvassa näytettävien tekstien analysoinnilla saadaan lisätietoa visuaalisen sisällöntunnistuksen tueksi. Menetelmät voivat olla hyvinkin innovatiivisia yhdistäen esimerkiksi suoran TV-lähetyksen videokuvaa ja web-castingin kautta saatavaa teksti-informaatio haluttujen tapahtumien tunnistamisessa.

Urheiluvideoiden kohdalla käytetään hyvin pitkälti täysin samoja sisällöntunnistuksen menetelmiä kuin muissakin videosisällöissä. Videoiden monimuotoisuus, suuri informaatiomäärä ja eri lajien erilaiset säännöt ja toimintaympäristöt luovat kuitenkin suuria haasteita useisiin lajeihin sopivien yhdenmukaisten menetelmien kehittämiseksi. Tähän on kuitenkin lupaavia yrityksiä jo tehty ja useampaankin erilaiseen lajiin sopivia menetelmiä on esitelty. Yleisimmin urheiluvideoihin kehitetyt sisällöntunnistuksen menetelmät ovat koskettaneet pallopelejä ja joukkuelajeja. Yksilölajeihin soveltuvia järjestelmiä on ainakin vielä tähän mennessä suunniteltu ja raportoitu valitettavan vähän.

Tutkielma antoi hyvän yleiskuvan videoiden sisällöntunnistuksessa ja luokittelussa käytettävistä menetelmistä sekä niiden hyödyntämisestä urheiluvideoissa. Tutkimusalue on kuitenkin erittäin laaja ja tämän tutkielman rajoissa sen tarkastelu jäi kuitenkin suhteellisen pintapuoliseksi. Esimerkiksi erilaisia matemaattisia menetelmiä sisällöntunnistukseen on kehitetty niin paljon, että yksinään jo pelkän visuaalisen informaation analysoinnista olisi myös pystynyt tutkielman kirjoittamaan. Tutkielman kohdealueen yleiskartoituksena työ kuitenkin toimi hyvin ja antaa selkeä perustan mahdollisille jatkotutkimuksille esimerkiksi joltakin kapeammalta videoiden sisällöntunnistuksen osa-alueelta.

LÄHTEET

Akiri Y, Kumano M. & Tsukada K. 2003. Highlight Scene Extraction in Real Time from Baseball Live Video. Teoksessa Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, Berkeley, California. ACM, New York, NY, USA, 209-214.

Arman F., Hsu A. & Chiu M.Y. 1993. Feature Management for Large Lideo Database. Teoksessa Proceedings of SPIE Storage and Retrieval for Image and Video Databases, vol. 1908, 2-12.

Assfalg J., Bertini M, Colombo C., Del Bimbo A. & Nunziati W. 2003. Automatic Extraction and Annotation of Soccer Video Highlights. Teoksessa Proceedings of the International Conference on Image Processing, September 14-17, vol. 2, 527-30.

Berkeley Multimedia Research Center 2008, MPEG-1 Video [online]. University of California, Berkeley [viitattu 7.12.2008]. Saatavilla www-osoitteessa: http://bmrc.berkeley.edu/frame/research/mpeg/faq/mpeg1.html.

Bertini M., Del Bimbo A., Cucchiara R. & Prati A. 2004. Semantic Video Adaption Based on Automatic Annotation of Sport Videos. Teoksessa Proceedings of the ACM MIR'04, New York, NY, USA, October 15-16. ACM, New York, NY, USA, 291-298.

Bertini M., Del Bimbo A. & Nunziati W. 2006. Automatic Detection of Player's Identity in Soccer Videos using Faces and Text Cues. Teoksessa Proceeding of the ACM MM'06, Santa Barbara, California, USA, October 23-27. ACM, New York, NY, USA, 663-666.

Brunelli R., Mich O. & Modena C. 1999. A Survey on the Automatic Indexing of Video Data. Journal of Visual Communication and Image Representation, 10(2), 78-112.

Chum O., Philbin J., Isard M. & Zisserman A. 2007. Scalable Near Identical Image and Shot Detection. Teoksessa Proceedings of the 6th ACM International Conference

on Image and video Retrieval, Amsterdam, The Netherlands, July 9-11. ACM, New York, NY, USA, 549-556.

Deb S. & Zhang Y. 2005. An Overview of Video Information Retrieval Techniques. Teoksessa Video Data Management and Information Retrieval, Hershey PA, USA: IRM Press, 282-292.

Duan L.-Y., Xu M., Chua T.-S., Tian Q. & Xu C.-S. 2003. A Mid-level Representation Framework for Semantic Sports Video Analysis. Teoksessa Proceedings of the ACM MM'03, Berkeley, California, USA, November 2-8. ACM New York, NY, USA, 33-44.

Ekin A., Tekalp M. & Mehrotra R. 2003. Automatic Soccer Video Analysis and Summarization. IEEE Transactions on image processing, 12(7), 796-807.

Fleischman M. & Roy D. 2007. Unsupervised Content-Based Indexing of Sports Video. Teoksessa Proceedings of the ACM MIR'07, Ausburg, Bavaria, Sermany, September 28-29. ACM New York, NY, USA, 87-94.

Harold E.R. 2001. XML bible, 2nd ed. John Wiley & Sons, Inc., New York, NY, USA.

Huayong L. Content-Based TV Sports Video Retrieval Based on Audio-Visual Deatures and Text Information. 2004. Teoksessa Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'04). Beijing, China, September 20-24. IEEE Computer Society, Washington, DC, USA, 481-484.

ISO/IEC 1992, JTC1/SC29/WG11N11172. MPEG-1 [viitattu 1.12.2008]. Saatavilla www-osoitteessa: http://www.chiariglione.org/mpeg/achievements.htm.

ISO/IEC 1994, JTC1/SC29/WG11N13818, MPEG-2 [viitattu 3.12.2008]. Saatavilla www-osoitteessa: http://www.chiariglione.org/mpeg/achievements.htm.

ISO/IEC 1999, JTC1/SC29/WG11N14496, MPEG-4 [viitattu 1.12.2008]. Saatavilla www-osoitteessa: http://www.chiariglione.org/mpeg/achievements.htm.

ISO/IEC 2001, JTC1/SC29/WG11N15938, MPEG-7 [viitattu 4.12.2008]. Saatavilla www-osoitteessa: http://www.chiariglione.org/mpeg/achievements.htm

ISO/IEC 2002, JTC1/SC29/WG11N21000, MPEG-21 Overview v.5 [viitattu 4.12.2008].
Saatavilla www-osoitteessa: http://www.chiariglione.org/mpeg/achievements.htm.

Koyama T., Kitahara I. & Ohta Y. 2003. Live 3D video in Soccer Stadium. Teoksessa Proceeding of the International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH 2003 Sketches & Applications, San Diego, California. ACM, New York, NY, USA, 1-1.

Li H., Doerman D. & Kia O. 2000. Automatic Text Detection and Tracking in Digital Video. Teoksessa IEEE Transactions on Image Processing, 9(1). IEEE Computer Society, Washington, DC, USA, 147-156.

Li D., Sethi I.K., Dimitrova N. McGee T. 2001. Classification on General Audio Data for Content-based Retrieval. Teoksessa Pattern Recognition Letters, 22. Elsevier Science Inc., New York, NY, USA, 533-544.

MikroPC.net IT-uutiset 2008. Uudella videontunnistuksella vältetään tekijänoikeustaistelut [online]. MikroPC.net [viitattu 22.11.2008]. Saatavilla www-osoitteessa: http://mikropc.net/uutiset/index.jsp?categoryId=atk&day=20081103&ref=w20081103081936135 97#w20081103081936135 97.

Mohan A., Papageorgiou C. & Poggia T. 2001. Example-based Object Detection in Images by Components. Teoksessa IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(4). IEEE Computer Society Washington, DC, USA, 349-361.

Motion Picture Experts Group 2008. The MPEG Home Page [viitattu 2.12.2008]. http://www.chiariglione.org/mpeg/index.asp

OpenDML AVI M-JPEG File Format Subcommittee 1996, OpenDML AVI File Format Extensions: Version 1.02 [viitattu 4.12.2008]. Saatavilla www-osoitteessa: http://www.morgan-multimedia.com/download/odmlff2.pdf.

Sadlier D.A. & O'Connor N.E. 2005. Event Detection in Field Sports Video Using Audio–Visual Features and a Support Vector Machine. IEEE Transactions on Circuits and Systems for Video Technology, 15(10), October, 1225-1233.

Snoek G. & Worring M. 2005. Multimodal Video Indexing: A Review of the State-of-the-art. Multimedia Tools and Applications, 25. Springer Science + Business Media, Inc. The Netherlands, 5–35.

Sudhir G., Lee J.C.M. & Jain A.K. 1998. Automatic Classification of Tennis Video for High-level Content-based Retrieval. Teoksessa Proceedings of the 1998 International Workshop on Content-Based Access of Image and Video Databases (CAIVD '98), Bombay, India. IEEE Computer Society Washington, DC, USA, 81-90.

Tjondronegoro D.W., Chen Y.-P.P. & Joly A. 2008. A Scalable and Extensible Segment-Event-Object-Based Sports Video Retrieval System. ACM Transactions on Multimedia Computing, Communications, and Applications, 4(2), Article 13.

Tjondronegoro D.W., Chen Y.-P.P. & Pham B. 2005. Content Based Video Indexing for Sports Applications using Integrated Multi-Modal Approach. Teoksessa Proceedings of the ACM MM'05, Singapore, November 6-11. ACM, New York, NY, USA, 1035-1036.

Tjondronegoro D.W., Chen Y.-P.P. & Pham B. 2006. Extensible Detection and Indexing of Highlight Event in Broadcasted Sports Video. Teoksessa Proceedings of the 29th Australasian Computer Science Conference, Hobart, Australia. Australian Computer Society Inc., Darlinghurst, Australia, 237-246.

Windows Hardware Developer Central 2001, DV Video Data and AVI Files [online]. Microsoft Corporation [viitattu 5.12.2008]. Saatavilla www-osoitteessa: http://www.microsoft.com/whdc/archive/dvavi.mspx

W3C 2006. Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26 November 2008 [viitattu 5.12.2008]. Saatavilla www-osoitteessa: http://www.w3.org/TR/xml/

W3C 2007. XQuery 1.0: An XML Query Language. W3C Recommendation 23 January 2007 [viitattu 4.12.2008]. http://www.w3.org/TR/xquery/

Xu C., Wang J., Wan K., Li Y. & Duan L. 2006. Live Sports Event Detection Based on Broadcast Video and Web-Casting Text. Teoksessa Proceedings of the 14th Annual ACM International Conference on Multimedia, October 23-27, Santa Barbara, California, USA, 221-203.

Yu X. & Farin D. 2005. Current and Emerging Topics in Sports Video Processing. Current and Emerging Topics in Sports Video Processing. Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, July 6-8, 526-529.

Zhang H.J., Tan S.Y., Smoliar S.W. & Yihong G. 1995. Automatic Parsing and Indexing of News Video. Multimedia Systems, vol. 2, 256-266.

Zhao Y., Zhou X. & Tang G. 2006. Audiovisual Integration for Racquet Sports Video Retrieval. Teoksessa Lecture Notes in Computer Science: Advanced Data Mining and Applications, vol. 4093/2006, Springer-Verlag Berlin/Heidelberg, 673-680.

Zhou W. & Dao S.K. 2001. Combining hierarchical classifiers with video semantic indexing systems. Teoksessa Proceedings of Advances in Multimedia Information Processing – PCM, 2nd IEEE Pacific Rim Conference on Multimedia, Beijing, China, October, 78-85.

Zhou W., Vellaikal A. & Kuo C.-C. 2000. Rule-Based Video Classification System for Basketball Video Indexing. Teoksessa Proceedings of the ACM Multimedia Workshop, Del Ray, CA, USA, 213-216.