

Mikko Tanni

TV-uutisten automaattinen jäsentäminen

Segmentointi ja jaksoiden tunnistaminen digitaalisen videon automaattisen indeksoinnin perustana

Mikko Tanni: TV-uutisten automaattinen jäsentäminen. Segmentointi ja jaksoiden tunnistaminen digitaalisen videon automaattisen indeksoinnin perustana [Automatic parsing of TV-news. Segmentation and episode recognition as the basis of automatic indexing of digital videos.] Informaatiotutkimus 22 (4), pp 121-127.

Digital video is getting increasingly common, and many broadcasters consider switching to digital formats. Video data requires appropriate indices to be available for effective use. Manual indexing methods are ineffective in high volume databases. This article is an introduction to the basics of automatic parsing of news videos, the first step of content based indexing. The article goes through shot detection methods and presents episode extraction by modelling, the two fundamental phases of recovering news video structure. Also, a news broadcast from Finnish Broadcast Company (YLE) is considered in the light of these content based segmentation methods.

Address: Mikko Tanni, Department of Information Studies, University of Tampere, Finland FIN-33014 University of Tampere. Email: mikko.tanni@uta.fi

1. Johdanto

TV-uutiset olisivat hyödyllisiä tiedonlähteitä, mutta niiden relevantti sisältö ei ole helposti tiedonhakijoiden saatavilla. Monet laajat elokuva- ja TV-arkistot perustuvat tekstipohjaiseen aineiston kuvailuun; indeksi on tekstimuodossa palvelimella ja varsinaisen videodata nauhoilla (Buscher 1999; Del Pero, Dimino & Stroppiana 1999). Videonauhurin käyttöliittymä sopii huonosti selailuun ja nauhat ovat kömpelö media, joten perinteiset videonauhoihin perustuvat arkistot eivät tarjoa tarkoituksenmukaista ratkaisua videotiedonhakuun (Del Pero ym. 1999). I Jotkut tutkijat ovat vielä epäilleet, että voidaanko tekstimuodossa esittää videosisältöjä ilman, että alkuperäisen datan keskeisiä ominaisuuksia ei kadotettaisi (ks. Bolle ym. 1998; Grosky 1997, 73; Gupta & Jain 1997, 72). Toisin sanottuna mitä erilaisemmiksi digitaalisiin videoarkistoihin, videoiden tallentamisen hallitsemiseen, säilyttämiseen, jakeluun

ja hakemiseen liittyvät tiedontarpeet muodostuvat sitä monipuolisempia tekniikoita vaaditaan niiden tyydyttämiseksi (Buscher 1999). Manuaalisen indeksoinnin periaatteellisten ongelmien ja videomateriaalin valtavan määrän vuoksi viime vuosina on alettu kiinnittämään yhä enemmän huomiota digitaaliseen muotoon muutettujen uutisvideoiden automaattiseen indeksointiin: esimerkiksi Euroopan yleisradioliitolla 2 on ollut useita projekteja aiheesta (ks. esim. Baker & Bulford 1996; Daubney 1996; Buscher 1999; Del Pero ym. 1999).

Idrisin ja Panchanathanin (1997, 146) mukaan nykyään videoiden indeksointia lähestytään kuvananalyysin ja -ymmärtämisen näkökulmasta. Heidän mukaansa nykytutkimuksessa pyritään kehittämään automaattisia ja aiheriippumattomia hahmopohjaisia (content based) tekniikoita, jotka mahdollistavat visuaalisen datan indeksoimisen ja hakemisen sisällön perusteella. Hahmopohjaisissa indeksointimenetelmissä algoritmit poimivat videodatasta matemaattisesti määriteltäviä piirteitä,

jotka esittävät kuvan visuaalisia ominaisuuksia, kuten värejä, tekstuureita ja muotoja. Piirteitä käytetään videosisällön edustamiseen (eli representaatioon) tekstuaalisten kuvailuterminien sijaan. (Idris & Panchanathan 1997, 146; ks. Del Bimbo 1999, 1.) Hahmopohjaisilla hakuvälineillä voidaan tehdä visuaalisia kyselyjä suoraan videoiden *sisältöön* (eli kuvainformaatioon) mutta ei videoista *tulkittuihin merkityksiin* (ks. Prabhakaran 1997, 54; Tanni 2003, 21–22).

Videoiden indeksointi eroaa tekstidokumenttien indeksoinnista. Video on sekventiaalinen ja ajallisesti jatkuva esitysmuoto, joka ei rakennu kirjoitusmerkkien ja sanojen kaltaisista selkeistä yksiköistä. Jotta videoiden – ja visuaalisen informaation yleensäkin – indeksointi olisi mahdollista, on tunnistettava sellaisia elementtejä, jotka sopivat indeksoitaviksi perusyksiköiksi.

Tässä artikkelissa esitellään, kuinka olemassa olevalla hahmopohjaisella indeksointiteknologialla voidaan yksinkertaisella tavalla jäsentää automaattisesti TV-uutisten sisältöä myöhemmässä indeksointivaiheessa tapahtuvaa selailua tai sisälöntunnistamista varten. Laajoja katsauksia videon hahmopohjaisista indeksointimenetelmistä, joissa käsitellään koko indeksointiprosessia, ovat esittäneet muun muassa Idris ja Panchanathan (1997), Bolle, Yeong ja Yeong (1998), Brunelli, Mich ja Modena (1999), Petkovic ja Jonker (2000) sekä Antani, Kasturi ja Jain (2002). Tämä artikkeli perustuu pro gradu -tutkielmaan, joka käsittelee videotiedonhakua, videotiedonhaun käyttöliittymiä ja videoiden indeksointia TV-uutisten näkökulmasta (Tanni 2003).

Artikkeli on jäsennetty seuraavalla tavalla. Luvussa 2 tarkastellaan videoiden ja uutisten ajallisaikallisia elementtejä. Tämän jälkeen luvussa 3 käydään läpi segmentointimenetelmiä käsittelevää kirjallisuutta. Luvussa 4 tarkastellaan käsiteltyjen segmentointimenetelmien soveltumista Yleisradion televisiuutisiin. Luku 5 on varattu johtopäätöksille.

2. Uutisvideoiden ajallisaikallinen rakenne

2.1 Jaksottaisuus videoissa

Video on sekvenssi erillisiä staattisia kuvia, joiden esittäminen nopeasti luo vaikutelman dynaamisuudesta, liikkeestä ja jatkuvuudesta. Vaikka video on esitysmuotona jatkuva, se on visuaaliselta

sisällöltään ja semantiikaltaan *jaksottainen*. Niinpä kukin videosekvenssi on jaettavissa alisekvensseihin eli ajallisiin yksiköihin, joiden väliset siirtymät ovat havaittavissa visuaalisina eroina ja ymmärrettävissä semanttisena jaksottaisuutena, esimerkiksi uutisvideoiden tapauksessa uutisaiheina ja niiden välisinä eroina. Visuaalisella sisällöllä tarkoitetaan piirteiden esiintymistä ja jakautumista kuvasekvenssissä. Semantiikalla tarkoitetaan tässä yhteydessä videon ajallisten yksiköiden eroja merkityksessä. Videon visuaalinen jaksottaisuus, eli siitä visuaalisesti havaittavissa olevat alisekvenssit ja niiden väliset erot, mahdollistaa TV-uutisten automaattisen jäsentämisen: itse asiassa videoiden jäsentämisessä on kyse jaksottaisuuden tunnistamisesta ja niiden merkityksellistämistä mallien avulla. TV-uutisten jäsentämisen kannalta keskeisimpiä ajallisia yksiköitä ovat *otto* ja *jakso*.

Otolla tarkoitetaan (yleensä lyhyttä) ajanjaksoa, jonka yksittäinen kamera nauhoittaa: otolla on siis selvä yhteys videoilmaisuun tekniseen ulottuvuuteen. Otto on videossa lyhin merkityksellinen ajallinen yksikkö – ja näin ollen luontevasti indeksoinnin perusyksikkö. 3 Ottojen väliset erot ovat havaittavissa visuaalisina eroina; visuaalisia eroja eli muutoksia kuvan pikseleiden värispektrissä voidaan pitää fyysisinä. (Ks. mm. Antani ym. 2002, 955; Bolle ym. 1998; Brunelli ym. 1999, 81; Del Bimbo 1999, 10; Idris & Panchanathan 1997, 154.)

Uutisvideoissa ottoja laajempia merkityksellisiä ajallisia yksiköitä ovat jaksot, jotka muodostuvat tietyistä sarjasta ottotyyppejä (Del Bimbo 1999, 10). Jaksojen välillä ja niiden sisällä ei ole välttämättä jatkuvuutta paikassa, ajassa ja toiminnassa; tämä tarkoittaa, että jaksojen otot eivät välttämättä muistuta toisiaan visuaalisesti. Jaksoja ovat esimerkiksi uutisjutut ja -sähkeet.

Video on siis ajallisesti jatkuva esitysmuoto, ja esitysmuodon välittämä videosisältö on jaksottaisista: ajallisten yksiköiden *sisällä* on sekä visuaalista että semanttista jatkuvuutta; vastaavasti ajallisten yksiköiden *välillä* jatkuvuus katkeaa. Video, joka ei sisällä fyysisiä tai semanttisia katkoja, muodostuu yhdestä ainoasta otosta, joka käsittelee samaa kohdetta ja aihetta.

2.2 Uutislähetysten ajallisaikalliset elementit

Uutislähetys alkaa *uutistunnuksella*; sen jälkeen vuorossa on *uutisankkurin tervehdys*; sitten vuor-

ossa on lähetyksen *ensimmäinen sisällysluettelo*, jossa käydään läpi uutislähetyksen tärkeimmät aiheet; sisällysluettelon jälkeen siirrytään lähes poikkeuksetta *illan pääuutisjuttuun*; ensimmäistä uutisjuttua seuraa *uutissähkeitä ja –juttuja*; uutislähetyksen puolivälissä on *toinen sisällysluettelo*, jossa käydään läpi lähetyksen lopun aiheita; lähetyksen loppupuolella on *säätiedotus*; viimeisiä jaksoja ovat *lopetuspuhe* ja *lopetustunnus*.

Keskeisiä paikkasidonnoisia (spatial) elementtejä uutislähetyksessä ovat uutistenlukija eli *uutisankkuri*, *uutisikkuna*, jossa näytetään uutisankkurin vasemmalla puolella käsiteltävän uutisjutun otsikko ja siihen liittyvä usein symbolinen tai metonyminen kuva, *nimipalkki*, jossa näytetään uutisankkurin nimi, ja *tausta*. Nämä elementit esiintyvät uutisstudiossa.

Paikkasidonnoisten elementtien esiintyminen ajallisissa elementeissä vaihtelee. Nimipalkki esiintyy vain uutisankkurin tervehdyksessä. Uutisikkuna on näkyvillä uutisstudiossa aina uutisjuttua aloitettaessa, ja joskus myös lopetettaessa palataan uutisstudioon ennen uutisjutun vaihtumista. Seuraava uutisjuttu alkaa kuitenkin uutisstudiossa, jos uutisikkunan sisältö on vaihtunut. Uutissähkeitä luettaessa uutisikkunaa ei ole, ja kuvausetaisyys on lähempää uutisankkuria.

3. Menetelmät

Tässä luvussa käsitellään uutislähetysten jäsentämiseen sopivia menetelmiä. Ottoja pidetään niiden teknisen, kameran käyttöön liittyvän luonteen takia indeksoitavina perusyksiköinä. Näin ollen uutisvideoiden indeksointi alkaa ottorajojen tunnistamisella eli segmentoinnilla. Varsinaisena haasteena automaattisessa indeksoinnissa voidaan kuitenkin pitää semanttisten jaksosten tunnistamista: sen lisäksi, että uutislähetys saadaan jäsennettyä ottoihin, on tiedonhaun kannalta oleellista, että uutisjutut ja -sähkeet saadaan eroteltua toisistaan ja asetettua käyttäjien saataville.

Hahmopohjaisen indeksointijärjestelmän näkökulmasta ottoja kutsutaan segmenteiksi, ja ottorajojen tunnistamiseen käytetyt menetelmät ovat segmentointialgoritmeja. Koska hahmopohjaiset menetelmät tunnistavat vain visuaalisia eroja, ne eivät ymmärrä jaksottaisuutta eli jatkuvuutta ja epä-jatkuvuutta semanttisella tasolla. Tästä syystä on luotava päättelysääntöjä, joilla videon visuaaliset piirteet voidaan kartoittaa semanttiseksi yksiköiksi tiettyä aihetta koskevan tietämyksen perusteella eli

mallintamalla olosuhteita ja käytäntöjä, joilla video on tuotettu. Myöhemmin luvussa esitetään, kuinka jaksot pyritään tunnistamaan sarjoina tietynlaisia ottotyyppisiä.

Hahmopohjaiset segmentointimenetelmät ovat videoiden indeksoinnin kannalta perustavassa asemassa, ja niitä on käsitelty kirjallisuudessa kattavasti.

3.1 Segmentointi

Ottojen väliset rajat voivat olla välittömiä (sharp) eli leikkauksia tai asteittaisia (gradual) eli siirtymätehosteita (Bolle ym. 1998; Brunelli ym. 1999, 81; Yeo & Yeung 1997, 47). Periaatteessa ottojen väliset erot tunnistetaan niiden peräkkäisten kuvien tiettyjen piirteiden välisestä epäjatkuvuudesta (Apers ym. 1997, 172). Oletuksena on, että ottojen sisällä kahden peräkkäisen kuvan välillä ei ole oleellisia eroja taustassa eikä hahmoissa (Del Bimbo 1999, 204–205). Mikäli ottojen välillä on suuria eroja tarkkailtavien piirteiden suhteen ja oton sisällä muutokset ovat vähäisiä, indeksointi on tarkempaa; mikäli ottojen väliset erot ovat pieniä ja kunkin oton sisällä on suuria muutoksia piirteiden suhteen, segmentointitarkkuus laskee. Del Bimbon (1999, 204–205) mukaan ongelmia aiheuttavat esimerkiksi äkilliset muutokset valaistuksessa ottojen sisällä (esimerkiksi salamavallo) ja hahmojen tai kameran liike kuvasarjassa. Tällaisissa tilanteissa voi olla vaikeaa määrittellä erot paikallisia, ottojen sisäisiä, vai johtuvatko erot ottojen vaihtumisesta. (Del Bimbo 1999, 204–205.)

Yksinkertaisimmillaan ottojen välisiä eroja etsitään vertaamalla kuvavirran peräkkäisten kuvien pikseleitä vastaavissa kohdissa kuvia: jos muuttuneiden pikselien määrä kuvien välillä on suurempi kuin ennalta asetettu kynnyks, ottoraja merkitään. Tämänkaltaiset menetelmät ovat kuitenkin erittäin herkkiä hälylle ja erityisesti kaikenlaiselle liikkeelle ottojen sisällä, sillä liike aiheuttaa myös muutoksia pikseleiden asemassa yksittäisten kuvien välillä. (Ks. Antani ym. 2002, 955–956; Idris & Panchanathan 1995, 155; Rui ym. 1999, 360; Xiong ym. 1997, 52.)

Pikselien vertailun sijaan ottojen tunnistamiseen voidaan käyttää histogrammeja eli väri-intensiteettijakaumia, jotka esittävät värien jakautumista kuvissa. Ideana on, että oton sisällä värijakauman muutokset kuvasekvenssin yksittäisten kuvien välillä ovat pienempiä kuin kahden eri oton välillä: kun erot kahden kehyksen välillä ovat enemmän kuin

asetettu kynnyks, ottoraja merkitään. Histogrammeihin perustuvat menetelmät ovat huomattavasti vähemmän herkkiä kameran ja hahmojen liikkeille kuin pikseleiden vertailuun perustuvat menetelmät: samaa kohdetta kuvattaessa värijakaumat eivät tietenkään muutu niin paljoa kuin yksittäiset pikselit. Ongelmat histogrammeissa tulevat esiin, kun kahden otton välillä on vain vähäisiä muutoksia; tällöin liian korkeaksi asetettu kynnyks estää ottorajojen tunnistamisen. (Ks. Brunelli ym. 1999, 81–82; Rui ym. 1999, 360; Idris & Panchanathan 1997, 156.)

Segmentointitarkkuutta voidaan parantaa jakamalla kokonainen kuva osiin eli lohkoihin ja tarkastelemalla erikseen kussakin lohossa tapahtuvia muutoksia piirteiden suhteen. Mikäli riittäviä muutoksia on tapahtunut tarpeeksi suuressa osassa lohkoja, ottoraja merkitään. (Ks. Antani ym. 2002, 956; Idris & Panchanathan 1997, 156; Xiong ym. 1997, 52; Prabhakaran 1997, 77.)

Videoiden editointivaiheessa käytettävät tehosteet levittävät siirtymän kahden otton välillä useamman kuvan mittaiseksi. Peräkkäisten kuvien eroja vertailevat segmentointimenetelmät eivät toimi asteittaisissa siirtymissä, koska muutokset kehysten välillä ovat niin pieniä, että muuten tarkoituksenmukaiset kynnykset eivät niitä havaitse (ks. Brunelli ym. 1999, 85). Liian herkän eli matalan kynnyksen käyttäminenhan saattaa aiheuttaa ottojen sisäisten muutosten tulkittamisen ottorajaksi. Periaatteessa ongelma ratkaistaan käyttämällä asteittaisia siirtymiä varten kahta kynnyksä: kahteen vertailuun perustuvissa menetelmissä oletetaan, että kehukset ennen ja jälkeen asteittaisen siirtymän ovat yleensä huomattavan erilaisia, vaikka erot siirtymässä itsessään ovat pieniä. Ensimmäisellä vertailukerralla käytetään korkeaksi asetettua kynnyksä välittömien siirtymien löytämiseksi ja toisella kerralla etsitään tehosteiden alkamiskohtia kuvasekvenssistä. (Ks. Rui ym. 1999, 360; Apers ym. 1997, 172–173; Brunelli ym. 1999, 88; Idris & Panchanathan 1997, 156.) Koska TV-uutiset ovat visuaalisesti samankaltaisia lähetyksestä toiseen, niissä käytettyjä tehosteita voidaan tunnistaa tuottamalla niistä erityisiä matemaattisia malleja (ks. Brunelli ym. 1999, 86–87; Idris & Panchanathan 1997, 157).

Uutisstudion sisäisten siirtymien havaitsemiseksi voi olla tarpeellista tarkastella ottorajojen vaihtumista kuvan hahmojen reunojen ilmestymisen ja katoamisen avulla. Periaatteena näissä menetelmissä on, että ottorajoissa ja joissain editointitehosteissa kuvassa esiintyvien hahmojen

ja alueiden reunoja ilmestyy kaukana edellisten kuvien hahmojen reunoista, ja vanhat reunat katoavat kaukana uusista. (Ks. Antani ym. 2002, 956; Brunelli ym. 1999, 87–88; Rui ym. 1999, 360.) Jos esimerkiksi kuvassa on ohiajava auton aariviivat muodostavat reunoja taustaa vasten; kun auto katoaa kuvasta kohtauksen vaihtuessa, reunat muuttuvat kahden kuvasekvenssin kuvan välillä kerralla enemmän kuin otton sisällä.

Hahmojen tunnistamiseen käytettäviä menetelmiä käsittelee tarkemmin muun muassa Prabhakaran (1997, 71–72). Käytännössä hahmojen tunnistamiseen käytettävät menetelmät perustuvat hahmojen välisten rajojen tunnistamiseen tai pikselien määrittämiseen hahmojen sisä- tai ulkopuolelle kuuluviksi.

Segmentointialgoritmien tarkkuus vaihtelee. Yleisesti ottaen voidaan sanoa kuitenkin, että leikkausten tunnistamiseen tarkoitettujen algoritmien luotettavuus on riittävä uutislähetysten jäsentämistä varten, mutta asteittaisten siirtymien tunnistaminen saattaa aiheuttaa ongelmia (ks. Brunelli ym. 1999, 90).

3.2 Jaksojen tunnistaminen

TV-uutisten indeksoinnin kannalta on järkevää pitää uutisjuttuja ja -sähkeitä tiedonhaun perusyksikköinä: koska ne eivät ole liian pitkiä selattavaksi ja ne käsittelevät tiettyä selvästi rajattua aihetta, videoiden semanttisesta sisällöstä kiinnostuneelle ne ovat ilmiselviä hakukohteita (ks. Mills ym. 2000, 4). Lisäksi on oletettavaa, että TV-uutiset kiinnostavat hakujärjestelmän käyttäjiä samasta syystä kuin ne kiinnostavat uutisten katsojiaakin. Toki hakija voi olla kiinnostunut yksittäisistä ostoistakin: hakujärjestelmän kannalta niiden järjestäminen käyttäjän saataville on helpompaa kuin jaksojen.

TV-uutislähetys on rakenteinen videotyyppi: se muodostuu joukosta silmiinpistäviä tietyllä tavalla järjestettyjä ajallisaikallisia elementtejä, joille indeksointi voidaan perustaa (ks. Hietala 1996, 63). Koska uutiset ovat rakenteellisesti samankaltaisia lähetyksestä toiseen, niiden jäsentämistä on luontevaa lähestyä uutisia varten luotujen erityisten mallien kautta. Mallien ideana on eksplikoita indeksointijärjestelmälle säännöt ja käytännöt, joilla ajallisten elementtien visuaalinen ulkoasu ja esitysjärjestys on tuotettu.

Malleihin perustuvissa menetelmissä uutisjuttuja ja -sähkeitä jäsennetään tiettyjen ottojen esiinty-

misjärjestyksen ja niistä tunnistettujen hahmojen esiintymisen ja sijainnin avulla. Toisin sanottuna semanttiset jaksot pyritään tunnistamaan tiettyjen visuaalisten piirteiden ja ottojen esiintymisen säännönmukaisuuksien avulla. Jäsentäminen aloitetaan segmentoimalla uutisvideo ja valitsemalla kustakin otosta *avainkuva* edustamaan koko ottoa; avainkuva voi olla esimerkiksi kuvasekvenssin ensimmäinen kuva. Avainkuvat luokitellaan uutisankkurin sisältäviin ottoihin (uutisaiheiden alustuksesta) ja varsinaisten uutisjuttujen ottoihin toimittajan osuudesta. Luokittelu tehdään *kuvamallien* ja *aluemallien* avulla: kukin kuvamalli esittää paikkasidonnaisesti järjestettyjen aluemallien sijainnin, ja avainkuvia verrataan kuvamalleihin. Aluemallit esittävät kuvan paikkasidonnaisia elementtejä, jotka avainkuviissa esiintyvät: uutisankkuria, uutisikkunaa, nimipalkkia ja taustaa. Kuvamallit muodostuvat siis sommitelmasta aluemalleja. (Ks. Bolle ym. 1998; Brunelli ym. 1999, 104; Del Bimbo 1999, 51–52, 227–228; Tanni 2003, 107–108.) Mikäli oton avainkuva ei sovi mihinkään uutisstudioti kuvaavaan kuvamalliin, ottoa ei oteta huomioon jäsentämisessä; vain uutisankkurin sisältäviä ottoja käytetään jaksoiden tunnistamiseen. Kuvamallien avulla päätellään, mistä kukin uutisjuttu tai -sähke alkaa ja koska se loppuu, sillä kukin uutislähetys muodostuu ajallisesti tietyssä järjestyksessä esiintyvistä kuvamalleista: uutisjuttu alkaa uutisstudiosta ja seuraavan kerran uutisstudioon siirryttäessä alkaa uusi uutisjuttu, *mikäli* uutisikkunan sisältö on erilainen kuin uutisjutun aloittaneen oton avainkuvassa. Kuvamallit ja niiden väliset siirtymät (eli leikkaukset tai tehosteet) osoittavat siis uutisjuttujen ja -sähkeiden alkamista ja loppumista.

4. Yleisradion TV-uutisista tehtyjä havaintoja

Tässä luvussa tiivistetään Yleisradion pääuutislähetyksestä tehtyjä havaintoja segmentoinnin ja jaksoiden tunnistamisen näkökulmasta (ks. Tanni 2003, 136: liite 8). Ensimmäinen, lähes itsestään selvä oletamus on, että Yleisradion TV-uutiset noudattavat ajalliselta rakenteeltaan ja visuaaliselta ulkoasultaan samaa kaavaa lähetyksestä toiseen. 4 Näin ollen yhdestä lähetyksestä tehdyt havainnot ovat yleistettävissä.

Automaattisen jaksoiden tunnistamisen perusoletus on, että (1) tietyntyyppiset otot ja niistä koostuvat jaksot esiintyvät ajallisesti tietyssä

järjestyksessä ja tietyillä paikoilla, ja (2) tietyntyyppisillä otoilla ja jaksoilla on yleensä samankaltaisia visuaalisia ominaisuuksia, joita esittävien mallien perusteella on mahdollista päätellä näiden ajallisaikallisten elementtien merkitykset uutislähetysten kontekstissa. Jäsentämisessä on otettava huomioon, että uutislähetyksessä esiintyy pitkällä ajanjaksolla (n. 25 minuuttia) visuaalisesti samankaltaisia ottoja (eli otot uutisstudiossa), jotka kuuluvat semanttisesti eri jaksoihin: uutisjuttujen juonnot lasketaan osaksi varsinaisia uutisjuttuja, sillä juonnot sisältävät uutisjuttuihin liittyvää informaatiota. Ongelmaksi automaattisen hahmopohjaisen indeksoinnin kannalta muodostuu, että usein uutisaiheissa visuaalisesti erilaiset otot kuuluvat semanttisesti yhtenevään otoryhmään eli jaksoon. Yksinkertaistavat oletukset visuaalisen samankaltaisuuden ja semanttisen samankaltaisuuden yhteyksistä eivät siis pidä paikkaansa kuin korkeintaan silloin, kun samankaltaiset otot ovat peräkkäin tai ajallisesti hyvin lähellä toisiaan. Ylimääräisen ongelman muodostavat vielä poikkeukset säännöistä: edellä jo mainittiinkin tilanteesta, jossa ankkuri kommentoi vielä uutisjuttua sen sijaan, että siirtymä uutisstudioon takaisin aloittaisi uuden jutun.

Pääuutisjutun 17:stä otosta 11 päättyy asteittaiseen siirtymään (ks. Tanni 2003: liite 8). Uutislähetyksessä käytettävät siirtymätehosteet saattavat tehdä ottojen välisistä rajoista vaikeasti havaittavat varsinkin, jos ottojen väliset erot eivät ole muutenkaan suuret. Siirtymätehosteiden havaitseminen voi olla erityisen vaikeaa esimerkiksi, kun otot käsittelevät samaa kohdetta – tällaisessa tapauksessa tosin ottojen huomauttamatta jättäminen ei ole uutislähetysten jäsentämisen kannalta kovinkaan haitallista. Päinvastaisen ongelman muodostavat esimerkiksi uutistunnuksessa kuvaan ilmestyvät hahmot, jotka aiheuttavat silmiinpistäviä muutoksia; segmentointialgoritmi saattaa tulkita suurien hahmojen esiintymisen kuvissa oton vaihtumiseksi. 5

Histogrammipohjaiset menetelmät eivät ehkä havaitse siirtymiä uutisstudion sisällä, koska väri- ja kaumioiden muutokset eivät ole tarpeeksi suuria; histogrammien erot varsinaisten uutisjuttujen ja uutisstudion välillä ovat kuitenkin yleensä selvät. Uutisstudioti koskevien ottotyyppien erottamiseksi toisistaan voidaan käyttää hahmojen reunojen tunnistamiseen perustuvia segmentointimenetelmiä, hahmontunnistusta paikallisten elementtien löytämiseksi ja lohkopohjaisia menetelmiä, joiden avulla tunnistetaan uutisikku-

nassa tapahtuvat muutokset tai sitten uutisikkunan puuttuminen (uutissähkeiden tapauksessa).

5. Johtopäätökset ja keskustelu

Uutisvideoiden jäsentäminen perustuu ottojen segmentoinnille ja uutisstudiota esittävien ottojen erottamiseen varsinaisten uutisjuttujen ostoista. Erottaminen tehdään vertaamalla kunkin otton avainkuvan tiettyjä piirteitä – lähinnä hahmoja ja värijakaumia – etukäteen tuotettuihin paikkasidonnaisiin malleihin.

Uutismateriaalia tarkasteltaessa eräs keskeinen havainto on, että jotta jäsentäminen toimisi joka tilanteessa, ajallisiin malleihin on rakennettava poikkeavat tilanteet: esimerkiksi joskus palataan uutisstudioon kesken uutisjutun käsittelyn esimerkiksi uutisankkurin lisätessä juttuun kommenttejaan tai ankkurin haastattellessa jotakuta. Uutislähetystä esittävät mallit ovat ikään kuin merkijärjestelmiä: ne sisältävät merkkejä (eli kuva- ja aluemalleja) ja erilaisia sääntöjä (eli ajallisia malleja), joilla ”merkkejä” järjestetään ajallisesti. Näiden säännönmukaisuuksien eksplikointi mahdollisesti yhteistyössä uutistuottajien kanssa olisi mielenkiintoinen tutkimuksellinen haaste.

Tässä artikkelissa lähestyttiin segmentointia yksinkertaisella visuaalisella tasolla. Eräs videon keskeinen ominaisuus, multimodaalisuus, äänen ja kuvan toimiminen yhdessä, jätettiin kokonaan huomiotta. Kuitenkin on oletettavaa, että parhaat tulokset saavutetaan erilaisten menetelmien yhteiskäytöllä – tässä on myös mielenkiintoinen aihe tutkimukselle.

Uutislähetysten ajallisen rakenteen tunnistaminen on varsinaisen indeksointiprosessin ensimmäinen vaihe. Myöhemmissä vaiheissa keskeiseen osaan nousevat videosisällön tunnistaminen ja sen visualisointi eli sisällön esittäminen paremmin selailua tukevassa muodossa. Aihetta koskeva tutkimus voisi kartoittaa tarkemmin käyttäjien tarpeita haettavien nimekkeiden suhteen: välittävätkö esimerkiksi videoilmaisun tekniset yksiköt eli otot videoinformaatiota käyttäjien tarpeita vastaavalla tavalla?

Hyväksytty julkaistavaksi 1.12.2003

Viitteet

1 Videotiedonhaussa selailun merkitys korostuu (ks. Tanni 2003, 81–105).

2 European Broadcasting Union eli EBU

3 Tekstejä indeksoitaessa yksittäisiä sanoja voidaan tarpeen mukaan käsitellä morfeemeina, ja vastaavalla tavalla myös videoita indeksoitaessa voidaan ottoja käsitellä yksittäisinä kuvina. Morfeemit ja ostoista poimitut avainkuvat eivät ole kuitenkaan samalla tavalla merkityksellisiä kuin kokonaiset sanat ja otot.

4 Uutislähetysten visuaalinen ilme uusitaan suhteellisen harvoin.

5 Joissain tapauksissa, esimerkiksi kuvattaessa ohi ajavia autoja, voi olla järkevääkin muodostaa segmenttejä objektien ilmestymisen ja katoamisen perusteella, jotta segmentit eivät muodostuisi kohtuuttoman pitkiä.

Lähteet

- Antani, S., Kasturi, R. & Jain, R. 2002. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern recognition* 2002 (35), 945–965.
- Apers, P. M. G., Blanken, H. M. & Houtsma, M. A. W. 1998. *Multimedia Databases in Perspective*. London: Springer-Verlag.
- Baker, I. & Bulford, B. 1996. A cost-effective strategy for the archiving of ageing film. *EBU Technical Review*, Summer 1996. Saatavilla pdf-muodossa: <URL: http://www.ebu.ch/trev_268-baker.pdf>. (Viitattu 10.11.2003.)
- Bolle, R. M., Yeo, B-L. & Yeung, M. M. 1998. Video query: Research directions. *IBM Journal of Research and Development* 42 (2). Saatavilla [www](http://www.research.ibm.com/journal/rd/422/bolle.html)-muodossa: <URL:<http://www.research.ibm.com/journal/rd/422/bolle.html>>. (Viitattu 14.3.2002.)
- Brunelli, R., Mich, O. & Modena, C. M. 1999. A Survey on the Automatic Indexing of Video Data. *Journal of Visual Communication and Image Representation* 10, 78–112.
- Buscher, I. 1999. Digital video archives – facing the facts. *EBU Technical Review*, Summer 1999. Saatavilla pdf-muodossa: <URL: http://www.ebu.ch/trev_280-buscher.pdf>. (Luettu 10.11.2003.)

- Daubney, C. 1996. Maintaining a valuable programme archive in the face of technological change – some pragmatic thoughts. EBU Technical Review, Summer 1996. Saatavilla pdf-muodossa: <URL: http://www.ebu.ch/trev_268-daubney.pdf>. (Viitattu 10.11.2003.)
- Del Pero, R., Dimino, G. & Stroppiana, M. 1999. Multimedia catalogue – the RAI experience. EBU Technical Review, Summer 1999. Saatavilla www-muodossa: <URL: http://www.ebu.ch/trev_280-stroppiana.pdf>. (Viitattu 10.11.2003.)
- Del Bimbo, A. 1999. Visual Information Retrieval. San Francisco, California: Morgan Kaufmann Publishers Inc.
- Grosky, W. I. 1997. Managing Multimedia Information in Database Systems. Communications of the ACM 40 (12), 73–80.
- Gupta, A. & Jain, R. 1997. Visual Information Retrieval. Communications of the ACM 40 (5), 71–79.
- Gupta, A., Santini, S. & Jain, R. 1997. In Search of Information in Visual Media. Communications of the ACM 40 (12), 35–42.
- Hietala, V. 1996. Ruudun hurma: johdatus TV-kulttuuriin. Jyväskylä: Gummerus.
- Idris, F. & Panchanathan, S. 1997. Review of Image and Video Indexing Techniques. Journal of Visual Communication and Image Representation 8 (2), 146–166.
- Mills, T. J., Pye, D., Hollinghurst, N. J. & Wood, K. R. 2000. AT&TV: Broadcast Television and Radio Retrieval. Paper presented at RIAO 2000 (Recherche d'Informations Assistée par Ordinateur; Computer Assisted Information Retrieval). April 2000. Paris.
- Petkovi, M. & Jonker, W. 2000. An Overview of Data Models and Query Languages for Content-based Video Retrieval. Paper presented at the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet. 31.7.–6.8.2000. L'Aquila, Italy.
- Pietilä, V. 1995. TV-uutisista hyvää iltaa. Tampere: Vastapaino.
- Prabhakaran, B. 1997. Multimedia Database Management Systems. Boston, London and Dordrecht: Kluwer Academic Publishers.
- Rui, Y., Huang, T. S. & Mehrotra, S. 1999. Constructing table-of-content for videos. Multimedia Systems 7, 359–368.
- Tanni, M. 2003. Digitaalisen videon automaattinen sisällönkuvailu TV-uutisissa ja videotiedonhaun käyttöliittymät. Tampere: Tampereen yliopiston kirjasto. <URL: <http://tutkielmat.uta.fi/pdf/gradu00189.pdf>>. (Viitattu 5.6.2003.)
- Xiong, W., Chung-Mong Lee, J. & Ma, R-H. 1997. Automatic video data structuring through shot partitioning and key-frame computing. Machine Vision and Applications, 1997 (10), 51–65.
- Yeo, B-L. & Yeung, M.M. 1997. Retrieving and Visualizing Video. Communications of the ACM 40 (12), 43–52.