DIGITAALISEN VIDEON AUTOMAATTINEN SISÄLLÖNKUVAILU TV-UUTISISSA JA VIDEOTIEDONHAUN KÄYTTÖLIITTYMÄT

Mikko Tanni

Pro gradu -tutkielma
Maaliskuu 2003
Informaatiotutkimuksen laitos
Tampereen yliopisto

Tampereen yliopisto
Informaatiotutkimuksen laitos
TANNI, MIKKO: Digitaalisen videon automaattinen sisällönkuvailu TV-uutisissa ja videotiedonhaun käyttöliittymät
Pro gradu -tutkielma, 143 s., 13 liitettä.
Informaatiotutkimus
Maaliskuu 2003

TIIVISTELMÄ

Tässä kirjallisuustutkielmassa käsitellään digitaalisten videoiden automaattiseen sisällönkuvailuun soveltuvien hahmopohjaisten indeksointimenetelmien periaatteita. Lähteet on valittu ja teksti jäsennetty TV-uutisten erityisvaatimusten ja -ominaisuuksien perusteella. Tarkastelu keskittyy metadatan tuottamiseen videodatan havaittavista piirteistä sisällönkuvailun eri tasoilla. Indeksoinnin lisäksi tutkielmassa käsitellään videontiedonhaun käyttöliittymiä. Tavoitteena on jäsentää videoiden automaattisen sisällönkuvailun ja visualisoinnin ongelmakenttiä koskevaa kirjallisuutta. Menetelmien toimivuutta pohditaan Yleisradion TV-uutislähetyksen kohdalla.

Tulosten perusteella voidaan esittää, että videoiden ajallisen rakenteen jäsentäminen – esimerkiksi uutisjuttujen tunnistaminen – on realistisesti toteutettavissa automaattisin menetelmin nykytietämyksen valossa, ja sitä on käsitelty kirjallisuudessa kattavasti. Sen sijaan sisällön tunnistaminen semanttisella tasolla – esimerkiksi havaittujen kasvojen nimeäminen – on edelleen ratkaisematon ongelma muuten kuin rajoitetuissa konteksteissa. Nykyisten indeksointimenetelmien suorituskykyä voitaisiin parantaa integroimalla kuvaan ja ääneen perustuvia menetelmiä entistä tiukemmin. Videotiedonhakua varten on kehitelty erilaisia hakuvälineitä, mutta ne ovat rajoittuneita eivätkä hyödynnä kuin osaa mahdollisista visualisointimenetelmistä. Hakujärjestelmää suunniteltaessa pitäisi ottaa huomioon indeksointimenetelmien rajoitukset.

Sisältö

1 JOHDANTO

1.1 Videot tiedonlähteinä

Videoiden automaattista hahmopohjaista indeksointia käsittelevässä kirjallisuudessa korostetaan, että digitaaliseen muotoon tallennettua videomateriaalia tuotetaan jatkuvasti lisää. Tällä perustellaan videomateriaalin hallitsemiseen, käsittelemiseen ja kuvailemiseen tarvittavien automaattisten menetelmien kehittämisen tarkoituksenmukaisuutta. [Ks. mm. Antani, Kasturi & Jain 2002; Bolle, Yeo & Yeung 1998; Brunelli, Mich & Modena 1999; Sheridan, Wechsler & Schäuble 1997; Xiong, Chung-Mong Lee & Ma 1997.] Käyttökohteita automaattiseen indeksointiin perustuville hakujärjestelmille löytyy useita: TV-uutisten urheilutoimittaja saattaa olla kiinnostunut jonkin jalkapallo-ottelun maalitilanteista, mutta ei haluaisi katsella koko ottelua läpi. Vastaavasti politiikkaan keskittyvä toimittaja saattaa olla kiinnostunut uutisjutuista, joissa esiintyy jokin nimetty ja tunnettu henkilö. Hakijalla saattaa olla valokuva jostain tunnetusta tapahtumasta, josta pitäisi löytää myös videomateriaalia. Sovelluskohteita voisivat olla myös valvontakameroiden kuvamateriaalin käsittely – lähinnä poikkeavien tilanteiden tunnistaminen – tai lääketieteelliset käyttötarkoitukset. Muita käyttötarkoituksia mainitsevat esimerkiksi Antani ja muut [2002, 945], Geisler, Marchionini, Nelson, Spinks ja Yang [2001, 58–59] sekä Yeo ja Yeung [1997, 44].

Tutkielmassa käsitellään digitaalisen videon automaattisia indeksointimenetelmiä ja videotiedonhaun käyttöliittymiä TV-uutisten ja -toimituksen muodostamassa viitekehyksessä. Vaikka TV-uutiset ja -dokumentit välittävät yhteiskunnallisesti relevanttia informaatiota ja ovat siten ilmeisiä tiedonlähteitä, tuo informaatio ei ole ollut yleensä tarpeeksi helposti saatavilla, sillä perinteiset videoarkistot eivät ole olleet miellyttäviä käyttää: Esimerkiksi Markkula [2002] mainitsee erääksi ongelmaksi videotiedonhakujärjestelmien käyttöä käsittelevän tutkimuksen esiraportissa, että Yleisradion TV-toimituksen videoarkistossa varsinaista visuaalista sisältöä ei ole dokumentoitu ollenkaan, vaikka sitä järjestelmän pitäisi käyttäjien tiedontarpeiden täyttämiseksi kuvailla. Toimittajat joutuvat pettymään, jos nauhoilla tilattu video ei vastaa odotuksia. Usein tarvitaan paljon ylimääräistä materiaalia, jotta varmistuttaisiin, että edes jotain käyttökelpoista löytyy. [Markkula 2002.] Lisäksi tiedonhaku kuvanauhurin kömpelöllä käyttöliittymällä selailemalla ei ole millään tavalla mielekäs vaihtoehto. Varsinkin äänimateriaalin hakeminen on koettu kunnollisten selausmenetelmien puuttuessa vaivalloiseksi.

1.2 Videoiden indeksoinnin tutkimus

Suoraviivainen tapa indeksoida videosisältöjä on kuvailla niitä tekstuaalisin termein [Idris & Panchanathan 1997, 146]. Tämänkaltaista lähestymistapaa indeksointiin kutsutaan käsitepohjaiseksi. Yksinkertaisimmillaan käsitepohjainen indeksointi on aina manuaalista, ja sen suorittavat ihmiset. Manuaalinen indeksointi on kuitenkin havaittu liian hitaaksi ja kalliiksi useimpiin tarkoituksiin. Lisäksi ihmiset tulkitsevat erityisesti visuaalisesta informaatiosta eri asioita. Markkula ja Sormunen [2000] sanomalehden digitaalista valokuva-arkistoa käsittelevässä artikkelissaan huomauttavat, että manuaalinen indeksointi on usein epäjohdonmukaista [mts. 17–19].

Manuaalisen lähestymistavan ongelmien ohittamiseksi on visuaalisen datan indeksointia lähestytty kuvananalyysin ja -ymmärtämisen tekniikoiden näkökulmasta. Alan tutkimuksessa on pyritty kehittämään automaattisia ja aiheriippumattomia tekniikoita, jotka mahdollistavat visuaalisen datan indeksoimisen ja hakemisen sisällön perusteella. [Idris & Panchanathan 1997, 146.] Viime vuosina tutkimuksessa on kiinnitetty huomiota videoiden automaattiseen indeksointiin [Brunelli et al. 1999, 79]. Hahmopohjaisessa ('content based') lähestymistavassa videodatasta poimittuja havaittavia piirteitä käytetään sisällön kuvailemiseen; näin mahdollistetaan videoiden hakeminen suoraan niiden sisällön perusteella. [Del Bimbo 1999, 1; Markkula & Sormunen 2000, 2.] Piirteistä voidaan edelleen pyrkiä johtamaan semanttisia käsitteitä. Hahmopohjainen indeksointi on aina automaattista, ja kun aineistoa on runsaasti, vain automaattinen indeksointi tulee kysymykseen. Kuitenkin muun muassa Markkula ja Sormunen [2000, 2] toteavat, että laajasta tutkimuksesta huolimatta, hahmopohjaiset menetelmät toimivat tällä hetkellä tehokkaasti vain videodatan yksinkertaisten matalan tason piirteiden tasolla. Semanttisella tasolla, yritettäessä nimetä ja luokitella objekteja, edistys on ollut huomattavasti hitaampaa. Automaattiseen indeksointiin ja videotiedonhakuun liittyy lukuisia ongelmia, joista osa on toistaiseksi täysin ratkaisemattomia. Automaattisten menetelmien kehittäminen varsinkin käsitteellisesti korkean tason sisällön indeksoimiseen on havaittu erittäin vaikeaksi tai jopa mahdottomaksi, vaikka monien tiedonhakuongelmien ratkaiseminen edellyttäisi juuri tällä tasolla toimivia järjestelmiä. Aihealue on kuitenkin aktiivisen tutkimuksen kohteena ja lähteitä löytyy runsaasti: Esimerkiksi Bolle ja muut [1998], Brunelli ja muut[1999], Del Bimbo [1999], Idris ja Panchanathan [1997] sekä Petković ja Jonker [2000] käsittelevät aihetta yleisluontoisena esityksenä. Kattavaa esitystä alan tutkimuksesta oppikirjana ei kuitenkaan ole saatavilla – varsinkaan suomeksi.

1.3 Tutkimuskysymykset ja jäsennys

Tutkielmassa käsitellään muun muassa konenäön ('computer vision'), hahmontunnistuksen, käyttöliittymien ja semiotiikan alojen kirjallisuuden avulla digitaalisten videoiden hahmopohjaisen sisällönkuvailun ja videontiedonhaun käyttöliittymien periaatteita. Keskeisimpänä tavoitteena on videoiden automaattisen sisällönkuvailun ja visualisoinnin ongelmakentän jäsentäminen ja alaan liittyvien osa-alueiden käsittely, kun niitä tarkastellaan TV-toimituksen näkökulmasta käsin [ks. Del Bimbo 1999, 13–15]. Tutkielman perustavaa laatua olevina kysymyksiä ovat: (1) mitkä ovat videoiden ja erityisesti TV-uutisten keskeisimmät ominaispiirteet indeksoinnin kannalta tarkasteltuna, (2) kuinka merkitys muodostuu havaittavien piirteiden pohjalta ja kuinka hahmopohjaiset indeksointimenetelmät pyrkivät mallintamaan tätä prosessia; lisäksi pyritään selvittämään (3) videoiden esittämisen ja videotiedonhaun käyttöliittymien periaatteita eli sitä, miten alkuperäinen sekventiaalinen videodata esitetään uudelleen tiedonhakuun paremmin sopivassa muodossa. Lopuksi näiden kysymysten pohjalta tuotettua tietämystä sovelletaan Yleisradion TV1:n uutislähetykseen.

Tutkielmaa ei ole tarkoitettu kattamaan digitaalisten videoiden indeksointi- ja hakujärjestelmiä koskevaa kirjallisuutta kokonaisuudessaan, vaan TV-uutiset muodostavat viitekehyksen aineiston tarkastelulle. Lähestymistapa on siinä mielessä käyttäjäkeskeinen, että kirjallisuus on jäsennetty tietyn potentiaalisen käyttäjäryhmän tiedonhakuongelmien pohjalta [ks. Markkula 2002]. Näkökulmasta ja jäsennyksen perusteista huolimatta tutkielma ei pyri varsinaisesti vastaamaan siihen, miten kokonainen indeksointi- ja hakujärjestelmä toimii tai millaisen järjestelmän TV-uutiset käytännössä vaatisivat, sillä se edellyttäisi näiden järjestelmien käyttäjien kattavampaa haastattelua. Tutkimuskohteena ei ole siis videoinformaation organisointi tai hallitseminen (näistä lisää ks. Prabhakaran 1997, 25–51) vaan indeksoinnin ja hakemisen periaatteet.

Tutkielma on jäsennetty seuraavalla tavalla: Television ja videon ominaispiirteitä mediana käsitellään luvussa 2. Automaattista semanttista päättelyä videodatasta lähestytään semiotiikan ja semanttisten mallien näkökulmasta luvuissa 2 ja 3. Indeksointitehtäviä, joita järjestelmän pitäisi tukea, käsitellään luvussa 4. Videotiedonhaun käyttöliittymiä käsitellään luvussa 5. Tutkielma on lisäksi osittain kvasikokeellinen, sillä Yleisradion TV1-kanavan illan pääuutislähetystä analysoidaan indeksointitehtävien suorittamisen ja teorian havainnollistamisen näkökulmasta luvussa 6. Teoreettisella tasolla tarkastellaan, kuinka automaattiset indeksointimenetelmät toimisivat TV1:n pääuutislähetyksessä. Luvussa 7 esitetään johtopäätöksiä

ja ehdotuksia jatkotutkimusta varten.

2 VIDEO JA TV-UUTISET

Tämän luvun semioottisella lähestymistavalla TV-uutisiin pyritään pohjustamaan luvussa 3 käsiteltävää automaattista semantiikan johtamista videosisällöstä. Automaattisten indeksointijärjestelmän toiminnan käsitteellistämiseksi semanttisella tasolla on välttämätöntä ymmärtää, kuinka videot tuottavat merkityksiä ihmisille. Semiotiikkaa lähestytään pääosin Kressin ja van Leeuwenin [1999] visuaalista kielioppia käsittelevän teoksen sekä Seiterin [1992], Ellisin [1992] ja Cornerin [1995] elokuvia ja televisiota käsittelevien teosten avulla. Tätä ennen kuitenkin määritellään, mitä tutkielmassa tarkoitetaan videolla ja mitkä ovat videon keskeisiä ominaisuuksia.

2.1 Videon ominaispiirteet

Perustavaa laatua olevilta ominaisuuksiltaan video on multimodaalinen ja sekventiaalinen yksittäisistä kuvista muodostuva kuvavirta. Videokuvan ajallinen ulottuvuus syntyy esitettäessä yksittäisiä kuvia peräkkäin; tästä muodostuvaa liikkeen tuntua tukee kuvasekvenssiin liittyvä ääniraita. Digitaalinen video on periaatteessa mikä tahansa elektroninen digitaalisessa muodossa oleva kuvavirta. Jokaisella viestintävälineellä on ominaispiirteitä, jotka rajoittavat ja mahdollistavat keinoja, joilla asioita voidaan esittää ja informaatiota välittää. Videota mediamuotona luonnehtivat seuraavat keskeiset ominaisuudet: (1) multimodaalisuus, (2) paikallisuus, (3) ajallisuus, (4) sekventiaalisuus ja (5) katkonaisuus.

Videoiden multimodaalisuus tarkoittaa, että ne muodostuvat useammasta rinnakkaisesta kommunikaation ja informaation kanavasta, joita ovat kuva, grafiikka (esim. kirjoitettu kieli, logot yms.), puhe, musiikki ja ääni [Grosky 1997, 74; Seiter 1992, 43; ks. Prabhakaran 1997, 7]. Video välittää informaatiota yksittäisten kuvien paikkasidonnaisen ('spatial') sisällön ja kuvasekvenssien tuottaman ajallisen ('temporal') ulottuvuuden avulla. Paikkasidonnainen sisältö muodostuu objekteista ja niiden sommittelusta videosekvenssin yksittäisessä kuvassa, ja ajallinen sisältö muodostuu videon alisekvensseissä kuten otoksissa ja kohtauksissa esiintyvistä paikkasidonnaisen sisällön muutoksista. Prabhakaran [1997, 8–9] määrittelee

videot kolmiulotteisiksi mediaobjekteiksi: teksti ja ääni ovat jatkuvia ja yksiulotteisia medioita, ja kuva on sommitelma kahdessa suunnassa paikallistettavia paikkasidonnaisia alueita; videossa yhdistyvät nämä kaksi ulottuvuutta ja muodostavat kolmannen. Video on ajallinen media, jossa värin, tekstuurin, muodon ('shape') ja liikkeen muutokset useamman kehyksen alueella merkitsevät enemmän kuin yksittäisten kehysten sisältö: sekvenssoiminen luo semanttisia sisältöjä, jotka eivät välttämättä ole tulkittavissa yksittäisistä kehyksistä [Del Bimbo 1999, 8; Petković & Jonker 2000; Lee & Smeaton 1999, 1].

Videon sekventiaalisuus tarkoittaa, että kaksi tai useampi samanaikaisesti esiintyvää prosessia on käytännössä esitettävä vuorotellen [Rui, Huang & Mehrotra 1999, 359, 362]. Tästä päädytään videon ehkä keskeisimpään rakenteelliseen ominaispiirteeseen, joka ilmenee katkonaisuutena fyysisellä tasolla siirryttäessä kehyksestä kehykseen ja otoksesta otokseen. Ihminen ei havaitse sekvenssin yksittäisten kuvien välistä katkonaisuutta – jos kuvia näytetään tarpeeksi nopeasti peräkkäin – mutta otosten välinen epäjatkuvuus on havaittavissa, vaikka se pyritäänkin peittämään. [Ks. Bolle et al. 1998.] Fyysisellä tasolla tarkasteltaessa video on kuvasekvenssi, joka muodostuu joukosta alisekvenssejä, joista osa on ulkonäöltään yhtenäisiä. Videosisältö eli sekvensseissä esiintyvät objektit ja tapahtumat välittyvät tämän ajallisesti katkonaisen rakenteen läpi. Videoiden indeksoinnin kannalta keskeistä on ottaa huomioon, että esitystavan fyysinen katkonaisuus ei ole sama asia kuin semanttinen katkonaisuus merkityksessä. Ihanteellisesti videotiedonhaku voi kohdistua sekä ajallisiin alisekvensseihin, joita ovat esimerkiksi otokset ja otosryhmät, että yhden tai useamman sekvenssin alueelle ulottuvaan videosisältöön, kuten tiettyihin objekteihin [Del Bimbo 1999, 8]. Videon ajalliseen rakenteeseen kuuluvat elementit voidaan esittää seuraavalla hierarkkisella tavalla:

- *Kehys* ('frame') eli kuvavirran yksittäinen kuva on informaation perusyksikkö videoissa samalla tavalla kuin sanat ovat tekstidokumenttien perusyksikköjä. Ajalliselta pituudeltaan yksi kehys on 1/25 (PAL) tai 1/30 (NTSC) sekuntia [Del Bimbo 1999, 4, 8; Petković & Jonker 2000; Prabhakaran 1997, 8]. Tiedonhakijat eivät ole yleensä kiinnostuneita videoiden yksittäisistä kehyksistä niiden suuren määrän vuoksi [Del Bimbo 1999, 9; Petković & Jonker 2000]. Avainkehys on yksittäinen kehys, joka on valittu edustamaan kokonaisen otoksen silmiinpistävää sisältöä [Rui et al. 1999, 359].
- *Otos* ('shot') on tauotta tallennettu sekvenssi peräkkäisiä kehyksiä. Perinteisesti otoksella on tarkoitettu sitä aikaa, joka kuluu katkeamattoman kameran toiminnan aikana, kun yksittäinen kamera nauhoittaa ja lopettaa nauhoittamisen. Otos edustaa siis jatkuvaa toimintaa ajassa ja paikassa. Yksittäisen kehyksen jälkeen otos on videon yksinkertaisin yksikkö; se on lyhin minimaalinen segmentti ja fyysinen olio ('entity'). Kukin videodo-

kumentti muodostuu useammasta otoksesta (eli alisekvenssistä), jotka ovat sisällöltään yhdennäköisiä alun ja lopun välillä ja joiden välillä on jatkuvuutta merkityksessä. Otosten loppukohdat eli siirtymät toisiin otoksiin voivat olla leikkauksia tai asteittaisia siirtymätehosteita. [Antani et al. 2002, 955; Apers, Blanken & Houtsma 1997, 172; Bolle et al. 1998; Brunelli et al. 1999, 81; Del Bimbo 1999, 10; Idris & Panchanathan 1997, 154; Lienhart, Pfeiffer & Effelsberg 1997, 55; Petković & Jonker 2000; Rui et al. 1999, 359; Seiter 1992, 45.] Otokset videon perusyksikköinä eli segmentteinä ovat merkityksellisiä ja ihmisen havaittavissa olevia. Otos on elokuvallinen rakennuspalikka ja kantaa vain minimaalisesti semanttista informaatiota: se on kuin lause – sillä on semanttista merkitystä, vaikka ei juurikaan kontekstista irrallaan. [Bolle et al. 1998; Del Bimbo 1999, 10; Xiong et al. 1997, 51.] Bolle et al. [1998] esittävät, että videotiedonhaussa riittää, jos otos on pienin haettavissa oleva yksikkö eli segmentti.

- *Ryhmä* on välittävä olio eli silta fyysisten otosten ja semanttisten kohtausten välillä. Ryhmät muodostuvat ajallisesti lähekkäisistä ja visuaalisesti samankaltaisista otoksista. [Rui et al. 1999, 359]. Ryhmät eivät perustu semanttiseen vastaavuuteen otosten välillä vaan ainoastaan visuaaliseen samankaltaisuuteen, joka ymmärretään semanttisella tasolla vasta kohtauksissa.
- *Kohtaus* ('scene') on ryhmä ajallisesti peräkkäisiä otoksia, joita yhdistävät ominaisuudet ajassa, paikassa ja toiminnassa sekä semanttisessa merkityksessä [Apers et al. 1997, 172; Del Bimbo 1999, 10; Bolle et al. 1998; Kender & Yeo 1998, 3; Lienhart et al. 1997, 57; Rui et al. 1999, 359]. Kunnolla leikattu video luo katsojalle tunteen merkityksen jatkuvuudesta otosten välillä, mikä ylittää varsinaisen esityksen epäjatkuvuuden eli leikkaukset, kuvakulmien vaihtelun ja vastaavat. Katsoja ymmärtää jatkuvuuden eli sen, että tietyt otokset kuuluvat samaan kohtaukseen joko tietoisesti tai tiedostamatta. [Bolle et al. 1998.] Vaikka otos on videon rakennuspalikka, kohtaukset välittävät videon semanttisen merkityksen [Rui et al. 1999, 359]. Kohtaukset liittyvät tarinoihin, ja ne voivat olla dynaamisia ja staattisia [Del Bimbo 1999, 10].
- *Jakso* ('episode') on sarja otoksia, jotka liittyvät toisiinsa erityisten otostyyppien sarjoina. Esimerkiksi uutislähetyksessä ankkurin juontoa seuraa uutiskatsaus, jota seuraa toimittajan osuus ja niin edelleen. [Del Bimbo 1999, 10]. Jakson otokset liittyvät toisiinsa semantiikkansa puolesta. Jaksot eroavat kohtauksista siinä, että jakson otoksien välillä ei välttämättä ole yhtenäisyyttä paikassa, ajassa ja toiminnassa [ks. Del Bimbo 1999, 10].
- *Tarina* on kokonainen ryhmä kohtauksia, jotka ovat kytkeytyneet toisiinsa merkityksessä: esimerkiksi kokonainen uutisjuttu.

Näiden lisäksi voidaan vielä ottaa huomioon:

- *Leike* ('clip') on kehyssarja, jolla on semanttista merkitystä. Leike on saatettu leikata mistä tahansa kohdasta kuvavirtaa ja minkä tahansa mittaisena. [Del Bimbo 1999, 10.] Leikkeet eivät siis ole sama asia kuin otokset tai kohtaukset, jotka on rakennettu kuvavirtaan tuotannon yhteydessä “luonnostaan”.

Ajallisen rakenteen lisäksi videoita voidaan tarkastella semanttisen sisällön tasolla, mikä on indeksoinnin kannalta huomattavasti haastavampaa kuin ajallisen rakenteen tunnistaminen. Videon sisältöä tarkasteltaessa voidaan ottaa huomioon Del Bimbon [1999, 10] mukaan:

- Elokuvalliset ominaisuudet, joihin kuuluvat näkökentän ('viewfield') leveys, näkökenttien määrä otoksessa, valaistus ('illumination'), värit ja muut vastaavat ominaisuudet.
- Kameran liike otoksessa, joka on tärkeä tekijä analysoitaessa esimerkiksi ohjaajan tyyliä.
- Äänen ominaisuudet, jotka ovat avuksi erotettaessa esimerkiksi dialogista koostuvat kohtaukset muista. Ääntä voidaan käyttää myös kuvaraidan sisällön semanttisessa päättelyssä.
- Objektin jatkuva läsnäolo ja sen liikkuminen.
- Objektien väliset suhteet ('situation').
- Kohtaukset ja tarinat videossa, jotka rakennetaan tietyillä tavoilla. Esimerkiksi keskustelua sisältävissä kohtauksissa käytetään otos–palautus–otos-tekniikkaa ('shot-reverse-shot').
- Värin ja liikkeen semantiikka. [Del Bimbo 1999, 10.]

Del Bimbon [1999, 10–11] mukaan yksittäisiä kehyksiä voidaan tarkastella myös paikkaa koskevan sisällön tasolla ottamalla huomioon

- valaistusolosuhteet
- havaittavat ominaisuudet (kuten värit, tekstuurit ja muodot)
- havaittavien ominaisuuksien ryppäyttämisen alueiksi
- objektien paikallistamisen ja tunnistamisen. [Del Bimbo 1999, 10–11.]

Edellä esitettyihin videoiden ominaisuuksiin ja niiden automaattiseen indeksointiin palataan seuraavissa luvuissa. Tätä ennen kuitenkin käsitellään visuaalista informaatiota ja merkityksiä semiotiikan ja taidehistorian teorioiden näkökulmasta. Tavoitteena on jäsentää merkityksen muodostuminen tasoihin, joilla indeksointimenetelmien toimintaa myöhemmissä luvuissa tullaan tutkimaan. Tämän luvun lopussa tarkastellaan vielä TV-uutisia ja sen konventioita erityisenä videosisällön muotona.

2.2 Representaatio ja merkityksen tasot

Seiterin [1992, 31] sekä Grossbergin ja muiden [1998, 128] mukaan semiotiikka on tutkimusala, joka tutkii merkitysjärjestelmien luonnetta, kaikkia kommunikoimiseen käytettäviä merkkejä ja niiden käyttöä ohjaavia sääntöjä sekä sitä, kuinka merkitys luodaan. Semioottinen tutkimus alkaa merkin eli minkä tahansa merkitysjärjestelmän pienimmän merkityksellisen alkeisyksikön tunnistamisella. [Grosberg et al. 1998, 128; Seiter 1992, 31.] Semiotiikan näkemyksessä merkki muodostuu (1) merkitsijästä eli merkin ulkoisesta materiaalisesta muodosta, esimerkiksi kuvasta, semioottisesti ilmaistusta objektista, äänestä tai väristä, ja (2) merkitystä eli merkityksellisestä käsitteestä, jonka merkitsijä tuo esille [Grosberg et al. 1998, 132–134; Seiter 1992, 33, 35]. Representaatio [1] on merkin tekemisen prosessi, jossa jonkin havaittavan muodon avulla tuodaan esille jostakin konkreettisesta objektista, tapahtumasta, abstraktiosta tai niitä edustavasta semioottisesta esityksestä eli edeltävästä representaatiosta metonymisesti poimitut tarkoituksenmukaiset piirteet [ks. Kress & van Leeuwen 1996, 6]. Kressille ja van Leeuwenille [1996, 7] merkin tekeminen on kahden askeleen metaforinen prosessi, jossa analogian avulla yhdistellään käsitteellisiä luokkia tai konkretiasta erotettuja osa-alueita ja tuotetaan näistä uusi esitys ominaisuuksia siirtämällä: x on kuin y ja y on kuin z – ympyrä paperilla on kuin rengas ja rengas on kuin auto. Merkillä ei tämän tutkielman yhteydessä tarkoiteta vain symbolisia kirjoitusmerkkejä ja sanoja vaan kaikkia visuaalisia, auditiivisia tai matemaattisia esityksiä, joilla tarkoituksenmukaisesti esitetään jotain. Representaation ideana on, että jotain mikä on ollut, tuodaan uudelleen esille jossakin toisessa muodossa; representaatio on esitys, joka edustaa jotain muuta. Tässä tutkielmassa ei ole kuitenkaan nähty tarpeelliseksi korostaa representaation luonnetta uudelleen esittämisenä: on selvää, että kuvan objektin esittäminen esimerkiksi joukolla matemaattisia vektoreita on oleellisesti jotain muuta kuin alkuperäinen esitys. Näin ollen kirjallisuudessa käytetty 'represent' on käännetty yksinkertaisesti esittämi-

1 Semiotiikassa käytettävällä representaation käsitteellä korostetaan, että mikään asia ei ole luonnostaan sellainen kuin miksi se on esitetty, vaan asiat merkitsevät jotain ainoastaan representaation eli uudelleen esityksen merkityksellistäminä.

seksi tai edustamiseksi.

Kukin merkki merkityksellistyy kahdella tasolla. Denotaatiolla tarkoitetaan merkityksellistämisen ensimmäistä tasoa ('the first order of signification'), jolla merkitsijä on kuva itsessään ja merkitty on se idea tai käsite, jota kuva esittää merkin puitteissa (esimerkiksi “a picture of”). Konnotaatio on merkityksellistävän järjestelmän toinen taso, joka käyttää ensimmäistä kokonaista merkkiä merkitsijänä ja liittää siihen ylimääräisen merkityksen, toisen merkityn, joka köyhdyttää merkityksen ensimmäisen tason (eli denotaation) merkin merkityspotentiaalin. Näin ollen, jos merkityksen ensimmäisellä tasolla kuvaa tarkastellaan muun muassa kuvakulman, värien, objektien muotojen ja niiden koon, valaistuksen ja sommitellun perusteella, toisella tasolla (eli konnotaatiossa) kuvailuun riittävät esimerkiksi “jalo”, “romanttinen”, “isänmaallinen”. [Seiter 1992, 39.] Konnotaatiolla tarkoitetaan kulttuurisessa merkitysjärjestelmässä olevia lisämerkityksiä; henkilökohtaiset lisämerkitykset ja mielikuvat ovat assosiaatioita. Kulttuurisesti toimivat konnotaatiot köyhdyttävät havaittavista piirteistä ja käsitteistä muodostuviin merkkeihin niiden yleisesti hyväksytyt merkitykset. Seiter [1992, 39] antaa esimerkkinä yksinkertaisesta denotaatiosta häivytyksen ('fade to black'), jonka merkitsijänä on kuvan asteittainen häipyminen ('disappearance') ruudulta ja merkittynä musta ruutu. Häivytys on merkkinä konventionalisoitu elokuvissa ja televisiossa, joten se toimii myös konnotatiivisesti: jos häivytys on merkitsijä, merkitty tarkoittaa kohtauksen tai ohjelman loppua. [Seiter 1992, 39.] [2]

Semioottista lähestymistapaa voidaan soveltaa myös esimerkiksi uutisstudion merkityksellistymisen ilmaisemiseksi. Jokin uutisstudioon liittyvä objekti tai alue voidaan esittää verbaalilla kielellä ilmaistulla käsitteellä tai vaikkapa otoksesta valitulla yksittäisellä avainkuvalla; vaihtoehtoisesti esitys voi muodostua matemaattisesti ilmaistuista piirteistä (usein vektoreista) ja niihin liittyvistä eksplisiittisesti ilmaistuista malleista, joissa määritellään objektien väliset suhteet ja niiden varaamat alueet ynnä muut ominaisuudet. Nämä tavat tuovat esiin esittämänsä kohteen tarkoitukseen sopivalla tavalla. Esimerkiksi uutisankkuri muodostuu tietynlaisesta hahmosta, alueista, väreistä ja pinnoista, jotka muodostavat merkitsijän, ja näiden piirteiden esiintuoma henkilö on merkitty, kohde, jota merkitsijä edustaa merkin sisällä. Uutisankkuri sijaitsee ruudun keskiosassa uutisikkunan oikealla puolella. Nämä kaksi silmiinpistävää objektia (yhdessä parin muun objektin kanssa) muodostavat uutisstudion eli kontekstia osoittavan tilan, kun aihetta tarkastellaan indeksointijärjestelmän näkökulmasta (ks. luku 6). Inhimillisen katsojan kannalta uutisstudio köyhdyttää uutisankkurin ja uutisikkunan merkkien visuaaliset piirteet, niiden paikkasidonnaiset sijainnit ja niiden merkitykset kon-

2 Seiter [1992, 39–40] viittaa Zettlin [1984, 596] televisiotuotannon käsikirjaan, jonka mukaan häivytys pitäisi liittää jokaisen ohjelman loppuun ja ennen jokaista mainoskatkoa.

notaation kautta: katsojan ei tarvitse kuin nähdä ne ja heti hänen mieleensä tulee uutisstudio.

Taidehistoriassa käytetään merkityksen muodostumisen jäsentämiseksi hieman erilaista käsitteistöä kuin semiotiikassa. Rasmussenin [1999, 177] mukaan *Erwin Panofsky* erotti semanttisessa merkityksessä kolme tasoa: esi-ikonografisen, ikonografisen ja ikonologisen. Esi-ikonografisella tasolla merkitysten käsitetään koskevan fyysisiä objekteja ja tapahtumia, mitkä voivat olla faktuaalisia tai ilmaisullisia: se, mitä kuva esittää, on yleensä itsestään selvää eli faktuaalista, mutta se, mitä kuvalla halutaan ilmaista eli mistä se kertoo, on subjektiivista. Itsestään selvästä faktuaalisesta merkityksestä käytetään termiä “ofness” (esim. “This is a picture is of a race car”) ja ilmaisullisesta “aboutness” (esim. “This is a picture about car racing”). “Ofness” on hyvin lähellä semiotiikan käsitettä denotaatio, kun taas “aboutness” lähestyy käsitettä konnotaatio [ks. esim. Fiske 1992, 113–115]. Esi-ikonografisella tasolla “ofness” tarkoittaisi kuvattujen asioiden yleistä esitystä ja “aboutness” esimerkiksi kuvan tunnelmaa ja siihen liittyviä henkilökohtaisia assosiaatioita [ks. Rasmussen 1999, 177]. Ikonografisella tasolla “ofness” tarkoittaisi asioiden nimeämistä faktuaalisesti (eli denotatiivisesti) ja “aboutness” tarkoittaisi kuvan tuottamia ilmaisullisia (eli konnotatiivisia) merkityksiä [ks. Rasmussen 1999, 178]. Näin ollen esi-ikonografisella tasolla kuvasta voidaan tunnistaa objekti ja sille faktuaalinen merkitys (“ofness”), esimerkiksi “auto”, jolle voidaan tunnistaa ilmaisullinen eli assosiatiivinen merkitys (“aboutness”), kuten “uusi”. Ikonografisella tasolla kuvan objekti voidaan nimetä esimerkiksi tietyn automerkin tietyksi malliksi: nyt objekti “auto” (esi-ikonografisella tasolla) onkin “Rolls Royce” (ikonografisella tasolla). Jos ikonografisella tasolla “Rolls Royce” on kuvan faktuaalinen merkitys (“ofness”), siihen voidaan liittää ilmaisullisesti esimerkiksi rahaan ja valtaan liittyviä konnotaatioita (“aboutness”). Ikonografisella tasolla vaaditaan kulttuurin tuntemusta, jotta asiat voidaan nimetä. Mitä korkeammalle merkityksen abstraktiossa mennään, sitä enemmän merkitys on kulttuurisidonnaisempi ja perustuu kuvan ulkopuolisiin merkityksiin. Ikonologisella tasolla liikutaan vapauden kaltaisten käsitteiden piirissä. Ikonologinen taso on lähellä semiotiikan käsitystä myytistä [ks. Fiske 1992, 158–162].

Markkula, Tico, Sepponen, Nirkkonen ja Sormunen [2001, 9–11] luettelevat kriteereitä, joita toimittajat käyttävät samankaltaisuuden arvioimiseen, järjestettynä edellä käsiteltyihin kolmeen abstraktion tasoon. Ensimmäisellä eli esi-ikonografisella tasolla kuvia lähestytään valoisuuden ja kontrastin, värien (esim. punainen alue mustalla taustalla) ja erilaisten alueiden sommittelun ja mittasuhteiden perusteella. Vuoden- ja vuorokaudenajat sekä jotkin paikat ovat läheisessä suhteessa matalan tason piirteisiin kuten väriin ja kirkkauteen. Myös rajaus, kuvausetäisyys ja kuvakulma, jotka koskevat kuvissa näkyviä objekteja, mutta eivät vaadi semanttista päättelyä, kuuluvat tälle tasolle. Toisella eli ikonografisella tasolla tar-

vitaan semanttista päättelyä, jotta havaittavista piirteistä voidaan johtaa ilmaistavia merkityksiä. Ikonografisella tasolla kuvasta tunnistetaan objektien tyyppejä (esim. junat), nimettyjä objekteja (esim. henkilöiden nimiä), toimia ja tapahtumia sekä nimettyjä paikkoja. [Markkula et al. 2001, 9–10.] Kyseessä on se denotatiivinen taso, jonka konnotaatio köyhdyttää lisämerkityksellään [ks. Seiter 1992, 39]. Kolmas taso käsittää abstraktit ideat (esim. väkivallan), tunteet ja symboliset merkitykset, jotka köyhdyttävät edellisen tason denotatiivisia merkityksiä. Näistä tasoista ensimmäinen on hahmopohjaisten indeksointialgoritmien ulottuvissa. Toisella tasolla ilmenee lukuisia eriävien tulkintojen aiheuttamia ongelmia. Perinteiset objektien tunnistusmenetelmät eivät pysty tunnistamaan yleisiä objektityyppejä ja luokittelemaan niitä, vaikka tiettyjen objektien tunnistaminen on kyllä mahdollista. [Markkula et al. 2001, 9–10.] Ikonologisella tasolla kuvaa ei voida indeksoida johdonmukaisesti edes manuaalisesti ja automaattinen sisällönkuvailu on täysin mahdotonta [Rasmussen 1997, 178]. Näihin huomioihin palataan myöhemmin.

2.3 Visuaalinen koodi

Semiotiikassa analyysimetodina käytetään usein strukturalismia, jonka mukaan jokainen merkki saa merkityksensä suhteistaan merkkijärjestelmän toisiin merkkeihin [Seiter 1999, 32]. Koodiksi kutsutaan systemaattista merkkien rakennetta ja merkkien yhdistelyn sääntöjoukkoa [Grossberg et al. 1998, 129; Seiter 1992, 33]. Sääntöjen säätelemää merkkien yhdistelmää, jossa merkit on järjestelty määrätyssä järjestyksessä, kutsutaan syntagmaksi. Merkit valitaan yhdistelmiin paradigmoista, jotka ovat luokkia samankaltaisia merkkejä, jotka voidaan korvata toisillaan syntagmassa. Syntagman merkitys juontuu ('derive') osittain toisten mahdollisten paradigmaattisten valintojen poissaolosta. [Seiter 1992, 46.] Paradigmoja ja niihin kuuluvien merkkien välisiä suhteita sekä syntagmojen tuottamista säädellään koodien avulla. Koodit voidaan jakaa verbaalin kirjoitusjärjestelmän kaltaisiin (1) symbolisiin koodeihin, joiden merkit edustavat kohdettaan sopimuksen perusteella, ja (2) kuvallisiin ('pictorial') koodeihin, jotka perustuvat ikonisiin ja indeksisiin merkkeihin [Seiter 1992, 33, 35]. Jos denotaation ja konnotaation käsitteet jäsentävät merkkien tuottamat merkitykset tasoihin, symbolinen, ikoninen ja indeksinen puolestaan kuvaavat merkin suhdetta viittaamaansa kohteeseen: Ikoninen merkki muistuttaa rakenteellisesti edustamaansa kohdetta. Indeksiset merkit sisältävät eksistentiaalisen linkin merkitsijän ja viitteen välillä, joiden yhteisestä ('joint') läsnäolosta merkki on riippuvainen jossain kohden aikaa. Riippuvuus voi olla esimerkiksi materiaalinen yhteys merkitsijän ja merkityn välillä. Merkkien suhdetta viittaamiinsa kohteisiin

kuvaavat kategoriat eivät ole toisensa poissulkevia, vaan merkit voivat olla ikonisia, indeksisiä ja symbolisia jopa samaan aikaan, ja TV:n kuva onkin näitä kaikkia. [Seiter 1992, 35–36.]

Seiter [1992, 46] mainitsee, että elokuvien leikkauksessa tehtyjä ratkaisuja voidaan tarkastella koodina. Otokset eli “minimaaliset segmentit” käsitetään paradigmoiksi, joista elokuva (eli semiotiikan näkökulmasta syntagma) muodostetaan tiettyjä yhdistelysääntöjä noudattamalla. [Ks. Seiter 1992, 46.] Voidaanko visuaalista informaatiota kuitenkaan tarkastella koodina siinä missä verbaalia kieltä? Perinteisen näkökulman mukaan tämä on ongelmallista. Kielessä pientä joukkoa eroteltavissa olevia yksiköitä – kirjaimia ja ääniä (mm. foneemeja) – käytetään luomaan monimutkaisempia merkityksellistämisiä ('significations'): sanoja, lauseita ja kappaleita [Seiter 1992, 42–43]. Seiterin [1992, 42–43, 45] mukaan televisioilmaisu ei kuitenkaan ole sopivasti rikottavissa erillisiin elementteihin tai merkityksen rakennuspalikoihin, koska sillä ole selvää aakkostoa. Indeksiset ja ikoniset merkit, kuten kuvat, eivät ole supistettavissa ('reducible') enää pienempiin yksiköihin, koska ne ovat jo itsessään eräänlaisia tekstejä eli yhdistelmiä merkkejä. Ikonisia merkkejä säännöstelevä koodi on vain heikko verrattuna kieliä säännösteleviin kielioppeihin. [Seiter 1992, 42–43, 45.] Toisin sanottuna indeksisissä ja ikonisissa koodeissa merkkien alkeisyksiköt eivät ole yhtä helposti erotettavissa toisistaan kuin verbaalin kielen kaltaisessa symbolisessa koodissa. Koodi edellyttää, että havaittavista piirteistä on muodostettavissa paradigmoja, joiden välille on tunnistettavissa rakenne. Kress ja van Leeuwen [1996, 16, 23] mainitsevatkin, että semiotiikassa valokuva on perinteisesti nähty viestinä ilman koodia, mihin myös Hall [1999, 143] viittaa kertoessaan, että kuva "antaa mahdollisuuden monille merkityksille, mutta sillä ei ole yhtä, oikeaa merkitystä” ja että kuvissa merkitys “'kelluu' eikä sitä voida kiinnittää lopullisesti”. Hall [1999, 143] lisää kuitenkin, että “representaation käytäntöjen tehtävänä on 'kiinnittää' merkityksiä jollakin tietyllä tavalla". Semiotiikan perinteisessä näkemyksessä kuvatekstin on katsottu ankkuroivan merkityksen kuvaan, jolloin merkityksen tuottaa ja kiinnittää kaksi diskurssia tai kanavaa: kuva ja teksti [Hall 1999, 144]. (Luvussa 4 käsitellään kahden viestintäkanavan välittämän informaation yhdistämistä kasvojen tunnistamiseksi videokuvasta.)

Vaikka kirjoitettua verbaalia tekstiä ja kuvia on tarkasteltu erillisinä koodityyppeinä, tavanomaisesta näkemyksestä poiketen, Kress ja van Leeuwen pitävät niitä kumpaakin visuaalisina koodeina: symbolisissa koodeissa merkit ovat vain tarkemmin jäsennetty ja kontrolloitu kuin kuvallisissa. [Kress & van Leeuwen 1996, 3–4, 40–44.] Kressin ja Leeuwenin [1996, 15, 23, 32] mukaan kaikki visuaalinen informaatio – mukaan lukien valokuvat ja piirrokset – on rakenteista ja koodattua. Kaikissa visuaalisissa koodeissa merkkien ulkoiset muodot, ääriviivat ja muut vastaavat ominaisuudet toimivat merkitsijöinä tietyille merkityille. Esimerkiksi siinä missä kielissä on mahdollista valita sanaluokista ja semanttisista rakenteista,

visuaalisessa viestinnässä paradigmaattiset valinnat tehdään esimerkiksi erilaisten värien ja sommitelmien välillä [Kress & van Leeuwen 1996, 2]. Semioottiset järjestelmät tarjoavat joukon ('array') valintoja niistä eri tavoista, joilla objektit voidaan esittää ja ne voivat olla suhteessa toisiinsa [Kress & van Leeuwen 1996, 40]. Visuaalinen rakenteistaminen siis luo merkityksellisiä väittämiä erityisen visuaalisen syntaksin avulla eikä vain toista todellisuutta [Kress & van Leeuwen 1996, 45]. Koodi saattaa vaikuttaa läpinäkyvältä, jos se tulkitaan huomaamatta erilaisten konventioiden ja lukutottumusten avulla. Sosiaalisiin ryhmiin ja institutionaalisiin konteksteihin liittyy erilaisia konventioita, abstrakteja periaatteita ja käytänteitä, joiden avulla tekstit koodataan [Kress & van Leeuwen 1996, 170]. Esimerkiksi TV-uutisissa esittämisen käytäntöjä normalisoimalla pyritään erilaisten merkitysten määrän rajoittamiseen, jotta halutunkaltaisen informaation välittäminen olisi mahdollista.

Kress ja van Leeuwen [1996, 46–48] käsittelevät formaalin taideteorian [ks. Arnheim 1974 ja 1982] kehittämiä keinoja, joilla voidaan tunnistaa (näennäisen) rakenteettomien kuvien objekteja ja elementtejä hyödyntämällä havainnointipsykologiaa ('psychology of perception'). Objektit havaitaan kuvista erillisinä olioina ('entity'), jotka ovat eriasteisesti silmiinpistäviä ('salient') niiden koon, muotojen, värien ja muiden vastaavien erojen takia. Esimerkiksi jotkin objektit voivat erottua muista niiden silhuettien ja valonlähteen välisen tonaalisen kontrastin takia. Samalla periaatteella, jolla taiteilijat supistavat havaittavan maailman yksinkertaisiin geometrisiin muotoihin, muodot voidaan havaita yksinkertaisten visuaalisten kaavioiden pohjalta niiden silmiinpistävien ominaisuuksien perusteella. [Kress & van Leeuwen 1996, 47.] Esimerkiksi lapset oppivat piirtämään kehittämällä repertuaarin perusmuodoista, jotka sitten asteittain sulautetaan ('fuse') toisiinsa [Arnheim 1974, Kress & van Leeuwenin 1996, 47 mukaan].

Kress ja van Leeuwen [1999, 79] jakavat visuaaliset rakenteet kerronnallisiin ja käsitteellisiin. Kerronnalliset rakenteet esittävät toimia ja tapahtumia eli paikkasidonnaisten asetelmien muutoksia. Käsitteelliset rakenteet esittävät objektit yleistettyinä luokan tai merkityksen mukaan, enemmän tai vähemmän vakaina ja ajattomina olioina. [Kress & van Leeuwen 1996, 79.] [3] Kuvallisissa koodeissa vektorit vastaavat verbaalin kielen teonsanoja. Paikan prepositioita (eli ”edessä”, ”takana” jne.) vastaavat ilmaisut toteutetaan kuvallisissa koodeissa formaaleilla piirteillä, jotka luovat kontrastin etu- ja taka-alan välille [Kress & van Leeuwen 1996, 44]. Kerronnallisilla rakenteilla on aina vektori mutta käsitteellisillä ei koskaan. Konteksti osoittaa millaista toimintaa vektori esittää. Esimerkiksi tie, joka kulkee viistoon kuvan avaruuden yli on vektori ja auto, joka ajaa tietä pitkin on toimija ('actor') ajamisen tapahtu-

3 Kress ja van Leeuwen [1996] käyttävät objekteista ja elementeistä termiä “osanottaja” ('participant') ja tapahtumista termiä “prosessi” [ks. mts. 46–47]. Yksinkertaisuuden vuoksi tutkielmassa käytetään edelleen termejä objekti ja tapahtuma.

massa (eli prosessissa). [Kress & van Leeuwen 1996, 56–57, 58.] Kerronnalliset kuvat sisältävät liikettä joko eksplisiittisesti kuvasarjojen muodossa tai sitten yksittäisen kuvan liikevektorien implikoimana. Kerronnalliset kuvat esittävät objektien välistä vuorovaikutusta. Käsitteelliset rakenteet muodostuvat paikkasidonnaisista elementeistä ja niiden välisistä rakenteellisista suhteista. Kress ja van Leeuwen [1996, 56–73, 79–119] käsittelevät tarkemmin kerronnallisia ja käsitteellisiä rakenteita.

Verbaali viestintä voi olla sumeaa ja monimerkityksellistä, mutta kuvallinen viestintä on niitä moninkertaisesti. Teoriassa on kuitenkin mahdollista johtaa automaattisesti semanttisia merkityksiä kuvallisesta koodista. Semanttinen päättely edellyttää kuitenkin jonkinlaisten säännönmukaisuuksien löytämistä kuvallisesta mediasta ja sen koodin tuntemista, jonka sääntöjen mukaan analysoitava syntagma on järjestetty, sekä näiden säännönmukaisuuksien ilmaisemista eksplisiittisesti indeksointialgoritmille. Tämän luvun loppuosassa käsitellään vielä hieman kuvallisia koodeja semiotiikan näkökulmasta, lähinnä kuvan eri osille annettuja merkityksiä, vaikka ne liittyvätkin pitkälti konnotaatioihin ja ikonologisiin merkitystasoihin, joiden automaattinen indeksoiminen on epärealistista nykytietämyksen valossa.

Kressin ja van Leeuwenin [1996] käsittelemistä kuvallisista koodeista voitaisiin automaattisen indeksoinnin näkökulmasta ottaa huomioon ainakin geometristen muotojen, kuvan koon, kuvausetäisyyden sekä perspektiivin merkitykset; näiden lisäksi kannattaisi ottaa huomioon sommittelu eli kuvan elementtien keskinäiset suhteet, elementtien yhdistäminen merkitykselliseksi kokonaisuudeksi sekä elementtien silmiinpistävyys, kehystäminen ja niiden informaatioarvo. [Kress & van Leeuwen 1996, 51–55, 130–148, 181–212.]

Informaatioarvo riippuu (1) elementtien sijoittelusta, mihin vaikuttavat kuvan alueet eli vasen ja oikea, yläosa ja alaosa, keskusta ja marginaali; (2) elementtien silmiinpistävyydestä ('salience'), joka johtuu niiden sijoittamisesta etu- tai taka-alalle tai sitten elementtien tarkkuuden eroista; (3) elementtien kehystämisestä eli niiden eristämisestä toisistaan esimerkiksi viivoilla, jotka osoittavat mikä kuuluu ja mikä ei kuulu yhteen. Silmiinpistävyys voi tehdä joistain elementeistä tärkeämpiä kuin niiden sijainti muuten mahdollistaisi. [Kress & van Leeuwen 1996, 181–212, 214–218.] Näitä elementtien informaatioarvoa koskevia havaintoja voidaan vielä edelleen tarkentaa, sillä Kress ja van Leeuwen [1996, 186–192] käsittelevät annetun ja uuden, vasemman ja oikean informaatioarvoa. Esimerkiksi televisiossa ja erityisesti haastatteluissa haastattelija sijoitetaan yleensä katsojan näkökulmasta vasemmalle eli henkilöksi, joka on katsojalle tuttu ja esittää katsojan puolesta kysymyksiä. Uuden ja vanhan suhdetta voidaan korostaa myös horisontaalisella kameran liikkeellä panoroinnilla. [Kress & van Leeuwen 1996, 186–192.] [4] Informaatioarvo vaihtelee myös keskustan ja marginaalin välillä:

4 Vasemman puolen käsittäminen tutuksi johtunee länsimaisesta kirjoitusjärjestelmästä.

keskusta on usein symbolinen piirros, jopa kuin logo, joka yhdistää sitä ympäröivää informaatiota ja toimii lisäksi vasemman ja oikean sekä ylä- ja alaosan välittäjänä [Kress & van Leeuwen 1996, 203–212].

Kuvan jakamista elementteihin ja niiden korostamista käytetään TV-ilmaisussa. Seiter [1992, 44] mainitsee, kuinka Yhdysvalloissa grafiikkaa käytetään kuvien merkitysten selventämiseksi. Erilaisia kaavioita asetetaan uutis- ja urheilulähetysten päälle ja houkutellaan näin katsojia tarkastelemaan niitä. Entuudestaan yksinkertaistettuja ('pared-down') kuvia reunustetaan ja kehystetään; sanoja ilmestyy kuvaruudulle ohjelmien, sponsoreiden, verkkojen, kaapeliasemien, tuotteiden nimien ja henkilöiden tunnistamisen helpottamiseksi. [Seiter 1992, 44.] Seuraavassa luvussa käsitellään tarkemmin televisioilmaisua koodina.

2.4 Televisio genrenä, koodina ja viestintävälineenä

Pietilän [1995, 172] mukaan genret eli lajityypit ovat tuottamista ja vastaanottoa sääteleviä tapoja. Genre on enemmän tai vähemmän implisiittisesti tuotettu sopimus, johon kuuluvat esteettiset ja kerronnalliset käytännöt, joita tuottajat, kriitikot ja lukijat määrittävät. Gripsrudin [1995, 164] sekä Grossbergin ja muiden [1998, 159, 160–161] mukaan genre on luokka tai kategoria tekstejä ja konventioita, jotka ovat jollain tavalla samankaltaisia, vaikkakaan mikä tahansa samankaltaisuus ei sijoita tekstejä samaan genreen. Feuer [1992, 138] huomauttaa, että jo pelkkä genre-termin käyttäminen vihjaa, että yksittäiset tekstit eivät ole ainutlaatuisia, jos ne voidaan ylipäätään luokitella. Genre on linkki tuottajien ja yleisöjen välillä; se on koodi ja konventio, joka säätelee tekstin tuottamista ja rajoittaa sen lukemisen mahdollisuuksia [Gripsrud 1995, 165]. Feuerin [1992, 142] mukaan genret antavat tuottajille helppokäyttöisen “työkalulaatikon”, jonka sisältämien genren tuottamista ja lukemista koskevien konventioiden avulla voidaan sanoa hyvin pienessä tilassa tai lyhyessä ajassa paljon. Genret ovat siis semioottisia järjestelmiä. Feuerin [1992, 143] mukaan genrejä voidaan tutkia formalisoituina merkkijärjestelminä, joiden säännöt on omaksuttu kulttuurista konsensuksesta. Kun teksti tuotetaan tietyn genren rajoissa, tekstin tulkintaa voidaan ohjailla genreen sisältyvien lukutapaa koskevien odotusten avulla.

TV-instituutiot ovat kehittäneet käyttöönsä sopivia genrejä, jotka koostuvat kerronnallisista ja esteettisistä sopimuksista, joiden enemmän tai vähemmän ääneenlausumattomien sääntöjen mukaan ohjelmia tuotetaan [Ellis 1992, 111]. Televisiossa genret on eroteltu selvästi toisistaan tunnuksin, joiden tarkoitus on kutsua katsojat tiettyyn katsomiskonventioon. Esimerkiksi uutislähetyksillä on omat tunnuksensa ja mainokset erotetaan muista ohjel-

mista mainoskatkon tunnuksella. [Pietilä 1995, 172.] Televisiokoodin perusyksikköjä ovat peräkkäin esitettävät merkityksiltään sisäisesti yhtenäiset ja itseselittävät segmentit, joista osa toimii linkkeinä toisiin segmentteihin, kuten esimerkiksi uutisankkurin uutisjuttua edeltävä juonto tai uutislähetyksen tunnus. Segmentti on itsenäinen ja kantaa jotain tiettyä tunnelmaa tai sanomaa, ja jokainen segmentti edustaa siirtymää ohjelman kantamassa argumentissa. Televisiokerronnassa siirtyminen segmenttien välillä ei tapahdu elokuvan tavoin kausaalisessa ketjussa, vaan tapahtumien seurauksia viivytetään. [Ellis 1992, 117–119, 148, 151.]

Cornerin [1995, 55] mukaan, vaikka TV-uutiset ovat saaneet kerronnan ja esittämisen vaikutteita muista genreistä, uutisten ilmaisu voidaan silti erottaa muista ohjelmatyypeistä. Monet uutisjutut esitetään kertomuksina. Uutisten ero fiktiiviseen tuotantoon ei ole niinkään kerronnan tyylikeinoissa vaan käytetyssä materiaalissa, sillä uutiset eivät lavasta kuvamateriaalia vaan järjestelevät "oikeasta elämästä" tallennettua materiaalia tarinan muotoon [Corner 1995, 59]. Lopputuloksena ei ole läheskään aina sujuvasti etenevä tarina, vaan materiaalin puuttumisen vuoksi siihen tulee katkoksia ajassa ja paikassa toisin kuin fiktiivisessä kerronnassa. Uutisjutuista voidaan kuitenkin usein löytää kertomuksen peruskaava, jonka elementtejä ovat tasapainotila, kriisi, ristiriita ja uusi tasapainotila [Pietilä 1995, 193]. Uutisten tarinat (eli uutisjutut) ovat lyhyitä, vaikka ne voivatkin olla osana laajempaa kertomusta. Näistä tarinoista puuttuu klassinen juonenkehitys, vaikka niissä voidaankin esittää arvoituksia ja konfliktejä [Corner 1995, 57].

Televisiotuotannon oppikirjat painottavat visuaalisten koodien käyttämistä kuvan yksinkertaistamiseen esimerkiksi symmetrisen sommittelun ('composition'), värien yhteensopivuuden ja valaistuksen avulla [Seiter 1992, 43]. Seiterin [1992, 43] mukaan lähikuvat kasvoista ja puhe – ns. “puhuvat päät” – ovat erityisen tärkeitä televisioilmaisussa pienen kuvaruudun takia. Pieni kuva pakottaa ohjaajat käyttämään myös nopeampaa leikkausta kuin elokuvissa [Ellis 1992, 131]. Televisiokerronnassa kuvassa keskitytään olennaiseen yksityiskohtien kustannuksella. Ääntä käytetään televisiossa huomion herättämiseen, yksityiskohtien tuomiseen esille sekä jatkuvuuden tunteen ylläpitämiseen. [Ellis 1992, 123, 129–132.] Corner [1995, 61] huomauttaa, että vaikka pelkkiä kuvia voidaan artikuloida yhteen muodostamaan monimutkaisia merkityksiä, käytännössä puhe muodostaa suurimman osan uutisen informaatiosta ja sitoo kuvat yhteen. Esimerkiksi kuvaruututekstit seuraavat yleensä ääniraidan puhetta [Seiter 1992, 44]. Seiter [1992, 44] esittää, että ääniraita on televisiossa niin yksiselitteinen, että televisio-ohjelmaa voidaan ymmärtää pelkästään kuuntelemalla. Mitä tulee kuvan ja äänen suhteeseen, Cornerin [1995, 60, 63] mukaan uutiskuva on usein assosiaatiosuhteessa käsiteltävään aiheeseen indeksisen suhteen sijaan, jolloin se, mitä näytetään ei ole suorassa yhteydessä siihen, mitä sanotaan. Pietilä [1995, 186] on myös esittänyt, että myös kuvan ja äänen

suhde voi olla paitsi indeksinen myös ikoninen tai symbolinen. Symbolinen kuvan ja äänen suhde on tyypillinen abstrakteja asioita käsiteltäessä kuten talousuutisoinnissa: esimerkiksi puhuttaessa viennin kasvusta näytetään kuvaa vilkkaasta satamasta. Suhteet eivät aina kuitenkaan ole selviä ja voivat olla useampaa tyyppiä samaan aikaan; automaattisen indeksoinnin kannalta riittää, että otetaan huomioon kuvan ja äänen epäsuoran yhteyden mahdollisuus.

3 HAHMOPOHJAISET INDEKSOINTIMENETELMÄT JA SEMANTTINEN PÄÄTTELY

Edellisessä luvussa käsiteltiin merkityksen muodostumista ja sitä ohjaavia koodeja. Samalla käsiteltiin merkityksen tasoja, joilla visuaalista informaatiota voidaan lähestyä. Tässä luvussa tarkastellaan automaattisia sisältö- eli hahmopohjaisia indeksointimenetelmiä ja semanttisia malleja, joiden avulla merkityksen muodostuminen ilmaistaan eksplisiittisesti ja joilla pyritään johtamaan merkityksiä automaattisesti videodatasta.

3.1 Visuaalisista piirteistä semantiikkaan

Guptan, Santinin ja Jainin [1997, 35] mukaan informaatio on dataa, jolla on semanttinen assosiaatio. Visuaaliseen dataan liittyy informaatiota data*sta* eli metadataa ja informaatiota datas*sa* itsessään, jolla tarkoitetaan visuaalisia piirteitä [Gupta ja Jain 1997, 72]. Metadata tuotetaan soveltamalla piirteitä poimivia algoritmeja mediaobjekteihin, joita ovat esimerkiksi teksti, kuva, video ja vastaavat [Prabhakaran 1997, 55]. Indeksoinnissa tuotettava metadata voidaan jaotella Del Bimbon [1999, 2] ja Prabhakaranin [1997, 54] mukaan seuraavalla tavalla:

- *Sisältöriippumaton metadata* tarkoittaa dataa, joka ei suoraan liity videosisältöön vaan on jossakin suhteessa siihen, esimerkkinä tallenteen fyysinen formaatti, tekijöiden nimet, tekijänoikeudet ja kuvauspaikka.
- *Sisältöä koskeva metadata* voidaan jakaa (a) *sisällöstä riippuvaan metadataan* ja (b) *sisältöä kuvailevaan metadataan*. Sisällöstä riippuva metadata koskee matalia ja välittäviä ('intermediate') havaittavia piirteitä, kuten värejä, tekstuureita, muotoja, paikkasidonnaisia suhteita ja liikettä. Sisältöä kuvaileva metadata viittaa havaittavista piirteistä johdet-

taviin semanttisiin tulkintoihin; tulkinnoilla tarkoitetaan kuvattujen asioiden suhdetta tosimaailman olioihin ja tapahtumiin sekä niiden assosioituihin merkityksiin ja tunteisiin. [Del Bimbo 1999, 2; Prabhakaran 1997, 54.]

Koska tutkielmassa keskitytään hahmopohjaisiin indeksointimenetelmiin eikä niinkään kokonaisiin videotietokantoihin, tarkastelu keskittyy sisältöä koskevaan metadataan ja sen tuottamiseen, vaikka jokainen järjestelmä tarvitseekin sisältöriippumatonta metadataa fyysisten dokumenttien kuvailuun ja tietokantojen järjestämiseen fyysisellä tasolla. (Multimediatietokantoja käsittelee tarkemmin mm. Prabhakaran 1997.) Sisältöä koskevaan metadataan liittyen on otettava huomioon ne muodot, joilla raaka videodata voidaan esittää tietokannassa: matemaattisesti ilmaistuina hahmoina ja tekstuaalisesti ilmaistuina käsitteinä. Videosisällön esittäminen uudelleen, tehtiinpä se tekstuaalisesti tai matemaattisesti, edellyttää aina tulkintaa: kaikki ajallispaikalliset rakenteet ja merkitykset on tulkittava videodatasta indeksoitaessa, sillä digitaalisen videon mediaobjektit ja ominaisuudet, kuten ääni ja kuva, ovat binaarisia, rakenteettomia ja tulkitsemattomia ilman ihmisen tai tietokoneohjelman niistä muodostamaa tulkintaa [Prabhakaran 1997, 53].

Käsitepohjaiseen indeksointiin liittyen Gupta, Santini ja Jain [1997, 37] kertovat, että visuaalinen data on paljon monitulkintaisempaa kuin teksti. Teksteissä käytetyillä sanoilla on rajattu määrä merkityksiä, joten sanan oikea semanttinen merkitys, jos se ei ole itsestään selvä, voidaan disambiguoida rajoitetusta määrästä vaihtoehtoja [mts. 37]. Idrisin ja Panchanathanin [1997, 146] mukaan on suoraviivaista indeksoida videoita tekstuaalisesti avainsanoilla ja käyttää niitä niihin liittyvän videodatan indekseinä tavanomaisessa tekstitietokannassa. Suurin ero perinteisen tekstitietokannan ja multimediatietokannan välillä on siinä, että viimeksi mainitun pitää hallita paitsi monta eri datatyyppiä myös niiden moninaiset tulkinnat, sillä varsinkin kuvallinen sisältö voidaan havaita, esittää ja ymmärtää useilla eri tavoilla [Antani et al. 2002, 946; Gupta et al. 1997, 42]. Useat tutkijat [ks. Bolle et al. 1998; Grosky 1997, 73; Gupta & Jain 1997, 72] ovat esittäneet, että visuaalisen datan esittäminen käsitepohjaisesti aiheuttaa ongelmia, joita tekstidokumenttien indeksoinnissa ei esiinny: Tekstidokumentteja indeksoidaan ja niitä haetaan samassa muodossa, josta dokumentit muodostuvat, mutta videodataa ei voida kuitenkaan esittää tekstuaalisesti ilman, että joitain alkuperäisen ei-tekstuaalisen datan keskeisiä ominaisuuksia ei kadotettaisi. Teksti ei pysty kunnolla esittämään kuvan havaittavia ominaisuuksia, visuaalisten piirteiden silmiinpistävyyttä ja niiden havaittavaa samankaltaisuutta [Del Bimbo 1999, 4]. Myös Rasmussen [1997, 176] huomauttaa, että jos välinettä pidetään viestinä, eli muotoa ja sisältöä erottamattomana, niin visuaalinen ilmaisu ei ole ongelmattomasti käännettävissä tekstuaaliseksi indeksointikieleksi ja käsitteiksi.

Esimerkiksi jos värejä edustetaan sanoilla tietokannassa, törmätään väistämättä siihen tosiasiaan, että ihmiset ymmärtävät värit eri tavalla: esimerkiksi ero sinisen ja vihreän välillä ei ole rajatapauksissa mitenkään selvä [ks. Fiske 1992, 63]. Videosisältöjen esittäminen tekstuaalisesti vaatii usein hankalaa erikoissanastoa, joka ei välttämättä edes edusta videosisältöä johdonmukaisesti, eivätkä hakutulokset yleensä voi olla tyydyttäviä, jos kysely perustuu piirteisiin, joita ei ole voitu kunnolla esittää tietokannassa [Idris & Panchanathan 1997, 146].

Uuden sukupolven tiedonhakujärjestelmät tukevat käsitteellisiin eli verbaaleihin ilmaisuihin perustuvan tiedonhaun lisäksi myös tiedonhakua visuaalisesti havaittavalla tasolla, jolla videoita esitetään objektiivisesti kuvankäsittelyn, hahmontunnistuksen ja konenäön avulla [Del Bimbo 1999, 4]. Hahmopohjaisissa indeksointimenetelmissä videodatasta tuotetaan metadataa poimimalla automaattisesti paikkasidonnaisista ja ajallisista ominaisuuksista erottelukykyisiä piirteitä, joita käytetään videon kuvailussa kuin sanoja tekstidokumenttien indeksoinnissa [Idris & Panchanathan 1997, 159; Markkula et al. 2001, 1; Prabhakaran 1997, 55; Rasmussen 1997, 170]. Videodokumentteja indeksoivan järjestelmän on tosin käsitettävä suurempi määrä piirteitä kuin tekstidokumentteja indeksoivan järjestelmän, joissa tyypillisesti sanat tai alisanat kuten foneemit ovat indeksoitavia piirteitä [ks. Ponceleon & Srinivasan 2002, 10]. Del Bimbon [1999, 22–23] mukaan videoiden havaittavien piirteiden kuten värin, tekstuurin, muodon, kuvan rakenteen, paikkasidonnaisten suhteiden ja liikkeen esittäminen ovat visuaalisen tiedonhaun keskeisimpiä ongelmia. Kuvananalyysimenetelmät ja hahmontunnistusalgoritmit tarjoavat keinoja poimia ('extract') numeerisia kuvaajia ('descriptor'), jotka antavat piirteille kvantitatiivisia mittoja. Konenäkö mahdollistaa objektien ja liikkeen tunnistamisen vertaamalla havaittuja ('extracted') kuvioita ennaltamääriteltyihin malleihin. [Del Bimbo 1999, 22–23.]

Koska hahmopohjaisissa hakujärjestelmissä videosisältöjä haetaan suoraan sisällön perustavaa laatua olevien piirteiden perusteella, ilman tekstuaalista ”välikättä”, Smithin [2001, 970] mukaan hahmopohjaiset menetelmät ratkaisevat ongelmat, jotka johtuvat avainsanapohjaisten järjestelmien riittämättömyydestä täydellisyyden ('completeness'), johdonmukaisuuden ja objektiivisuuden suhteen. Vaikkakin värijakauman ('color distribution') ja sommittelun ('spatial layout') kaltaiset piirteet antavatkin vain varsin rajoittuneen luonnehdinnan kuvien semanttisesta sisällöstä, hahmopohjaiset menetelmät on havaittu toimiviksi haettaessa nopeasti kuvia niiden visuaalisen samankaltaisuuden perusteella. [Smith 2001, 970.] Joitain semanttisia primitiivejä, kuten objekteja, tapahtumia ja kerronnallisia rakenteita, voidaan johtaa analysoimalla matalan tason piirteiden yhdistelmiä sopivien mallien mukaan. Semioottisen analyysin avulla voidaan tehdä eksplisiittiseksi, kuinka visuaaliset primitiivit välittävät merkityksiä havaitsijalle. [Del Bimbo 1999, 22–23.] Mallien tarkoitus on eksplisiittisesti ilmaista,

kuinka merkitys muodostuu ihmissubjektille.

Guptan ja muiden [1997, 35] mukaan tämän hetken tosiasia on, että informaatiojärjestelmät toimivat parhaiten, jos käsiteltävä data on rakenteisessa muodossa. Niissä sovelluksissa, joissa käsiteltävällä datalla ei ole eksplisiittisesti ilmaistua ihmisten tuottamaa rakennetta, järjestelmän pitäisi itsessään poimia semanttisia assosiaatioita raa'asta datasta eli tuottaa informaatiota. Sellainen hakujärjestelmä, jonka pitää poimia ja tuottaa informaatiota pelkästään raa'asta datasta, on luonnostaan heikompi kuin järjestelmä, jolle on erikseen ilmaistu, mitkä semanttiset assosiaatiot ovat. [Gupta et al. 1997, 35–36.] Automaattista semanttista päättelyä vaikeuttaa se, että informaatiota poimivat hahmopohjaiset algoritmit toimivat hyvin matalalla visuaalisella abstraktiotasolla toisin kuin käyttäjät, jotka hakevat tietoa suhteellisen korkealla abstraktion tasolla. Videoiden havaittavista piirteistä, eli niiden perustavaa laatua olevista ominaisuuksista, johdetaan semanttisia käsitteitä mallintamalla kontekseja, joissa piirteet esiintyvät ja merkityksellistyvät. Piirteitä poimivat algoritmit ja mallit vastaavat siis tavallaan ihmisen havainto- ja päättelykykyä. Piirteitä, niiden poimimista, semanttista päättelyä ja malleja käsitellään lisää seuraavaksi.

3.2 Visuaaliset piirteet ja samankaltaisuuden arvioiminen

Piirre on ominaisuus ('attribute'), joka on johdettu ('derive') alkuperäisestä visuaalisesta objektista jonkin kuvananalyysialgoritmin avulla ja joka luonnehtii jotakin kuvan tiettyä ominaisuutta ('property') [Gupta et al. 1997, 37]. Informaatiota visuaalisessa objektissa on objektin arvo eli kuvaus siitä, mitä se sisältää [ks. Gupta et al. 1997, 35–36]; yksinkertaisimmat piirteet, joita voidaan laskea, perustuvat raakadatan pikseliarvoihin. Yksittäisillä pikseleillä ei voi olla mitään muuta arvoa kuin niiden värisävy: yhden bitin värisyvyydellä pikseli voi olla RGB-värijärjestelmässä päällä (valkoinen) tai pois päältä (musta), mutta kahdeksan bitin tarkkuudella yksi pikseli voi saada jo 256 eri arvoa värisävyille. Niinpä matalimmalla mahdollisella tasolla voidaan pyytää järjestelmää hakemaan kaikki ne kuvat, joissa tietyllä alueella esiintyy valkoista väriä, jos valkoinen määritellään RGB-järjestelmässä niin, että kaikkien värikomponenttien arvot ovat 245–255. Jos käyttäjät tyytyisivät tämänkaltaisiin kyselyihin, visuaalinen tiedonhaku olisi tavattoman yksinkertaista [Gupta & Jain 1997, 72]. Gupta ja Jain [1997, 72] kuitenkin kertovat, että näin yksinkertaiset piirteet ovat liian herkkiä hälylle kuvassa, ja ne eivät ota huomioon kuvissa esiintyvien objektien eri asentoja tai eroja kuvien valaistuksessa.

Hahmopohjaiset indeksointimenetelmät toimivat raa'asta informaatiosta tuotta-

millaan abstrakteilla piirteillä [Lu 1999, 184]. Piirteet esitetään (1) yleensä matemaattisesti joukkona numeroita, joita kutsutaan piirrevektoreiksi, (2) kuvasta mitattujen muuttujien välisenä jakaumana ('distribution') eli histogrammina tai (3) joukkona pisteitä piirreavaruudessa [Gupta et al. 1997, 37–38]. Vektorit muodostavat yhtenäisen perustyypin piirteille, jotka esittävät videon sisältöä [Lu 1999, 184]. Piirteet ovat visuaalisen datan esitystavoista käsitteellisesti matalin. Gupta ja muut [1997, 37] sekä Lu [1999, 184] käsittelevät piirteille suoritettavia matemaattisia operaatioita.

Olipa joku piirre vektori tai ei, se ”asuttaa” muuttujiensa määrittämää aluetta piirreavaruudessa. Esimerkiksi tekstuuri voidaan esittää kolmella lukuarvolla: satunnaisuudella, jaksoittaisuudella ('periodicity') eli toisteisuudella ja johdattavuudella ('directionality'). Kuvalla, jossa on kymmenen aluetta ja niissä eri tekstuuri, on kymmenen pistettä satunnaisuutta, jaksoittaisuutta ja johdattavuutta koskevassa koordinaatistossa. Kun tietokantaan lisätään kuvia, kolmiulotteinen piirreavaruus täyttyy pisteillä jokaista uutta teksturoitua aluetta varten. Tätä avaruutta voidaan kohdella informaatio-objektina ja sille voidaan tehdä kyselyjä erilaisten operaatioiden avulla. [Gupta et al. 1997, 38.] Gupta ja muut [1997, 38–39] käsittelevät piirreavaruudessa käytettäviä operaatioita.

Ryhmittelemällä useita piirteitä saadaan yksi monimutkainen vektori, joka on ilmaisuvoimaisempi kuin yksittäiset piirteet. Esimerkkinä tästä Gupta ja muut [1997, 39] viittaavat kirjallisuudessa käsiteltyyn menetelmään, joka tunnistaa paljasihoisia ihmisiä yhdistämällä ihonvärin tunnistamiseen tarkoitetun piirreryhmän sylinterimäisiin geometrisiin piirteisiin. [Gupta et al. 1997, 39–40.] Antanin ja muiden [2002, 954–955] mukaan monet tekstuuripiirteisiin perustuvat lähestymistavat yhdistävät erilaisia piirteitä yhdeksi piirrevektoriksi. Gupta ja muut [1997, 39–40] käsittelevät piirreryhmille suoritettavia operaatioita.

3.2.1 Yleisimmät piirteet ja niiden esittäminen

Antanin ja muiden [2002, 951] mukaan kirjallisuudessa kuvaillut menetelmät käyttävät pääasiassa kolmen-tyyppisiä piirteitä: värejä, tekstuureja ja muotoja. Näitä käsitellään lyhyesti seuraavaksi.

3.2.1.1 Väreihin perustuvat piirteet

Väri on välittömästi havaittava, keskeinen ja käytetyin visuaalinen piirre hahmopohjaisissa indeksointimenetelmissä [Antani et al. 2002, 951; Del Bimbo 1999, 81; Idris & Panchanathan

1997, 148]. Värien etuna on riippumattomuus kuvakulmasta, kuvan kääntymisestä ('translation') ja resoluutiosta. [Antani et al. 2002, 951; Idris & Panchanathan 1997, 148.] Värillä on kaksi muuttujaa: sävy ('hue') ja kyllästys ('saturation'). Sävy tarkoittaa värien spektriä eli värihavainnon tosiasiallista aallonpituutta ja kyllästys (tai värikylläisyys) sitä, kuinka paljon harmaata on lisätty puhtaisiin väreihin. Puhtaissa eli kyllästyneissä väreissä ei ole lainkaan valkoista väriä. Kirkkaus ('brightness') edustaa värin intensiteettiä. [Del Bimbo 1999, 82; Gupta & Jain 1997, 73–74.]

Kromaattisia piirteitä eli väriärsykkeitä ('colour stimuli') esitetään yleisesti geometristen värimallien avulla, jolloin värit esitetään pisteinä kolmiulotteisessa väriavaruudessa, vaikka värihistogrammit ovatkin perinteisin ja useimmiten käytetty tapa esittää matalan tason väriominaisuuksia ja niiden jakautumista kuvissa. Värihistogrammi on moniulotteinen vektori, joka tuotetaan erottamalla kuvan värit ja laskemalla kuinka monta pikseliä on minkäkin värisiä. Kehyksistä puretut väripiirteet tallennetaan värilaarien ('color bins') muodossa. Jokainen laari on yleensä kuutio kolmiulotteisessa väriavaruudessa (joka vastaa RGB-järjestelmän perusvärejä). Histogrammin arvo ilmaisee niiden pikseleiden prosenttimäärää, jotka ovat mahdollisimman samankaltaisia tiettyyn väriin nähden. Mitkä tahansa kaksi pistettä samassa laarissa edustavat samaa väriä. Myös mustavalkokuvien harmaatasoja voidaan tallentaa ja esittää histogrammien muodossa. Kahden kuvan välistä samankaltaisuutta voidaan verrata esimerkiksi niiden histogrammeja käyttämällä. [Del Bimbo 1999, 24, 81, 94; Idris & Panchanathan 1997, 148–149; Prabhakaran 1997, 76–77.]

Yleisesti ottaen väripiirteet ovat varsin helposti purettavissa ja täsmäytettävissä. Värikuvaajan ('descriptor') määritteleminen ('specification') vaatii väriavaruuden kiinnittämistä ('fixing') ja sen osittamista ('partitioning'). Väriavaruuden osittaminen on välttämätöntä, jotta piirreavaruuden moniulotteisuutta ('dimensionality') voitaisiin vähentää yhdistämällä ('aggregate') samankaltaisia värejä ja erottamalla havaittavasti erilliset värit. [Smith 2001, 971.] Kuvissa on niin paljon väri-informaatiota, että samankaltaiset ja havaittavasti erilaiset värit erotetaan eri ryhmiin, jotta niiden käsitteleminen olisi helpompaa. Ponceleonin ja Srinivasanin [2002, 12] mukaan värejä mallinnetaan 8*8 pikselin lohkoina. Värejä tarkasteltaessa huomiota voidaan kiinnittää muun muassa vallitseviin ja keskimääräisiin väreihin, sommitteluun ('layout') ja alueisiin [Ponceleon ja Srinivasan 2002, 10]. Ihmisen havaintoelimiä kuvailevien mallien mittaamien etäisyyksien väriavaruudessa pitäisi vastata ihmisen havaintoja värien eroista. Lisäksi värikuviot pitäisi esittää niin, että silmiinpistävät kromaattiset ominaisuudet saadaan poimittua ('capture'). [Del Bimbo 1999, 81.] Idris ja Panchanathan [1997, 148–-150] esittelevät useita eri lähestymistapoja kuvien värijakaumien esittämiseen sekä laskennan vähentämiseen. Antani ja muut [2002, 951–953] käsittelevät myös väripohjaisia piirteitä ja

useita erilaisia täsmäytysmenetelmiä värejä piirteinään käyttäviä kyselyitä varten.

Saatavilla on useita erilaisia värimalleja, jotka voidaan jaotella eri tavoin. Esimerkkinä yleisistä värimalleista voidaan mainita laitteistosuuntautunut RGB, joka perustuu optisten laitteiden kuten näyttöjen ominaisuuksiin, ja käyttäjäsuuntautunut HSV, joka pohjautuu ihmisten havaitsemiin väreihin. RGB on yleisin värimalli digitaalisissa kuvissa; se perustuu jossain määrin verkkokalvon fysiologiaan. Värit RGB-järjestelmässä perustuvat punaisen, vihreän ja sinisen lisäämiseen. Sekoittuneet värit saadaan lisäämällä valkoista valoa. [Antani et al. 2002, 951; Del Bimbo 1999, 81–82, 84.]

3.2.1.2 Tekstuureihin perustuvat piirteet

Tekstuuri on objektin havaittavan pinnan keskeinen piirre, jolla tarkoitetaan pinnaltaan toisteisia (tai kvasitoisteisia) samankaltaisesti kuvioituja alueita, joissa on suuria eroja pikseleiden kirkkausarvoissa ja joita väripiirteet eivät riitä kuvaamaan. Psykologisesta näkökulmasta ihmiset kiinnittävät huomiota tekstuurin rakeisuuteen, suuntaan ja toisteisuuteen. Tekstuurin koko, siitä eroteltavissa olevien harmaatason väriprimitiivien määrä ja näiden primitiivien paikkasidonnainen sijoittuminen ovat kaikki keskinäisessä suhteessa toisiinsa ja kuvaavat tekstuuria. Tekstuureilla voi olla tilastollisia ja rakenteellisia ominaisuuksia ja ne voivat muodostua toisteisista rakenteisista tai satunnaisista elementeistä. Tekstuurien ominaisuuksiin kuuluvat: kontrasti, yhdenmukaisuus ('uniformity'), karheus ('coarseness'), rosoisuus ('roughness'), täsmällisyys ('regularity'), taajuus ('frequency'), tiheys ('density') ja suuntaisuus ('directionality'). Koska tekstuureita on mahdotonta ilmaista sanoin, niitä esitetään yleensä numeerisilla vektoreilla. [Antani et al. 2002, 954; Del Bimbo 1999, 25, 117; Idris & Panchanathan 1997, 150; Smith 2001, 971.] Tekstuurin mallintamisen ja luokittelemisen tekniikat voidaan jakaa Idris ja Panchanathanin [1997, 150–151] sekä Ponceleonin ja Srinivasanin [2002, 15] mukaan seuraavasti:

1. *Rakenteiset tekniikat*: Tekstuurit luokitellaan heikkoihin tai vahvoihin riippuen niiden primitiivien (eli piirteiden) välisestä paikkasidonnaisesta vuorovaikutuksesta. Primitiivillä tarkoitetaan tässä yhteydessä ryhmää soluja, joita kuvaavat harmaatasot, muodot ja homogeenisyys.
2. *Tilastolliset* ('statistical') *tekniikat*: Tekstuurit luokitellaan tasaisiksi ('smooth'), hienoiksi ('fine'), karkeapintaisiksi ('coarse'), rakeisiksi ('granular'), laineileviksi ('ripple'), säännönmukaisiksi ('regular'), epäsäännöllisiksi ('irregular') tai lineaarisiksi.

3. *Spektraaliset* ('spectral') *tekniikat*: Fourier-muunnoksiin perustuvilla tekniikoilla tekstuuri analysoidaan tunnistamalla spektrin ('spectrum') energian korkeat ja matalat tasot.

Antani ja muut [2002, 954] sekä Idris ja Panchanathan [1997, 150–151] käsittelevät tarkemmin menetelmiä, joilla tekstuureja on mallinnettu, luokiteltu ja ilmaistu.

3.2.1.3 Muotoihin perustuvat piirteet

Muodolla ('shape') tarkoitetaan objektin profiilia ja fyysistä rakennetta. Jos objektit eivät ole yksinkertaisia kaksiulotteisia geometrisia kuvioita, niiden kuvaileminen tekstuaalisesti on mahdotonta. Tekstuualisten kuvailujen sijaan muotoja voidaan esittää piirrejoukoilla, esimerkiksi Idrisin ja Panchanathanin [1997, 151–152] käsittelemillä tekniikoilla. Muotoon liittyvät piirteet voidaan jakaa globaaleihin, jotka kuvaavat muotoa yleisesti, ja paikallisiin, jotka kuvaavat muodon paikallisia elementtejä. Globaalit piirteet johdetaan ('derive') koko muodosta ja paikalliset piirteet johdetaan paikallisista ominaisuuksista, kuten esimerkiksi objektin reunoista. [Del Bimbo 1999, 25–26; Idris & Panchanathan 1997, 151.] Idris ja Panchanathan [1997, 151] mainitsevat vielä hahmotelmista ('sketch'), joilla he tarkoittavat abstrakteja kuvia, jotka sisältävät objektien rajat ('outline'). Hahmotelma luodaan käyttämällä reunojen ('edge') tunnistusmenetelmiä ja ohentamalla ('thinning') sekä kutistamalla ('shrink') kuvaa [Idris & Panchanathan 1997, 151.]

Muotoja käsitellään pisteinä muotopiirreavaruudessa. Samankaltaisuus kahden muodon välillä lasketaan pikseleiden (eli pisteiden) välisinä etäisyyksinä. [Del Bimbo 1999, 25–26.] Antanin ja muiden [2002, 953] mukaan muotojen täsmäyttämistä varten on olemassa useita eri lähestymistapoja. Muotoja voidaan käyttää täsmäytysvälineenä kysely esimerkillä -tyyppisissä kyselyissä tai käyttäjän piirtämiin hahmoihin perustuvissa kyselyissä. Viimeksi mainitun lähestymistavan puolesta puhuu se, että käyttäjän piirtämässä hahmossa ihmisen havaitsema samankaltaisuus on luonnostaan eikä täsmäytysjärjestelmään tarvitse kehittää malleja ihmisen samankaltaisuusarvioita varten. Koska käyttäjän piirtämä hahmo ei yleensä vastaa tietokantaan tallennettuja hahmoja, täsmäytysmenetelmän on oltava joustava. [Antani et al. 2002, 953.] Idrisin ja Panchanathanin [1997, 152] mukaan tiedonhaku hahmojen samankaltaisuuden perusteella on vaikeaa, sillä hahmojen samankaltaisuudesta ei ole matemaattisesti tarkkaa määritelmää. Idris ja Panchanathan [1997, 151–152] käsittelevät tarkemmin hahmojen tunnistamista ja niiden perusteella tehtäviä hakuja. Antani ja muut [2002, 953–954] käsittelevät yksityiskohtaisemmin muotoihin perustuvia menetelmiä ja täsmäyttämistä.

3.2.2 Samankaltaisuusoperaatiot

Del Bimbon [1999, 5] mukaan visuaalisen kuvamateriaalin hakeminen edellyttää joukkoa esilaskettuja ('precomputed') erottelevia piirteitä eli malliparametreja, joiden samankaltaisuutta verrataan kyselyn piirteisiin. Visuaalisessa kyselyssä tai muuten kuvien piirrejoukkoja vertailtaessa käytetään samankaltaisuusmittaa [Antani et al. 2002, 949; Del Bimbo 1999, 5]. Del Bimbo [1999, 5–6] huomauttaa, että täsmäyttämisessä on kysymys havaitun objektin vastaavuudesta etukäteen määriteltyyn malliin, millä konenäön yhteydessä tarkoitetaan objektin tunnistamista ja luokittelemista, ja samankaltaisuuden mittaamisessa on kysymys enemmänkin tietokannan kuvien asettamisesta järjestykseen kyselyyn nähden niiden mitatun samankaltaisuuden perusteella. [Del Bimbo 1999, 5–6.] Piirrevektoreilla ja niiden yhdistelmillä voidaan suorittaa monia erilaisia operaatioita, joista eräs yleisimmistä on vektorien välisen etäisyyden laskeminen: mitä suurempi etäisyys sitä vähemmän ne täsmäävät. Jos jokin piirre ei ole vektori, etäisyys lasketaan täsmäytettäessä jakaumien eli histogrammien välillä. Yleisessä käytössä olevia histogrammeja pidetään samankaltaisina, jos niiden etäisyys on sama tai vähemmän kuin asetettu etäisyyden kynnys. [Antani et al. 2002, 949; Del Bimbo 1999, 25–26; Gupta, Santini & Jain 1997, 37]. Antani ja muut [2002, 949] esittelevät joitain yleisiä histogrammien vertailutestejä. Eräs yleisimmistä etäisyysmitoista on euklidinen etäisyys, jota esimerkiksi Satoh, Nakamura ja Kanade [1999, 26, 31] käyttävät kasvojentunnistuksessa [Antani et al. 2002, 949; ks. myös Gupta et al. 1997, 37].

Samankaltaisuusmallien pitäisi olla yhdenmukaisia ihmisen havaintoärsykkeiden ('sensorial stimuli') kanssa [Antani et al. 2002, 950; Del Bimbo 1999, 30]. Del Bimbo [1999, 31] viittaa tunnettuun teoriaan, joka väittää ('postulate'), että ihmisen havainnot samankaltaisuudesta perustuvat tarkoituksenmukaisen etäisyyden mittaamiseen mitallisessa ('metric') psykologisessa avaruudessa. Teoriassa oletetaan, että joukko piirteitä mallintaa havaintoärsykkeen ominaisuuksia niin, että se voidaan esittää pisteenä piirreavaruudessa. Piirrevektorimalli ja metriset mitat ovat kuitenkin osoittautuneet puutteellisiksi psykologisissa tutkimuksissa, sillä muotojen samankaltaisuuden mittaamisessa piirrevektorit ja metriset mitat (kuten euklidinen etäisyys) eivät täysin vastaa ihmisen käsitystä samankaltaisuudesta. [Del Bimbo 1999, 31.] Myös Antani ja muut [2002, 950] ottavat huomioon euklidisen metriikan ongelmat viitatessaan psykologiseen kirjallisuuteen. Piirrevektorimallilla on kuitenkin useita etuja, jotka tekevät siitä suosituimman mallin visuaalisissa hakujärjestelmissä: (1) se on riittävä tietynlaisia inhimillisiä samankaltaisuusarvioita varten (kuten värejä koskevat arviot); (2) se on yhdenmu-

kainen ('consistent') piirrepohjaisen kuvailun kanssa, mikä on yleinen lähestymistapa hahmontunnistuksessa ja konenäössä; (3) piirrevektorimallissa indeksi voidaan rakentaa klassisten moniulotteisten hakumenetelmien ('access methods') mukaan, joita ovat esimerkiksi kolmioepäyhtälöön ('triangle inequality') perustuva samankaltaisuusmittaus, kiinteät ruudukot ja ruudukkotiedostot ('fixed grids', 'grid files'), *K-d*-puut, erilaiset R-puut ja SS-puut. [Del Bimbo 1999, 31–32, 38–46; ks. Apers et al. 1997, 222–224.]

3.3 Semanttinen päättely

Käyttäjien tarpeisiin vastatakseen videotiedonhakujärjestelmän olisi toimittava semanttisella tasolla, jolla käyttäjien tiedontarpeet yleensä muodostuvat – paitsi, jos käyttäjä on tietoisesti kiinnostunut vain havaittavista piirteistä. Petković ja Jonker [2000] kutsuvat kuitenkin semanttisella tasolla hakevia käyttäjiä naiiveiksi. Nykyisten järjestelmien rajoitukset huomioon ottaen olisi realistisempaa miettiä, minkälaisia piirteitä missäkin semanttisissa käsitteissä saattaisi esiintyä ja hakea niiden perusteella – eli toisin sanoen tällä hetkellä olisi parasta vain sopeutua järjestelmien rajoituksiin. Ponceleon ja Srinivasan [2002, 4] mainitsevat, kuinka käyttäjät mielellään tekisivät kyselyitä, joissa haetaan esimerkiksi videokuvaa syöksylaskusta, avaruussukkulan laskeutumisesta tai puhujasta puhumassa Yhdysvaltain lipun edessä. Mainitut esimerkit ovat haastavia semanttisen tason kyselyitä. Realistisempana vaihtoehtona ehdotetaan “punaisen objektin liikkumista vasemmasta yläreunasta oikeaan alareunaan valkoisella taustalla” ja tämän ilmaisemista visuaalisesti piirtotyökaluilla [Ponceleon & Srinivasan 2002]. Indeksointijärjestelmän pitäisi kuitenkin pystyä kuvailemaan videodataa niin korkealla käsitteellisellä tasolla, että järjestelmä voi vastata käyttäjien tarpeisiin heidän tiedontarpeidensa edellyttämällä tavalla – ja se ei ole Ponceleonin ja Srinivasanin esittämä realistinen taso.

Jos edellä käsitellyillä visuaalisilla primitiiveillä tarkoitetaan esimerkiksi jotain tiettyä väriarvoa ja piirteellä sitä kuvaavaa vektoria, semantiikalla tarkoitetaan tässä tapauksessa tuon tietyn väriarvon merkitystä tietyssä kontekstissa ja tietyllä käsitteellisellä tasolla: RGB-arvo (255, 255, 255) on yhtä kuin valkoinen, joka taas on yhtä kuin lumi. Lu [1999, 184] sijoittaa piirteiden kerroksen päälle primitiivejä, joilla tarkoitetaan kokoelmaa vektoreita, jotka edustavat kuvainformaation yksittäistä kategoriaa, kuten väriä, tekstuuria ja muotoa. Primitiivi on siis kuvan semanttisesti merkityksellinen piirre, kuten väri, tekstuuri tai muoto. [Lu 1999, 1984.] Semanttinen sisältö pitää tulkita videoista, sillä videoista poimittavat piirteet saavat semanttisen merkityksen vasta ikonografisella tasolla, esimerkkinä jonkin värin nimeäminen. Kirjallisuudessa lähdetään siitä, että semanttinen sisältö on pitkälti videon havaittavien

piirteiden välittämää, vaikka aihealueen ymmärtämiseen kuuluvaa tietämystä ei voida unohtaa. Kullakin indeksointijärjestelmällä on joukko paradigmaattisia piirteitä, joita se voi tunnistaa videosekvensseistä; kun järjestelmä tunnistaa tietyistä piirteistä muodostuvan olion (eli syntagman), tämä joukko voidaan nimetä jollakin aihealueella vaikka jalkapalloksi. Niinpä piirrejoukkoja voidaan kutsua myös semanttisiksi primitiiveiksi, koska semanttista eli käsitteellistä sisältöä voidaan johtaa videodatasta poimituista piirrejoukoista [Rasmussen 1997, 183]. Videodatan monimutkaisuuden takia semantiikan johtaminen vaatii monimutkaista prosessointia [Grosky 1997, 74]. Bolle ja muut [1998] käyttävät termiä automaattinen videon kommentointi ('automatic video annotation') semanttisten merkityksien liittämisestä automaattisesti videosegmentteihin.

Petkovićin ja Jonkerin [2000] mukaan videotietokanta tarvitsee perinteisistä tekstitietokannoista poikkeavan tietomallin, joka ottaa huomioon paitsi videon sisäiset ('inherent') rakenteelliset ominaisuudet myös videon semanttista sisältöä esittävät käsitteet. Kerrostetut videomallit yhdistävät piirteisiin ja semantiikkaan pohjautuvat lähestymistavat. Petković ja Jonker [2000] esittävät kirjallisuudessa yleisesti käytetyn jäsennyksen: videoita voidaan tarkastella (1) raakana datana, (2) matalan tason visuaalisena sisältönä eli piirteinä sekä (3) semanttisena sisältönä eli käsitteinä. Kerrostetuissa videomalleissa tyypillisesti pohjalla oleva raaka videodata muodostuu videodokumentin ominaisuuksista kuten pakkausformaatista, ruudunpäivitysnopeudesta, bittisyvyydestä, värimallista yms. Tällä matalimmalla eli fyysisellä tasolla video on sekvenssi kehyksiä ja pikseleistä muodostuvia alueita ilman suoraa yhteyttä semanttiseen sisältöön. Seuraava taso muodostuu alueriippumattomista piirteistä eli visuaalisista primitiiveistä, jotka voidaan purkaa automaattisesti: tähän kuuluvat staattiset eli paikkasidonnaiset piirteet kuten muodot, tekstuurit, värihistogrammit sekä dynaamiset piirteet kuten ajallisuus ja liikevektorit. Visuaalisten primitiivien päällä on käsitetaso, joka johdetaan aihealuetta koskevan tietämyksen ('domain knowledge') avulla piirretasosta. [Petković & Jonker 2000.] Semanttinen taso koostuu abstraktiotasoltaan suhteellisen korkeista käsitteistä kuten nimetyistä objekteista ja tapahtumista, joita voidaan kuvata luonnollisen kielen sanoilla. (Tietenkään piirteistä ei voida kirjoittaa ilman, että niistäkin käytettäisiin sanoja.) Esimerkiksi objektilla voidaan tarkoittaa semanttisella tasolla kuvassa esiintyvää autoa ja tapahtumalla auton liikettä.

Del Bimbo [1999, 28] lähestyy videoiden semantiikan tarkastelemista kahdesta näkökulmasta. Perinteisessä perspektiivissä semanttiset primitiivit, kuten objektit, roolit, toimet ja tapahtumat, tunnistetaan ('identify') visuaalisten merkkien abstraktioiksi havaitsemalla ('recognize') ja tulkitsemalla. Havaitseminen perustuu matalan tason piirteistä muodostuvan joukon valitsemiseen ja hahmojen tunnistamiseen objektien luokittelemiseksi. Tulkinta perus-

tuu päättelyyn ('reasoning'), jossa käsinkoodattuja päättelysääntöjä ja aihetietämystä käytetään toimien ja tapahtumien rekonstruointiin. Näiden semanttisten primitiivien tunnistamiseen tarkoitetut ratkaisut ovat yleensä aiheriippuvaisia, esimerkkinä kasvontunnistusjärjestelmät. [Del Bimbo 1999, 28.] Semioottisessa lähestymistavassa tarkastellaan semantiikkaa yhdistettynä sopivaan yhdistelmään visuaalisia merkkejä. Esimerkiksi vaikutelmat ja tunteet voidaan asettaa yhteyteen värien läsnäolon ja yhdistelmien kanssa. Semioottisen analyysin näkökulmasta merkitys muodostuu merkkien yhdistelmästä, ja se voidaan päätellä ohjelmallisesti matalan tason piirteiden, kuten leikkausten, liuotusten ('dissolve'), kameran suunnan, värien läsnäolon ja liikkeen avulla [mts. 230]. Merkkien käsitetään välittävän merkityksiä yleisesti sovittujen konventioiden avulla, joten merkit ovat suhteessa niiden merkityksiin kulttuurisen taustan mukaan. [Del Bimbo 1999, 29, 230.] Merkityksen muodostumisessa on kaksi askelta:

- *Narratiiviset rakenteet*: Abstrakti taso sisältää kaikki ne perusmerkit, jotka tuottavat merkitykset videosekvensseissä: kameran tauot, värit, leikkaustehosteet, rytmi, kuvakulmat yms. Lisäksi tälle tasolle kuuluvat näistä merkeistä koottujen yhdistelmien määrittämät arvot.
- *Diskursiiviset rakenteet*: Konkreettinen taso kuvaa tapoja, joilla narratiivisia elementtejä käytetään tarinan luomiseen. [Del Bimbo 1999, 29.]

Diskursiiviset rakenteet ovat liian monimutkaisia, jotta ne voitaisiin purkaa hahmontunnistuksen tai konenäön avulla. Narratiiviset rakenteet voidaan tunnistaa automaattisesti, kun matalan tason piirteet puretaan visuaalisesta datasta ja käännetään korkean tason semantiikaksi ottamalla huomioon niiden väliset suhteet ja havaitsijoiden niille antamat arvot. [Del Bimbo 1999, 29.]

Yleisten videoiden automaattinen sisällöntunnistaminen ei vaikuta kovinkaan realistiselta vaihtoehdolta edes tulevaisuudessa. Gupta ja muut [1997, 37] huomauttavat että kahdella suhteellisen samanmuotoisella objektilla voi olla erilainen semanttinen merkitys; lisäksi sama objekti voidaan esittää monella eri tavalla. Esimerkiksi joukko samankaltaisia piirteitä, jotka tietyssä kontekstissa voidaan tulkita jalkapalloksi, saattavat jossain toisessa tilanteessa tarkoittaa jotain muuta objektia. Semantiikan johtaminen piirteistä on huomattavasti vaikeampaa kuin pelkästään piirteiden purkaminen datasta ja niiden uudelleenesittäminen tietokannassa [Petković & Jonker 2000]. Antani ja muut [2002, 946] sanovat, että nykyiset järjestelmät pystyvät visuaalisen sisällön osittaiseen kuvailuun. Rasmussenin [1997, 183] mukaan hahmopohjaisilla järjestelmillä ei yleensä edes yritetä analysoida kuvien semanttista sisältöä muutoin kuin rajoitetuissa konteksteissa, sillä automaattisten indeksointimenetelmien

käyttämät algoritmit ovat varsin yksinkertaisia. Bolle ja muut [1998] ottaessaan huomioon konenäön nykytilan, eivätkä pidä ideaa automaattisesta semantiikan johtamisesta kovinkaan toteuttamiskelpoisena tällä hetkellä. Myös Petković ja Jonker [2002] jatkavat samalla linjalla artikkelissaan: Vaikka raakavideodatan automaattinen kartoitus piirrekerrokseksi on jo saavutettu, piirteiden muuttaminen automaattisesti käsitekerrokseksi on yhä haastava ongelma. Semanttisen informaation purkaminen on erittäin monimutkaista ja se vaatii aihetietämystä käsiteltävästä alasta. Visuaalisten piirteiden purkaminen voidaan kuitenkin tehdä automaattisesti, ja se on yleensä alueriippumatonta. [Petković & Jonker 2000.] Semanttisesti korkean tason käsitteet ovat vaikeita ja nykyisellään mahdottomia indeksoida automaattisesti varsinkin, jos semanttinen merkitys on metaforista tai assosiatiivista. Bolle ja muut [1998] ovat varsin pessimistisiä ja epäilevät voidaanko koskaan kehittää järjestelmiä korkean tason semanttisten merkitysten tunnistamiseksi kunnolla.

3.4 Videomallit

Bollen ja muiden [1998] mukaan videodatan käsitteellistämiseksi ('video data abstraction') ja tiivistelmän tai indeksin luomiseksi videosta on tunnistettava syntaktiset ja semanttiset komponentit videomateriaalissa ja määriteltävä tietomalleja ('data model'), jotka suppeasti kuvaavat videon rakenteellisia ominaisuuksia ja semanttista sisältöä. Videosisältö voi olla staattista, kuten jonkin objektin läsnäolo kuvassa, tai dynaamista, jolloin objekti ilmestyy tai menee pois näkyvistä, sekä liikkeeseen perustuvaa, jolloin kyseessä on objektien toiminta tai niiden välinen vuorovaikutus. [Bolle et al. 1998.] Paikkasidonnaisten elementtien sijainnin lisäksi on siis otettava huomioon minkälaiset ajalliset rakenteet liittyvät minkäkinlaisten semanttisten merkitysten tuottamiseen. Esimerkiksi saippuaoopperoissa saatetaan seurata useaa rinnakkain tapahtuvaa tarinaa vuorotellen, mutta TV-uutisten uutisjutuissa ei koskaan. Mikäli videon otokset ovat lyhyitä ja ääniraita on kompressoitu lähelle täyttä äänenvoimakkuutta voidaan päätellä, että kyseessä voi hyvinkin olla nopeatempoinen musiikkivideo. Mallintamisessa hyödynnetään visuaalisen informaation koodaamista tiettyjen konventioiden mukaisesti, mitä varten luvussa 2 käsiteltiin TV:n käyttämiä esittämisen keinoja.

Malli muodostuu piirrejoukosta ja aihetiedosta, joiden avulla edustetaan jotain videovirran ajallista tai paikkasidonnaista elementtiä, objektia tai tapahtumaa. Malleissa ilmaistaan eksplisiittisesti minkälaisia syntaktisia rakenteita ja havaittavia piirteitä liittyy minkäkinlaisten semanttisten merkitysten tuottamiseen erilaisissa konteksteissa. Ajallista rakennetta koskevat mallit ovat sääntöjoukkoja, joissa on ilmaistu videon leikkauksessa tehdyt rat-

kaisut; ne sisältävät tiedon siitä, millä tavalla luvussa 2.1 käsitellyt ajalliset elementit on järjestetty. Sisältöä koskevat mallit koskevat joko objekteja, niiden liikettä tai paikkasidonnaisia sommitelmia. Objekteja koskevat mallit sisältävät tiedon siitä, että minkä muotoisiin ja värisiin sekä minkälaisen tekstuurin sisältämiin alueisiin liitetään minkäkinlaisia semanttisia merkityksiä. Petkovićin ja Jonkerin [2000] mukaan objekteja määrittävät niiden semanttiset, geometriset ja paikkasidonnaiset ominaisuudet sekä suhteet toisiin objekteihin. Semanttiset merkitykset, kuten objektin tyyppi, nimi ja sisältöinformaatio, voidaan yhdistää objektiin aihetietämyksestä. [Petković & Jonker 2000.] Mallissa paikkasidonnaisiin sommitelmiin liitetään tieto siitä, että minkälaisia merkityksiä tietynlaiset objektit saavat sijaitessaan tietyssä paikassa [ks. Petković & Jonker 2000]. Esimerkiksi kun kuvasta tai sen liikealueesta tunnistetaan tiettyyn liikemalliin sopiva joukko piirteitä, voidaan päätellä, että kyseessä on jokin tietty objekti, joka voidaan nimetä, jos piirrejoukkoon voidaan yhdistää ihmisten havaittaville piirteille antamat merkitykset. Mallien luominen edellyttää apriorista kontekstuaalista tietoa eli aihetietämystä siitä aihealueesta, joka halutaan mallintaa. Jotta objekteja tunnistettaessa aihetietämystä voidaan soveltaa objekteihin, on ensiksi tiedettävä konteksti, jossa objektit esiintyvät.

Del Bimbon [1999, 47] mukaan tietomallit ('data models') ja tietämysrakenteet tarjoavat muodollisen tuen visuaalisen sisällön indeksien järjestämiseen abstraktion eri tasoilla. Hän tekee eron (1) yleistettyjen mallien, jotka eivät ota huomioon videoiden erikoisominaisuuksia vaan keskittyvät visuaalisen sisällön esittämiseen käsitteellisellä tasolla, ja (2) dokumenttispesifisten mallien välille, jotka on suunniteltu kunkin videodokumenttityypin erikoisominaisuudet huomioon ottaen. [Del Bimbo 1999, 47.]

3.4.1 Hierarkkiset ja objekteja koskevat yleiset mallit

Del Bimbo [1999, 47–51] käsittelee useita erilaisia yleistettyjä malleja. Eräs yleisimmin kirjallisuudessa mainittuja malleja on Guptan, Weymouthin ja Jainin [1991] sekä Jainin ja Hampapurin [1994] alunperin käsittelemä *VIMSYS* (eli *Visual Information Management System*). VIMSYS on yleinen tietomalli, joka käyttää hierarkkista dataesitystä ja mahdollistaa visuaalisen sisällön abstraktien tasojen eksplisiittisen esittämisen. Mallissa on Del Bimbon [1999, 47–48] ja Lun [1999, 181–183] mukaan seuraavat kerrokset (lueteltuna matalimmasta korkeimpaan):

1. *Kuvanesitysten* taso, joka tallentaa raakaa kuvadataa. Tällä tasolla voidaan käsitellä vain pikseleihin perustuvia kyselyitä, mutta kuvaesitysten taso välittää raakaa dataa myöhemmille tasoille, joilla määritellään ja poimitaan korkeamman tason piirteitä.
2. *Kuvaobjektien* taso on jaettu segmentaation ja piirteiden alatasoihin. (a) Segmentaation alataso tiivistää informaatiota koko kuvasta tai videosta paikkasidonnaisiksi tai ajallisiksi ryppäiksi yhteenvedettyjä ominaisuuksia. (b) Piirrealataso sisältää piirteitä, joita voidaan laskea, organisoida ja joiden välisiä etäisyyksiä voidaan mitata. Yleisiä piirteitä ovat värijakaumat, objektien muodot ja tekstuurit. Näitä piirteitä kohdellaan syntaktisina elementteinä ilman tulkintaa.
3. *Alueobjektitaso* tarjoaa semanttisen tulkinnan datasta aihetietämyksen avulla. Alueobjekti on määritelty olio ('entity'), joka edustaa fyysistä olemusta tai käsitettä, joka voidaan johtaa yhdestä tai useammasta aiempien tasojen piirteistä. Taso assosioi olioita (eli piirteitä), jotka tunnistettiin edellisellä tasolla, objekteihin.
4. *Aluetapahtumataso* määrittää objektien välisiä suhteita ja tapahtumia, joita voidaan kysellä. Tapahtumat voidaan määritellä liikkeen, objektien paikkasidonnaisten ja ajallisten suhteiden, objektien ilmestymisen, häipymisen ja muiden vastaavien ominaisuuksien perusteella. Tässä vaaditaan liikkeentunnistamista. Jokainen objekti tällä tasolla voidaan yhdistää ('associate') minkä tahansa muun objektin kanssa käyttämällä alueobjektia ('domain entity') aluetapahtumien tasolla. [Lu 1999, 181–183; Del Bimbo 1999, 47–48.]

Kaksi ensimmäistä tasoa ovat alueriippumattomia ja selittävät vain videodatan syntaktisia piirteitä ('aspect'); loput kaksi tasoa ovat semanttisia. VIMSYS-tietomallin avulla Del Bimbon [1999, 48] mukaan korkean tason semanttiset käsitteet voidaan helposti kääntää matalan tason visuaalisiksi piirteiksi.

Ponceleonin ja Srinivasanin [2002, 36] jäsennys muodostuu

1. media- ja datakerroksesta
2. piirrekerroksesta
3. objektien tasosta, joka muodostuu lohkoista ('block') ominaisuuksia ja piirteitä
4. käsitteiden tasosta. [Ponceleon ja Srinivasan 2002, 36.]

Malli muistuttaa VIMSYS-mallia: niiden ensimmäiset ja toiset tasot vastaavat toisiaan, mutta Ponceleonin ja Srinivasanin mallin tasot 3 ja 4 sisältyvät VIMSYS-mallin tasoon 3. (Ponceleonin ja Srinivasanin esittämässä mallissa käytetään termiä ”objekti” samassa merkityksessä kuin ”oliota” VIMSYS-mallissa.) VIMSYS-malli sopii siis paremmin tarkoituksiin,

joissa objektien, niiden välisten suhteiden ja liikkeen tarkempi mallintaminen on tarpeellista. Smithin [2001, 971] esitys visuaalisen informaation tasoista sopii erityisesti paikkasidonnaiseen mallintamiseen:

1. *Piirteet*: Havaittavat piirteet välittävät semanttisia merkityksiä. Piirteet ovat kuvien matalan tason havaittavia ominaisuuksia. Niistä tärkeimpiin kuuluvat väri, pinta ('texture') ja muoto.
2. *Rakenne*: Kuvan alueet ('regions') ovat suhteessa toisiinsa rakenteisten suhteiden kautta, joihin kuuluvat kasautuminen ('aggregation'), sommittelu ja assosiointi – esimerkiksi kuvan osa voi sijaita vasemmalle jostain toisesta osasta.
3. *Semantiikka*: Semantiikka on usein kontekstiriippuvaista – esimerkiksi objektilla (esim. jalkapalloilijalla) on suhde (potkiminen) toiseen objektiin (palloon) – joten aihetietämys ja semanttiset ontologiat ovat tärkeitä sisällönkuvailussa. Yleisesti objektit ('entities') [5] ovat suhteessa toisiinsa erilaisten semanttisten suhteiden kautta, jotka kuvaavat objektien keräytymistä ('aggregation'), erikoistumista ('specialization'), assosiaatiota ja vuorovaikutusta. [Smith 2001, 971.]

Edellisten mallien koostamiseksi voidaan esittää, että videoiden ajallispaikallista rakennetta voidaan esittää hierarkkioiden avulla, jotka mallintavat otosten peräkkäisyyttä, objektien ja tehosteiden aiheuttamaa liikettä sekä objektien ja hahmojen sijoittumista paikassa. Kirjallisuudessa käsiteltyjen mallien väliset erot selittyvät muun muassa sillä, että erilaiset käyttötarkoitukset vaativat erilaisia malleja. Edellä käsitellyt mallit koskevat videosisällön yleisiä käsitteellisiä tasoja. Ne kuitenkin käsittävät vain piirteiden johtamista semanttisesta sisällöstä yleisellä tasolla. Videoiden ajallisen rakenteen huomioon ottamiseksi Del Bimbon [mts. 48] mukaan Tonomura, Akutsu, Taniguchi ja Suzuki [1994] ovat kehittäneet kaksitasoisen videomallin, jossa on

- linkkien rakenteen ('link structure') taso, joka tallentaa otosten välisiä suhteita (niiden sekvensointia ja siirtymien tyyppejä)
- sisällön rakenteen taso, joka käsittelee otosten sisältöä, tapahtumia, kameran liikettä ym.

Tämä malli sopii hyvin erityisesti ajallisten elementtien (kuten otosten ja kohtausten) ja niiden ominaisuuksien mallintamiseen. Otosten välisten linkkien mallintaminen on tärkeää, sillä

5 Kirjallisuudessa käytetään termejä 'entity' ja 'object' joskus samassa merkityksessä ja joskus eri merkityksessä. Jos niiden välille halutaan tehdä ero, olio on yleisnimitys paikkakäsitteellisille elementeille, kuten objekteille, pisteille, alueille ja viivoille [ks. Del Bimbo 1999, 26].

kaikki sisältö on videon rakenteellisen epäjatkuvuuden välittämää: objektit ja niiden liike esiintyvät otosten katkonaisuuden ylitse. Otoksista voidaan muodostaa käsitteellisesti korkeampia ajallisia yksikköjä ohjelmatyyppejä koskevien mallien avulla.

3.4.2 Uutislähetyksen spesifi ajallispaikallinen malli

Edellä mainittiin, että kaikki semanttiset merkitykset ovat havaittavien piirteiden välittämiä. Lisäksi esitettiin yleisiä tietomalleja, jotka jäsentävät ne tasot, joilla piirteistä voidaan johtaa semanttisia käsitteitä. Tiettyjä dokumenttityyppejä varten voidaan rakentaa apriorisia malleja analysoimalla näiden dokumenttityyppien esittämisen käytäntöjä: esimerkiksi TV-uutisten sisäistä esteettistä ja kerronnallista logiikkaa tutkimalla voidaan tehdä niiden rakennetta kuvaavia malleja. Bolle ja muut [1998] mainitsevat, että TV-uutisten ritualisoitua rakennetta hyödyntäviä malleja on onnistuneesti käytetty tunnistamaan yksittäisiä uutisia uutislähetyksistä. Del Bimbon [1999, 51] mukaan Swanberg, Shu ja Jain [1993] ovat määrittäneet videomallin, joka perustuu uutislähetysten ajalliseen järjestykseen. Tässä mallissa video on käsitteellistetty raa'aksi ja rakenteiseksi videovirraksi. Raaka video käsitetään kuvasekvenssiksi, joka muodostuu leikkeistä ja yksittäisistä kehyksistä elementteinä. Rakenteisessa videovirrassa erotetaan otokset, otoksen tyypit, kohtaukset ja jaksot. Jokaista otostyyppiä varten rakennetaan malli, jossa osoitetaan keskeisten hahmojen paikkasidonnainen asema avainkehyksissä ja näiden kehysten ajallinen järjestys. Del Bimbon [1999, 52] esimerkissä uutisankkuria koskeva otos muodostuu kahdesta avainkehyksestä, joista ensimmäisessä uutisankkuri sijaitsee keskellä ruutua, ja ruudussa on uutisankkurin nimi ja uutislähetyksen otsikko; toisessa otoksessa ankkuri sijaitsee hieman reunempana, uutisotsikko on eri kohdassa, ja uutisikkuna on ankkurista hieman yläviistoon. Tietty jakso uutislähetyksessä tunnistetaan otostyyppien järjestyksen avulla. [Del Bimbo 1999, 51–52.]

Bolle ja muut [1998] kirjoittavat tarinan yksikköjen tyyppien tunnistamisesta ja siitä, että ne voidaan luokitella automaattisesti esimerkiksi dialogiksi tai toiminnaksi toisteisuuden tai sen puuttumisen perusteella – tässä voidaan käyttää apuna juuri edellä mainittua otostyyppien toistumista. Dialogilla tarkoitetaan keskustelua ja keskustelunkaltaista kahden tai useamman samanaikaisen prosessin montaasiesitystä. Toiminnalla tarkoitetaan esimerkiksi toimintaelokuvien kohtauksia, joihin kuuluvat muun muassa nopeat leikkaukset ja muut vastaavat piirteet. Jo näiden kahden tyypin avulla voidaan tehdä päätelmiä ohjelmatyypistä. [Bolle et al. 1998.] Tällä tavalla uutislähetyskin voidaan tunnistaa: se sisältää sellaista otostyyppien toistumista, mitä ei löydy monista muista ohjelmatyypeistä. Bolle ja muut [1998] li-

säävät vielä, että leikkaajilla on käytössään useita peukalosääntöjä, joiden avulla he liimaavat otokset yhteen jatkuvaksi kertomukseksi. Esimerkiksi kahden henkilön puhuessa toisilleen kasvokuvien välityksellä, he katsovat suoraan toisiinsa; tämän lisäksi ohjaaja vielä muistuttaa paikkasidonnaisesta asemasta käyttämällä laajempaa kuvaa. [Bolle et al. 1998.] Lisäksi Del Bimbon [1999, 225] mukaan kokonaisuuksien tunnistamista helpottaa se, että leikkaamisessa tehosteita käytetään usein tuottamaan merkityksiä: esimerkiksi häivytyksiä, pyyhkäisyjä ja himmennyksiä käytetään osoittamaan muutosta kontekstissa (eli ajassa ja paikassa), ja liuotuksia käytetään kuin sivuhuomauksia tekstuaalisessa mediassa, esimerkiksi takaumien osoittamiseen. [Del Bimbo 1999, 225.] Siihen pitääkö tämä paikkansa TV-uutisten yhteydessä, palataan tarkemmin luvussa 6. Uutisten mallintamisen menetelmiä käsitellään tarkemmin luvussa 4.

3.5 Indeksointijärjestelmän pääpiirteet

Karkeasti digitaalisen videon indeksointiprosessi voidaan jakaa seuraaviin osiin: (1) yksittäisistä kuvista muodostuvan kuvavirran jäsentämiseen (segmentointiin) otoksiin ja avainkuvien valitsemiseen otoksista, (2) otosten järjestämiseen hierarkiaksi ja (3) sisällön tunnistamiseen eri tasoilla poimimalla videon havaittavista ominaisuuksista piirteitä ja mahdollisesti johtamalla niistä käsitteitä. [Ahanger & Little 1995; Gupta ja Jain 1997, 75; Xiong et al. 1997, 51.]

Tallennusprosessissa ('storage process') videodatasta poimitaan piirteitä, jotka edustavat videon semantiikkaa. Piirteet organisoidaan ja tallennetaan tietokantaan. Ponceleon ja Srinivasan [2002, 5] erottelevat ilmaisuun eli esimerkkikuviin perustuvat haut, jotka suoritetaan matalan tason piirteillä, ja semantiikkaan pohjautuvat haut, joissa pyritään ratkaisemaan piirteiden ja semantiikan välistä käsitteellistä etäisyyttä. Hakuprosessissa, jos kysely tehdään esimerkkikuvalla tai hahmotelmalla, järjestelmä analysoi kyselyn ja purkaa tarkoituksenmukaiset piirrevektorit, jonka jälkeen kyselyn ja tietokannan videodatan piirrevektorit täsmäytetään niiden samankaltaisuutta eli etäisyyttä vertailemalla. Osittaistäsmäyttävä järjestelmä järjestää tulokset ja tämän jälkeen arvottaa tulosjoukon samankaltaisuusarvon ('similarity score') perusteella ja typistää sen tietyssä katkaisu-arvossa ('cutoff value'). [Idris & Panchanathan 1997, 147; Ponceleon & Srinivasan 2002, 5; Smith 2001, 970; Sormunen, Markkula & Järvelin 1999, 2–3.] Hakuoperaatioita ei siis suoriteta kuville suoraan vaan indeksoiduille visuaalisille piirteille, jotka on abstraktoitu kuvista kuvankäsittelymenetelmien avulla [Rasmussen 1997, 183].

Kirjallisuudessa esitetään videohakujärjestelmälle useitakin erilaisia arkkitehtuu-

reja. Del Bimbo [1999, 11–12] käsittelee videotiedonhakujärjestelmän arkkitehtuuria, joka sisältää alijärjestelmiä ja välineitä muun muassa seuraavia tarkoituksia varten:

1. Leikkaustehosteiden purkaminen ja luokittelu; kuvan- ja hahmonanalysointimoduuleja käytetään segmentoimiseen ja siirtymätehosteiden havaitsemiseen
2. Matalan- ja kestitason piirteiden poimiminen otoksista kuvananalysointivälineillä
3. Kohtauksien ja tarinoiden poimiminen sääntöjen ja aikaisemman tietämyksen avulla
4. Selailu- ja visualisointiväline
5. Visuaalinen kyselyväline graafisia ja visuaalisia kyselyitä varten
6. Indeksointirakenne, joka mahdollistaa valikoivan pääsyn videon ja sen sisällön eri elementteihin
7. Sisältöpohjainen hakumoottori. [Del Bimbo 1999, 11–12.]

Idris ja Panchanathan [1997, 147–148] esittelevät yleisen mallin, joka koostuu seuraavista hierarkkisesti järjestetyistä komponenteista:

1. *Käyttöliittymä*: Kyselynkäsittelijä ('query processor'), joka tarjoaa useita menetelmiä ja käyttöliittymiä ('interface') kyselyjä varten, ja selain, jolla voidaan selata kyselyn tuloksia, ovat käyttöliittymän osia.
2. *Sisältöpohjainen hakumoduuli*: Moduuliin kuuluu (1) otosten- ja kohtausten rajojen tunnistaminen, (2) kuvan esikäsittely (kuvaa käsitellään piirteiden purkamisen helpottamiseksi) ja (3) piirteiden purkaminen ja esittäminen.
3. *Organisointi*: Kyselyjen tehokas käsitteleminen edellyttää videoindeksien järjestelemistä. Perinteiset indeksointirakenteet eivät sovi videoihin. Niiden sijaan Idris ja Panchanathan [1997, 148] mainitsevat joustavat R- ja R*-puut, nelihaaraisen puumallin ('quad-tree') ja ruudukkotiedostot. Eri indeksointirakenteilla on etunsa ja haittansa, joten niitä käytetään eri tilanteissa [Idris & Panchanathan 1997, 148].
4. *Tietokannanhallintamoduuli*: Moduuli tarjoaa fyysisen tallennusrakenteen ja pääsyn tietokantaan. [Idris & Panchanathan 1997, 147–148.]

Tästä Idrisin ja Panchanathanin [1997, 147–148] yleisestä mallista käsitellään tässä tutkielmassa kahta ensimmäistä tasoa. Del Bimbon [1999, 11–12] aiemmin luettelemia elementtejä käsitellään tarkemmin seuraavissa luvuissa.

4 INDEKSOINTITEHTÄVÄT

Edellisissä luvuissa esiteltiin videoiden automaattiseen indeksointiin liittyvää käsitteistöä ja teoreettista taustaa. Automaattista TV-uutisten indeksointijärjestelmää suunniteltaessa on selvitettävä mihin tarkoitukseen järjestelmä tulee ja mitä tehtäviä sen halutaan suorittavan. Tehtävällä voidaan tarkoittaa yksinkertaisimmillaan esimerkiksi videosekvenssin jakamista otoksiin. Monimutkaisimmat tehtävät liittyvät semanttisten tulkintojen tekemiseen videon objekteista. Kun tiedetään vaadittavat tehtävät, otetaan selvää, mitä niiden toteuttaminen vaatii. Brunellin ja muiden [1999, 79] mukaan käyttötarkoitukset ja niiden toteuttamisen ongelmat ovat moninaisia, koska hakijat ovat usein kiinnostuneet varsin erilaisista asioista: videon rakenteellisista ominaisuuksista eli siitä, kuinka video jakautuu otoksiin ja kohtauksiin, otoksia parhaimmin edustavista avainkehyksistä eli yksittäisistä kuvista tai kuvasekvensseistä, niissä esiintyvistä ihmisistä, heidän liikkeistään ja vuorovaikutuksestaan sekä näihin liittyvästä musiikista ja äänitehosteista [Brunelli et al. 1999, 79]. Suurin vaikutus indeksointijärjestelmän luonteeseen tulee siitä, onko indeksoitava materiaali valmiiksi tuotettua (kuten TV-ohjelmat) vaiko raakamateriaalia, jossa ei ole leikkaustehosteita, grafiikkaa tai muita indeksoimista helpottavia tehosteita. Tässä tutkielmassa keskitytään nimenomaan valmiisiin uutisiin. Käyttäjiä ja heidän erilaisia vaatimuksiaan käsitellään lisää luvussa 5.

Tutkielmassa edetään samankaltaisella tavalla kuin Whittaker, Hirschberg, Choi, Hindle, Pereira ja Singhal [1999, 27], jotka puhedokumenttien selailun ongelmia käsittelevässä artikkelissaan tutkivat käyttäjien äänipostin ('voicemail') käyttöä ja tekivät johtopäätöksiä siitä, millaisen käyttöliittymän puhedokumenttien selailu tarvitsee. Markkulan [2002] toistaiseksi julkaisemattoman TV-toimituksen nykyisiä käytäntöjä ja niihin liittyviä ongelmia esittelevän tutkimuksen esiraportin avulla on muodostettu indeksointitehtäviä. Tässä luvussa käsitellään näitä indeksointitehtäviä ja niiden toteuttamista automaattisia indeksointimenetelmiä koskevan kirjallisuuden avulla seuraavan jäsennyksen avulla:

1. *Segmentointi ja ajallisen rakenteen jäsentäminen*: Videovirrasta on eroteltava (eli segmentoitava) otokset ja niitä laajemmat yksiköt haettavien yksiköiden muodostamiseksi.
2. *Objektityyppien jäsennys ja objektien tunnistaminen*: Järjestelmän on eroteltava objektit ja tausta toisistaan, jäsennettävä erityyppiset objektit, tunnistettava kasvoja ja kuvatekstejä (eli tärkeimmät objektit uutisvideoissa) sekä tunnistettava suhteellinen kuvakulma

ja -etäisyys.

3. *Liikkeen ja tapahtumien havaitseminen ja tunnistaminen*: Järjestelmän on havaittava liike otoksissa eli luokiteltava otokset staattisiksi tai dynaamisiksi ja tunnistettava liikkeen tyyppi (kameran liike vaiko havaittujen objektien liike) sekä liikkeen suunta.
4. *Ääniraidan jäsennys ja tunnistaminen*: Järjestelmän on jäsennettävä ääniraita tyyppeihin (eli puheeseen, taustaääniin ja musiikkiin) ja tunnistettava puhetta, puhuja ja käytetty kieli.
5. *Visualisointi*: Järjestelmän on tiivistettävä videodata ja mahdollistettava sujuva tiedonhaku erilaisten visuaalisten menetelmien avulla ja tuettava myös ääniraidan selailua.

Tässä luvussa tarkastellaan näistä vaatimuksista neljää ensimmäistä. Videosisältöjen visualisointia käsitellään videotiedonhaun yhteydessä luvussa 5.

4.1 Segmentointi ja ajallisen rakenteen jäsentäminen

> Videovirrasta on eroteltava (eli segmentoitava) otokset ja niitä laajemmat yksiköt haettavien yksiköiden muodostamiseksi.

Videoindeksin pitää pystyä esittämään videon ajallispaikallinen sisältö, johon kuuluu yksittäisten kehyksien paikkasidonnaisten elementtien (eli objektien sijainnin ja sommittelun) lisäksi ajallisia ominaisuuksia kuten kameran ja objektien liikettä sekä kehyksien välisiä objektien suhteita. Ennen sisällön ja sen objektien tunnistamista video on jäsennettävä lyhimpiin mahdollisiin merkityksellisiin ja jatkuvana toimintana näkyviin alijaksoihin; kuvavirran osittaminen ja videon syntaktisen rakenteen tunnistaminen on ensimmäinen askel videon ymmärtämisessä ja videoiden indeksoinnin edellytys [Apers et al. 1998, 172; Bolle et al. 1998; Idris & Panchanathan 1997, 154; Roth 1998; Xiong et al. 1997, 52]. Segmentointi on prosessi, jossa videon jatkuva ja rakenteeton kuvavirta jäsennetään merkityksellisiin ja hallittavissa oleviin yksiköihin: otoksiin ja niiden korkeamman tason yhdistelmiin ('aggregates') kuten jaksoihin ja kohtauksiin. Segmentointia voi verrata fraasien tunnistamiseen tekstidokumentista. Samalla tavalla kuin tekstidokumenttien avainsanat ja -fraasit viittaavat lauseisiin, kappaleisiin, sivuihin ja dokumentteihin, videon indeksoinnissa avainkehykset ja -sekvenssit viittaavat korkeamman tason yksiköihin kuten kohtauksiin ja tarinoihin. [Brunelli et al. 1999, 80; Del Bimbo 1999, 203; Idris & Panchanathan 1997, 146, 154.] Automaattiset segmentoinnin menetelmät perustuvat kuvananalyysitekniikoihin ja sääntöjoukkoihin, joiden tarkoituksena on mal-

lintaa olosuhteet, joissa tietyt asiat tapahtuvat [Del Bimbo 1999, 203]. Koska asiat tapahtuvat ennen kaikkea ajassa, segmentoinnilla tarkoitetaan siis videon ajallisen rakenteen jäsentämistä. (Paikkasidonnaisesta jäsentämisestä lisää myöhemmin objektien tunnistamisen yhteydessä.) Yksinkertaisimmillaan segmentoinnilla tarkoitetaan otosrajojen löytämistä videovirrasta. Myös korkeamman tason semanttisten elementtien tunnistamista on tutkittu hieman [Del Bimbo 1999, 204].

4.1.1 Otosten tunnistaminen

Otosten tunnistaminen on perustava vaihe videoiden indeksoinnissa. Perinteisessä lähestymistavassa raaka videovirta segmentoidaan sarjaksi otoksia automaattisten otosrajojentunnistamismenetelmien avulla, minkä jälkeen segmentoiduista otoksista poimitaan avainkehyksiä, joista muodostetaan videon sisällysluettelo [Rui et al. 1999, 360; ks. Del Bimbo 1999, 203]. Otosten väliset rajat voivat olla välittömiä ('sharp') eli leikkauksia tai asteittaisia ('gradual'), ja ne voidaan tunnistaa automaattisten menetelmien avulla [Bolle et al. 1998; Brunelli et al. 1999, 81; Del Bimbo 1999, 203–204; Yeo & Yeung 1997, 47]. Apersin ja muiden [1997, 172] mukaan useimmat algoritmit tunnistavat otosrajat videon peräkkäisten kehysten välisten tiettyjen ominaisuuksien epäjatkuvuudesta. Menetelmät voidaan jakaa suoraan pakatussa kuin pakkaamattomassa videodatassa toimiviin [Yeo & Yeung 1997, 47]. Varsinkin MPEG-pakkausjärjestelmään perustuvat menetelmät ovat käytännöllisiä, koska suuri osa videodatasta on valmiiksi pakatussa muodossa. Seuraavaksi käsitellään otosrajojen tunnistamista.

4.1.1.1 Välittömien siirtymien tunnistaminen

Välitöntä siirtymää kahden otoksen välillä kutsutaan leikkaukseksi. Leikkaukset ilmenevät äkillisinä muutoksina kirkkauskuviossa ('brightness pattern') kahden peräkkäisen kehyksen välillä. Periaatteellisella tasolla automaattinen leikkausten tunnistaminen perustuu informaatioon, joka poimitaan ('extract') peräkkäisistä otoksista, joiden välillä leikkaus tapahtuu. Oletuksena on, että otoksen sisällä kahden peräkkäisen kehyksen välillä ei ole suuria eroja taustassa ja objekteissa, joten yleinen kirkkauden jakautuminen eroaa vain vähän otoksien sisäisten kehysten välillä. Otosrajojen tunnistaminen äkillisten kirkkauden muutosten perusteella on helppoa, jos peräkkäisissä otoksissa on vähän liikettä ja tasainen valaistus; ongelmia voi ilmaantua, jos otoksen sisällä esiintyy jatkuvaa objektien liikettä, kameran liikettä tai muutoksia valaistuksessa (esimerkiksi salamavalo). Tällaisissa tilanteissa on vaikeaa määritellä johtu-

vatko muutokset paikallisten olosuhteiden muutoksesta otoksen sisällä vaiko otoksen vaihtumisesta. Tätä ongelmaa pyritään kiertämään korostamalla otoksen visuaalisia ominaisuuksia kuvankäsittelymenetelmillä ja jakamalla yksittäiset kehykset alikehyksiin, joita käsitellään erikseen. [Del Bimbo 1999, 204–205.] Otosrajojen välittömiä siirtymiä tunnistavat algoritmit voidaan jakaa karkeasti paitsi pakkaamattomalla ja pakatulla datalla toimiviin myös globaale- ja paikallisia piirteitä käyttäviin algoritmeihin. Algoritmeja käsitellään seuraavaksi edellä mainituissa järjestyksissä.

Pakkaamattomalla datalla toimivat ja välittömiä siirtymiä tunnistavat algoritmit voidaan luokitella (1) värien intensiteettikaavioiden ('intensity/color template') täsmäytysmenetelmiin, jotka perustuvat kehysten välisten pikseleiden vertailuun, (2) histogrammipohjaisiin menetelmiin, jotka esittävät kuvan arvojen eli värin tai valotiheyden ('luminance') jakautumista kehyksessä ja (3) lohkoihin ('block') eli kehyksen alialueisiin perustuviin tekniikoihin. [Brunelli et al. 1999, 81; Idris & Panchanathan 1997, 154–157; Ponceleon & Srinivasan 2002, 22; Rui et al. 1999, 360.]

Värien intensiteettikaavioita käytetään kahden kehyksen välisen paikkasidonnaisen samankaltaisuuden vertailuun. Jokaista tietyssä kohdassa kehystä sijaitsevaa pikseliä verrataan seuraavan kehyksen vastaavaan kohtaan: jos muuttuneiden pikseleiden määrä on suurempi kuin määritelty kynnys, otosraja tunnistetaan. Pikseleiden vertailuun perustuvat menetelmät ovat herkkiä hälylle, objektien liikkeelle ja kameran käytölle, koska on vaikeaa erottaa suurella alueella tapahtuva pieni muutos pienellä alueella tapahtuvasta suuresta muutoksesta. Liike aiheuttaa usein vääriä otosrajan tunnistuksia. [Antani et al. 2002, 955–956; Del Bimbo 1999, 205; Idris & Panchanathan 1997, 155; Ponceleon & Srinivasan 2002, 22–23; Rui et al. 1999, 360; Xiong et al. 1997, 52.]

Otosrajoja voidaan etsiä myös vertailemalla kahden kehyksen välisten harmaataso- ja värihistogrammien eroja. (Histogrammit ovat väri-intensiteettijakaumia.) Vertailussa otetaan huomioon, että otoksen objektien ja taustan ollessa samat, histogrammissa on vähän muutoksia; jos ero ylittää asetetun kynnyksen, otosraja havaitaan. Kuvan pyöriminen, kuvakulman ja skaalan ('scale') muutos ja tukkeutuminen ('occlusion') aiheuttavat vain pieniä muutoksia histogrammiin, sillä ne eivät ota huomioon ollenkaan kuvien paikkasidonnaisia ominaisuuksia. Histogrammit ovat siis vähemmän herkkiä kameran ja objektien liikkeille kuin pikseleihin perustuvat menetelmät. Histogrammipohjaiset menetelmät käyttävät otosrajojen tunnistamisessa myös etukäteen säädettyjä kynnyksiä, jotka asetetaan sisällön perusteella. Kynnyksistä hieman lisää asteittaisten siirtymien käsittelyn yhteydessä. [Antani et al. 2002, 956; Idris & Panchanathan 1999, 155; Prabhakaran 1997, 77; Xiong et al. 1997, 52.] Histogrammipohjaiset segmentointi menetelmät toimivat hyvin ja niitä käytetään paljon, koska niiden globaali

luonne altistaa menetelmät vähemmän otosten sisäisille muutoksille [Brunelli et al. 1999, 81–82; Ponceleon & Srinivasan 2002, 23–25; Rui et al. 1999, 360]. Idris ja Panchanathan [1997, 156] tosin huomauttavat, että histogrammeihin perustuvat menetelmät saattavat jättää tunnistamatta otosrajan, mikäli intensiteettijakaumissa otosten välillä on vain vähäisiä muutoksia. Histogrammeihin perustuvat menetelmät eivät siis ole välttämättä ylivoimaisesti parempia kuin intensiteettikaavioiden täsmäyttämiseen perustuvat menetelmät.

Edellä käsitellyt tilastolliset menetelmät käyttävät otosrajojen tunnistamiseen kehyksien välisiä globaaleja piirteitä, kuten pikseleiden välisiä eroja tai värihistogrammeja. Lohkoihin pohjautuvat menetelmät käyttävät paikallisia piirteitä hälyn ja kameran salamavalojen vaikutusten vähentämiseksi. Ideana on osittaa kehys lohkoihin eli alikehyksiin ja sen sijaan, että vertailtaisiin kokonaisia kehyksiä, vertaillaan kahden kehyksen vastaavia alikehyksiä keskenään. Otosraja tunnistetaan mikäli muuttuneiden lohkojen määrä on tarpeeksi suuri. [Antani et al. 2002, 956; Idris & Panchanathan 1997, 156; Xiong et al. 1997, 52.] Ponceleon ja Srinivasan [2002, 23] mainitsevat esimerkiksi, että pikseleiden vertailuun perustuvia menetelmiä voidaan parantaa käyttämällä 3*3 pikselin keskiarvottavaa ('averaging') suodinta ennen vertailua ja jakamalla kuva esimerkiksi kahteentoista erikseen vertailtavaan alueeseen. [Ponceleon & Srinivasan 2002, 22–23; ks. Rui et al. 1999, 360.] Del Bimbo [1999, 208] ja Prabhakaran [1997, 77] mainitsevat menetelmästä, jossa globaalien histogrammipohjaisten menetelmien toimintaa on parannettu osittamalla kehys lohkoihin ja vertaamalla kahden peräkkäisen kehyksen vastaavien (myös ikkunoiksi kutsuttujen) alikehyksien välisten värihistogrammien eroja. Lohkoihin pohjautuvat menetelmät vähentävät ylitunnistamista eli ylimääräisten otosrajojen löytämistä; vääriä leikkauksia voidaan kuitenkin edelleen tunnistaa sellaisten kehysten välille, joilla on samankaltaiset pikseliarvot, mutta erilaiset tiheyden ('density') funktiot [Idris & Panchanathan 1997, 156].

Ongelmien ratkaisemiseksi on ehdotettu myös intensiteettitilastoja ('intensity statistics') otosrajojen tunnistamisen mittarina [Rui et al. 1999, 360]. Lohkopohjaisiin menetelmiin kuuluu myös todennäköisyyskerroin ('likelihood ratio'). Tässä menetelmässä tarkastellaan kahden kehyksen välisiä vastaavia alikehyksiä ja niiden intensiteettiarvojen toisen tason ('second order') tilastoja. Otosraja tunnistetaan, jos suurin osa lohkoista osoittaa ('exhibit') suurempia todennäkyisyyskertoimia kuin mitä ennalta määritelty kynnys. [Antani et al. 2002, 956; Brunelli et al. 1999, 82; Del Bimbo 1999, 206; Xiong et al. 1997, 52.] Antani ja muut [2002, 956] viittaavat vielä erääseen kirjallisuudessa käsiteltyyn menetelmään, jossa käytetään histogrammierojen mittaa ('histogram difference metric' eli HDM) ja paikkasidonnaisten erojen mittaa ('spatial difference metric' eli SDM). Menetelmä olettaa, että otos vaihtuu, kun kummatkin näistä arvoista ovat suhteellisen suuria. Brunelli ja muut [1999, 82–84] ja Del

Bimbo [1999, 205–211] käsittelevät yksityiskohtaisesti pakkaamattomassa datassa toimivia algoritmeja leikkausten tunnistamiseen.

Pakatussa datassa toimivat leikkauksentunnistamisalgoritmit hyödyntävät videoformaattien pakkausta. Algoritmit jakautuvat diskreetteihin kosinimuunnoskertoimiin ('Discrete Cosine Transform' eli DCT), liikevektoreihin, liikevektoreiden ja DCT-kertoimien yhdistelmään sekä alikaistan hajotukseen ('subband decomposition') [Idris & Panchanathan 1997, 157].

JPEG:n ja MPEG:n kaltaiset kuvan- ja videon pakkausstandardit ovat DCT-pohjaisia tekniikoita. DCT-kertoimet kantavat mukanaan informaatiota, jota voidaan käyttää leikkausten havaitsemiseen. Kuvanpakkausmenetelmissä (MPEG ja MJPEG) videokehys ryhmitellään 8 * 8 -pikselin yksiköihin, joihin DCT-kertoimia sovelletaan. Kertoimet ovat matemaattisesti suhteessa paikkasidonnaiseen alueeseen, ja siten ne edustavat kehysten sisältöä. Tästä syystä DCT-kertoimia voidaan käyttää otosrajojen tunnistamiseen pakatussa videodatassa vertailemalla otosten DCT-kertoimia: jos kertoimet eroavat toisistaan paljon, ne eivät todennäköisesti kuulu samaan otokseen. [Antani et al. 2002, 956; Brunelli et al. 1999, 84; Idris & Panchanathan 1997, 157; Prabhakaran 1997, 77–78; Rui et al. 1999, 360.] Jotkin DCT-pohjaiset algoritmit ovat herkempiä asteittaisille muutoksille kuin toiset ja jotkut, vaikka ovatkin nopeita laskea, ovat taipuvaisia tunnistamaan väärin leikkauksia samalla tavalla kuin pikseleiden vertaamiseen perutuvat menetelmät [Del Bimbo 1999, 211–212; Idris & Panchanathan 1997, 158].

Otosrajojen tunnistamiseen voidaan käyttää myös MPEG-virtaan sulautettuja liikevektoreita [Rui et al. 1999, 360]. Liikevektorilla tarkoitetaan vektoria, jonka havaittava piirre luo siirtyessään kehyksen yhdestä paikasta toiseen paikkaan seuraavassa kehyksessä. Liikkeen arvioiminen ja kompensaatio ovat keskeisessä roolissa videon pakkaamisessa, jossa hyödynnetään ajallista toisteisuutta peräkkäisten kehyksien välillä. Otosraja tunnistetaan arvioimalla kahden peräkkäisen kehyksen välisten samankokoisten lohkojen paikaltaan siirtymistä eli liikevektoreita: liikevektoreiden muutokset otoksen sisällä ovat yleensä jatkuvia, mutta jatkuvuus katkeaa otoksen vaihtuessa. [Idris & Panchanathan 1997, 158; Prabhakaran 1997, 79–80.]

Idris ja Panchanathan [1997, 159] esittelevät menetelmän, jossa ajallista segmentointia sovelletaan pakatun videon alimmalla alikaistalla. Tätä varten on tutkittu neljää mittaa ('metrics'): (1) histogrammien eroja, jotka eivät ole herkkiä objektien liikkeelle, mutta herkkiä kameran liikkeelle kuten panoroinnille ja zoomaukselle; (2) erokehyksien ('difference frame') histogrammeja, jotka ovat herkkiä objektien liikkeelle; (3) lohkohistogrammien eroja ('block histogram difference'), jotka ovat herkkiä paikalliselle objektien liikkeelle; (4) lohkovaihtelun

eroja ('block variance difference'), jotka ovat herkkiä paikalliselle objektien liikkeelle. [Idris & Panchanathan 1997, 159.] Brunelli ja muut [1999, 84–85] käsittelevät lisää pakatussa datassa toimivia leikkauksien tunnistamisen algoritmeja.

4.1.1.2 Asteittaisten siirtymien tunnistaminen

Videon leikkauksessa käytettävät tehosteet levittävät siirtymän kahden otoksen välillä useamman kehyksen alueelle, jolloin siirtymätehoste muodostaa oman lyhyen videosekvenssinsä aloitus- ja lopetuskehyksineen [Del Bimbo 1999, 212]. Siirtymätehosteita kohdellaan sekvensseinä itsessään, ja ne pyritään tunnistamaan samalla tavalla kuin otokset. Yleisimpien asteittaisten tehosteiden eli häivytysten ('fade'), liuotusten ('dissolve'), himmennysten ('matte') ja pyyhkäisyjen ('wipe') lisäksi on olemassa kymmeniä erilaisia tehosteita. Asteittaiset siirtymät eivät kuitenkaan ole yhtä yleisiä kuin leikkaukset. [Brunelli et al. 1999, 85; Del Bimbo 1999, 203–204.]

Häivytys on optinen prosessi, jossa otos vähitellen tummennetaan kunnes viimeisestä kuvasta tulee täysin musta – tai päinvastoin valaistaan musta ruutu valkeaksi [Del Bimbo 1999, 212; Ponceleon & Srinivasan 2002, 20]. *Liuotus* on voimistamisen ('fade-in') ja häivyttämisen ('fade-out') päällekkäiskuvaus. Kumpikin tehoste näkyy samaan aikaan, mutta aiempi otos himmenee mustaksi ja uusi tulee näkyviin samanaikaisesti [Del Bimbo 1999, 213].

Pyyhkäisyt ('wipe') ja *himmennykset* ('matte') ovat paikkasidonnaisia tehosteita. Pyyhkäisyssä – joka voi olla horisontaalinen tai vertikaalinen – seuraavan otoksen ensimmäisen kuva työntää edellisen otoksen viimeisen kuvan asteittaisesti ulos ruudulta [Del Bimbo 1999, 219–220; Ponceleon & Srinivasan 2002, 20]. Himmennyksessä visuaalinen kenttä tummennetaan tai sameutetaan asteittaisesti yleensä ruudun täyttävän ympyrän muotoisesti. Himmennyksiä käytettiin paljon mykkäfilmeissä ja animaatioissa, mutta nykyään ne ovat varsin harvinaisia. [Del Bimbo 1999, 220.]

Peräkkäisten kehysten välisiä eroja vertailevat menetelmät eivät toimi asteittaisissa siirtymissä, koska muutokset kehysten välillä ovat leikkaustehosteita käytettäessä liian pieniä [Brunelli et al. 1999, 85]. Lee ja Smeaton [1999, 10] mainitsevat joidenkin indeksointijärjestelmien mahdollistavan kynnyksen ('threshold') säätämisen ennen automaattista segmentointia, mikä tarkoittaa, että otosrajoja havaitsevan algoritmin herkkyyttä säädetään indeksoitavalle materiaalille sopivaksi [Lee & Smeaton 1999, 10]. Asteittaisia siirtymiä varten käytetään yleensä kahta kynnystä: toinen on globaali ja tarkoitettu välittömiä siirtymiä varten, ja

toisen avulla etsitään siirtymätehosteita. [Ponceleon & Srinivasan 2002, 23–25; Rui et al. 1999, 360.] Seuraavaksi käsitellään yleisimpiä erityyppisten asteittaisten siirtymien tunnistamiseen tarkoitettuja menetelmiä.

Kaksoisvertailumenetelmässä otetaan erityisesti huomioon, että kehykset ennen ja jälkeen asteittaisen siirtymän ovat yleensä huomattavasti erilaisia, vaikka erot kehysten välillä siirtymän sisällä ovatkin pieniä. Histogrammien erot peräkkäisten reunojen välillä asteittaisessa siirtymässä ovat pienemmät kuin välittömässä siirtymässä, mutta suuremmat kuin normaalisti otoksen sisällä. Ensimmäisellä vertailukerralla käytetään korkeaksi asetettua kynnystä välittömien siirtymien tunnistamiseksi. Tässä vaiheessa tunnistettuna on otoksen ja toistaiseksi tunnistamattoman asteittaisen siirtymän yhdistelmä. Toisella vertailukerralla edellä tunnistettuun sekvenssiin sovelletaan vähennettyä kynnystä ja pyritään tunnistamaan potentiaalinen aloituskohta asteittaiselle siirtymälle, jota verrataan seuraaviin kehyksiin kasaantuvien erojen mittaamiseksi. [Apers et al. 1997, 172–173; Brunelli et al. 1999, 88; Del Bimbo 1999, 215; Idris & Panchanathan 1997, 156.] Idris ja Panchanathan [1997, 158] kertovat menetelmästä, joka on liikevektoreiden ja valotiheyden ('luminance') DCT-kertoimien risteytys. Kaksivaiheisen ('two-pass') algoritmin ensimmäisellä käyttökerralla merkitään epäillyt otosrajat, ja toisella kerralla kaikki liuotusalueelle sijoittuvat epäillyt kehykset jäävät merkitsemättömäksi. Kaikki merkityt kehykset tutkitaan. Jos merkittyjen kehyksien ja edellisen otosrajan välinen ero ylittää kynnyksen, merkitty kehys on oikea otosraja. [Idris & Panchanathan 1997, 158–159.]

Tasanteen ('plateau') tunnistamisessa otetaan huomioon se, että asteittaisissa siirtymissä peräkkäisten kehysten välinen ero ei ole tarpeeksi suuri otosrajojen tunnistamiseksi. Ongelma pyritään ratkaisemaan valitsemalla vertailtavaksi kaksi otosta pitkältä aikaväliltä ja suorittamalla niille yhdenmukainen ('uniform') ajallinen alinäytteistäminen ('subsampling') vertailemalla kehyksiä tietyn matemaattisen kaavan mukaan sen sijaan, että verrattaisiin vain peräkkäisiä kehyksiä keskenään [Brunelli et al. 1999, 85].

Otosrajojen tunnistaminen mallintamalla ('detection by modelling') perustuu asteittaisten siirtymätehosteiden matemaattiseen mallintamiseen. Malleja on kehitelty muun muassa liuotuksia, häivytyksiä, pyyhkäisyjä ja himmennyksiä varten. [Brunelli et al. 1999, 86–87; Del Bimbo 1999, 216–217, 220; Idris & Panchanathan 1997, 157].

Otosrajoja voidaan tunnistaa myös seuraamalla objektien reunojen ilmestymistä ja katoamista (kuvankäsittelymenetelmillä käsitellyissä) kuvasarjoissa; Brunelli ja muut [1999] kutsuvat tätä piirrepohjaiseksi tunnistamiseksi. Otosrajoissa tai häivytyksissä uusia reunoja (“intensiteettireunoja” Brunellin ja muiden [1999] mukaan) ilmestyy kaukana vanhoista ja vanhat reunat katoavat kaukana uusista. Objektien reunapikseleitä ('edge pixels'), jot-

ka ilmestyvät tai katoavat kaukana vanhoista reunapikseleistä, tarkastellaan tulevina ('entering') tai lähtevinä ('exiting') reunapikseleinä. Otokset tunnistetaan laskemalla tulevien ja lähtevien pikseleiden määrä kahdessa peräkkäisessä kehyksessä – eli niiden (mustien) pikselien määrä, jotka ovat läsnä kehyksessä, mutta katoavat seuraavassa. Jos erot kehyksien välillä ovat tarpeeksi suuret, oletetaan, että kyseessä on otosraja. Leikkaukset, häivytykset ja liuotukset havaitaan laskemalla tulevat ja lähtevät reunapikselit, kun taas pyyhkäisyt havaitaan tarkastelemalla reunapikseleiden paikkasidonnaista jakaumaa. Menetelmä toimii hyvin myös raskaasti pakatuissa videoissa. [Antani et al. 2002, 956; Brunelli et al. 1999, 87–88; Del Bimbo 1999, 209–211, 222–223; Ponceleon & Srinivasan 2002, 22; Rui et al. 1999, 360.] Teoreettisesta näkökulmasta pyyhkäisyjä ei voida erottaa kameran liikkeistä kuten jäljittämisestä ('tracking') ja noususta ('boom'), sillä kummassakin kehyksien pikselit käyvät läpi yksinkertaisen käännöksen ('translation'). Niinpä kyseiset paikkasidonnaiset tehosteet ('edits') havaitaan vertaamalla niitä niiden abstrakteihin malleihin. [Del Bimbo 1999, 220.]

4.1.1.3 Segmentointimenetelmien luotettavuus

Antani ja muut [2002, 958–959] esittelevät histogrammipohjaisten algoritmien suorituskykyä otosrajojen tunnistamisessa: paras algoritmi saavutti RGB-väriavaruudessa parhaimmillaan 68 %:n saannin 95 %:n tarkkuudella. Toisin sanottuna algoritmi tunnistaa 68 % otosrajoista, kun tarkkuus on vielä riittävät 95 %. MPEG-pohjaisista algoritmeista parhaan algoritmin saanti oli 79 % ja tarkkuus 88 %. Jos otokset tunnistetaan väärin, se vaikuttaa muun muassa tarinan yksiköiden muodostamiseen ja niiden esittämiseen. Brunelli ja muut [1999, 90] toteavat, että leikkausten tunnistamiseen käytettävien algoritmien luotettavuus on hyvä. Sen sijaan asteittaisten siirtymien, erityisesti liuotusten, tunnistaminen on osoittautunut huomattavasti vaikeammaksi. [Brunelli et al. 1999, 90.] Del Bimbon [1999, 223] mukaan pakatussa videodatassa toimivien menetelmien luotettavuus leikkausten havaitsemisessa on lähes yhtä hyvä kuin pakkaamattomassakin datassa toimivien menetelmien, mutta pakkaamattomassa datassa toimivien menetelmien luotettavuus heikkenee monimutkaisten siirtymätehosteiden kohdalla. Pakkaamattomassa datassa toimivat menetelmät ovat luotettavampia, mutta ne vaativat enemmän tallennustilaa ja enemmän laskentatehoa. [Del Bimbo 1999, 223.] Leikkausten tunnistamisessa histogrammipohjaiset menetelmät ovat epäherkkiä hitaille kameran liikkeille kuten panoroinnille ja zoomaukselle. Niillä on mahdollisuus saavuttaa lähes 95 %:n tarkkuusarvoja. Useimmat histogrammipohjaiset menetelmät ovat kuitenkin herkkiä nopeille kameran liikkeille ja suurille tai nopeasti liikkuville objekteille. Nopeat muutokset kirkkaudessa heikentävät

histogrammipohjaisten algoritmien suorituskykyä. Joka tapauksessa kynnysten asettaminen on tärkeää lähes kaikissa tekniikoissa. Histogrammipohjaiset menetelmät vaativat toisaalta vain vähän prosessoritehoa. [Del Bimbo 1999, 223.]

Reunojen muuttumiseen perustuvat menetelmät häiriintyvät pahasti kahden peräkkäisen kehyksen välisestä heikosta kontrastista. Ne ovat myös erittäin herkkiä kynnysten säädöille: huonosti sopivat säädöt tekevät algoritmista herkän hälylle. Lisäksi reunojen muuttumiseen perustuvissa menetelmissä häivytykset ja liuotukset menevät helposti sekaisin, varsinkin jos liuotuksiin liittyy liikettä. Mallintamiseen perustuva menetelmä on luotettava väline asteittaisten siirtymien tunnistamiseen, vaikkakin häivytykset ja liuotukset saattavat joissakin tilanteissa mennä sekaisin. [Del Bimbo 1999, 224.] Del Bimbon [1999, 224] mukaan kaikilla asteittaisten siirtymien tunnistamiseen tarkoitetuilla menetelmillä on vaikeuksia havaita liuotuksia, jos kaksi hyvin samankaltaista otosta liuotetaan samoilla tai samankaltaisilla häivytysasteilla ('fade rate'). Vastaavasti nopeat häivytykset ('fade') ja liuotukset aiheuttavat myös ongelmia, koska ne eivät eroa tarpeeksi leikkauksista. [Del Bimbo 1999, 224.]

4.1.1.4 Avainkehysten valitseminen

Kunhan otosten ja kohtausten vaihdokset on tunnistettu ja video jäsennetty tarkoituksenmukaisiin segmentteihin, niistä valitaan avainkehyksiä. Avainkehys on videosekvenssistä valittu yksittäinen kehys, joka edustaa kokonaista sekvenssiä. Avainkehysten avulla esitetään videosisältöä tiivistetyssä muodossa (mistä lisää luvussa 5), mutta niitä tarvitaan myös objektien ja paikkasidonnaisten ominaisuuksien tunnistamisessa, sillä paikkasidonnaiset piirteet poimitaan avainkehyksistä. [Antani et al. 2002, 957; Del Bimbo 1999, 230, 255; Idris & Panchanathan 1997, 159; Rui et al. 1999, 360.] Xiongin ja muiden [1997, 58] mukaan on järkevää esittää mahdollisimman paljon videon sisällöstä käyttämällä mahdollisimman vähän avainkehyksiä. Näin ollen on tarkoituksenmukaista valita avainkehyksiä vain silmiinpistävästi muuttuneista otoksista [Antani et al. 2002, 957].

Avainkehyksien poimimiseksi segmenteistä on ehdotettu useita erilaisia algoritmeja. Yksinkertaisimmillaan avainkehykseksi voidaan valita segmentin ensimmäinen kehys, ja tätä kehystä voidaan käyttää viitteenä varsinaista avainkehystä valittaessa. Muut kirjallisuudessa esitellyt avainkehyksiä poimivat algoritmit perustuvat esimerkiksi optisen virtauskentän analyysiin, videovirran merkittäviin taukoihin tai liikekuvioiden käyttämiseen. Liikepohjainen avainkehysten tunnistaminen perustuu siihen, että ohjaajat toistuvasti käyttävät kameran liikettä – ja näyttelijät eleiden korostamista – rakentaakseen monimutkaisia viestejä yksittäisiin

otoksiin. Avainkehys voi olla esimerkiksi otoksen kymmenes kehys, otoksen ensimmäinen ja viimeinen tai, puhuttaessa kehittyneemmistä menetelmistä, avainkehykset voidaan valita sisällön monimutkaisuuden, otoksen aktiivisuuden tai liikkeen perusteella. [Brunelli et al. 1999, 98; Idris & Panchanathan 1997, 159; Rui et al. 1999, 360–361; Xiong et al. 1997, 58–60.] Jos paljon toimintaa sisältävästä otoksesta tai kohtauksesta otetaan vain pari avainkehystä, ne eivät välttämättä riitä esittämään videosisältöä kunnolla, vaan lopputuloksena on semanttinen katkos ('semantic discontinuity'). Toisaalta jos otoksissa on vähän toimintaa, avainkehyksissä voi olla liikaa toisteisuutta. Kummatkin näkökohdat on otettava huomioon. [Xiong et al. 1997, 60.] Xiongin ja muiden mukaan ryppäytystekniikoita ('clustering techniques') voidaan käyttää niin, että etsitään kehysryppäitä otoksesta ja valitaan avainkehyksiksi ne kehykset, jotka ovat lähinnä ryppäiden keskustaa. Toinen tapa on verrata otoksen ensimmäistä kehystä sekventiaalisesti seuraaviin kehyksiin ja valita ensimmäinen oleellisesti erilainen kehys avainkehykseksi. [Xiong et al. 1997, 60.]

4.1.2 Uutisjuttujen tunnistaminen

Segmentointi on suhteellisen ongelmatonta niin kauan, jos tyydytään vain erottelemaan otoksia toisistaan. Otokset ovat kuitenkin usein liian lyhyitä [6], jotta ne voisivat välittää semanttisia merkityksiä: jos uutisvideossa ääni on keskeisin informaatiota välittävä komponentti, kuinka paljon ehditään sanoa esimerkiksi viisi sekuntia kestävässä otoksessa? Pelkästään otoksia segmentoimalla ei pystytä vastaamaan useimpiin tiedonhakuongelmiin: samalla tavalla kuin tekstit muodostuvat tietyllä tavalla järjestetyistä lauseista, informaatiota välittävät tarinat muodostuvat tietyllä tavalla järjestetyistä otoksista. Bolle ja muut [1998] kutsuvat tarinan yksiköksi ryhmää ketjutettuja otoksia, jotka kuvaavat ajallisesti jatkuvaa ('continuous') tapahtumaa. Del Bimbo [1999, 225] kutsuu otoksia korkeamman tason ajallisia yksiköitä makrosegmenteiksi. Vaikka tapahtumat ovatkin ajallisesti jatkuvia ja ne halutaan sellaisena esittää, video on välineenä sekventiaalisuutensa takia epäjatkuva esitystapa. Epäjatkuvuus ei esiinny pelkästään kehysten ja otosten välillä, vaan myös tarinan yksikköjen välillä on ajallista katkonaisuutta. Bollen ja muiden [1998] mukaan kunnolla leikatun videon tuottama jatkuvuus merkityksessä peittää alleen varsinaisen esitystavan epäjatkuvuudet. Merkitykset välitetään videoissa montaasiksi kutsutun menetelmän avulla, jossa kuvallisia yksittäisiä otoksia, jotka ovat neutraaleja merkityksessä erillään, yhdistellään tuottamaan haluttuja merkityksiä. Videoiden indeksoinnin

6 Liitteessä 8 analysoidun uutisjutun otosten pituudet vaihtelevat parista sekunnista aina ensimmäisen uutisjutun juonnon kahdeksaantoista sekuntiin. Useimpien otosten pituus vaihtelee viiden ja kymmenen sekunnin välillä.

haasteena on löytää merkityksen epäjatkuvuus eli semanttista sisältöä välittävien tarinoiden yksiköiden väliset katkokset – esimerkiksi siirryttäessä uutisjutusta toiseen – tai vastaavasti todentaa merkityksen jatkuvuus otosryhmien välillä. [Bolle et al. 1998.] Näiden katkosten perusteella muodostetaan kohtauksia ja muita semanttisesti merkittäviä tarinan yksikköjä.

Videon ajallisen hierarkian tunnistamiseksi vaaditaan pelkkää otosten tunnistamista kehittyneempiä menetelmiä. Ongelmana on se, että mitä korkeammalle tasolle rakenteellisessa hierarkiassa siirrytään, sen vaikeammaksi tunnistaminen käy. Otosten välisten rajojen tunnistaminen on suhteellisen helppoa, mutta varsinaiset semanttiset merkitykset välittävät kohtaukset ovat jo huomattavasti vaikeammin poimittavissa. Varhaiset videomallinnuksen lähestymistavat eivät tukeneet segmenttien järjestämistä semanttisesti merkittäviksi yksiköiksi [Petković & Jonker 2000]. Video saatettiin järjestää vain peräkkäisiin segmentteihin kuluneen ajan perusteella sen sijaan, että samaa asiaa käsittelevät segmentit olisi järjestetty ryhmiksi niiden samankaltaisuuden ja ajallisten suhteiden perusteella [Bolle et al. 1998]. Rui ja muut [1999, 360] tosin lisäävät, että ongelmat tutkimuksen rajoittumisessa otosten ja avainkehyksien tasolle pätevät myös visuaalisesti samankaltaisista otoksista muodostettuihin ryhmiin, sillä nekään eivät vielä varsinaisesti välitä videon semanttista sisältöä.

Petkovićin ja Jonkerin [2000] mukaan videon segmentointikriteerit voivat olla syntaktisia tai semanttisia. Syntaktiset kriteerit, joita käsiteltiin edellä, tuottavat videosta kiinteitä segmenttejä, jotka on tallennettu yhtäjaksoisen kameran toiminnan aikana. Syntaktisten kriteerien ongelmia ovat kuitenkin niiden joustamattomuus ja rajoitukset: kiinteästi segmentoidut videot mahdollistavat vain yhden esityksen alkuperäisestä datasta. Lisäksi on otettava huomioon, että segmentien käsitteleminen toisistaan irrallaan jättää kontekstuaalisen ulottuvuus täysin huomiotta, mikä taas vaikeuttaa entisestään oikeiden semanttisten merkitysten tuottamista. [Petković & Jonkerin 2000.] Ongelmana syntaktisissa menetelmissä on siis se, että segmentoinnin jälkeen lopputuloksena on iso joukko lyhyitä, toisistaan irrallaan olevia videosekvenssejä, joiden väliset suhteet on hukattu. Bolle ja muut [1998] kutsuvat tämänkaltaista syntaktista segmentointia yksinkertaisesti otosrajojen tunnistamiseksi ('shot-boundary detection'). Kohtauksien kaltaisten semanttisten (vrt. fyysiset sekvenssit) yksiköiden tunnistamista voidaan kutsua loogiseksi segmentoinniksi. Bolle ja muut [1998] toteavat, että otosten välisten semanttisten tulkintojen tuottaminen on ainakin yhtä tärkeää kuin otosten sisäinen analyysi. Otosten välisten suhteiden (tai niiden puuttumisen) tunnistaminen mahdollistaa videon rakenteen uudelleenlöytämisen ja semanttisen merkityksen assosioimisen otossarjoihin [Rui et al. 1999, 360]. Segmentointiin liittyy vielä yleisten ajallisten tapahtumien tunnistaminen, esimerkiksi ”puhuvien päiden” välisen keskustelun erottaminen muusta videosisällöstä, ja otoksia pidemmillä aikaväleillä esiintyvien objektien tunnistaminen, joiden segmentointia

voidaan pitää semanttisena. [Bolle et al. 1998; Brunelli et al. 1999, 100.]

4.1.2.1 Ajallinen hierarkia

Otosten välisen prosessoinnin tarkoituksena on korkean tason videorakenteen johtaminen [Bolle et al. 1998]. Petkovićin ja Jonkerin [2000] sekä Ruin ja muiden [1999, 360] mukaan tutkimuksessa ollaan päädytty videon rakenteen esittämiseen hierarkiana, jossa monimutkaisia videoyksikköjä tuotetaan yhdistelemällä segmentoituja alkeisyksikköjä (eli käytännössä otoksia) korkeamman tason yksiköiksi. Useimmiten käytetään hierarkiaa, jossa ylimmällä tasolla on tarina (esimerkiksi uutisjuttu), sitten kohtaukset, otokset ja lopulta kehykset. Siirryttäessä videoesityksen hierarkiassa korkeammalle tasolle toisiinsa liittyviä ('relate') otoksia yhdistetään ryhmiksi, joista tuotetaan puu-muotoinen esitys selailua varten [Petković & Jonker 2000; Rui et al. 1999, 360].

Videon ajallinen hierarkia voidaan esittää yksinkertaisesti laajentamalla kerroksiin perustuvaa lähestymistapaa periytymisellä ('nesting'), jolloin videon elementit esitetään hierarkkisesti organisoituina solmuina puussa. Käyttäjä voi tällöin tutkia solmujen suhteita ja niitä konteksteja, joissa kukin solmu esiintyy. Semanttisesti suuntautuneet lähestymistavat voivat hyödyntää myös hierarkkista esitystapaa: videoista poimitaan objekteja ja tapahtumia, joiden esiintymisen perusteella video segmentoidaan. Jokainen objekti, ryhmä objekteja tai tapahtumia yhdistetään joukkoon kehyssekvenssejä. Jos käyttäjä haluaa nähdä johonkin hierarkian sisäiseen solmuun yhdistetyn videosekvenssin, hakujärjestelmän pitäisi luoda sommitelma ('composition') videosekvenssejä, jotka ovat yhteydessä tämän solmun perillisiin ('children'). [Petkovi ć & Jonker 2000.]

Jos kaksi otosta ovat visuaalisilta ominaisuuksiltaan samankaltaisia, on todennäköistä, että ne kuuluvat samaan ryhmään. Ruin ja muiden [1999, 360–361] mukaan otokset voidaan ryhmitellä esimerkiksi (1) pelkästään ajan perusteella ilman, että visuaaliseen sisältöön kiinnitetään mitään huomiota, tai (2) pelkästään visuaalisen sisällön perusteella, jolloin aikaa ei oteta huomioon. Valitettavasti kummassakaan mainitussa otosryhmiin perustuvassa lähestymistavassa videon rakenteistaminen ei tapahdu edelleenkään tarpeeksi korkealla semanttisella tasolla. [Rui et al. 1999, 360–361.] Seuraavaksi käsitellään tarkemmin otosryhmien tuottamista.

4.1.2.2 Otosten ryhmittely: mallit ja säännöt

Kohtausten tasolla tapahtuvaan videon jäsentämiseen on kaksi lähestymistapaa: mallit ja yleiset säännöt. Mallipohjaisessa lähestymistavassa ensin rakennetaan apriorinen malli nimenomaisesta sovelluksesta tai alueesta, joka halutaan mallintaa. Malli määrittää kohtausrajojen tuntomerkit ('characteristics'), joiden avulla rakenteeton videovirta voidaan abstraktoida rakenteiseksi esitykseksi. [Rui et al. 1999, 361.] Brunellin ja muiden [1999, 85] mukaan asteittaisia siirtymäefektejä on käytetty jäsentämisen ('punctuation') välineinä: esimerkiksi 1950-luvulle asti liuotuksia käytettiin kuvaamaan muutosta paikassa tai aikatakaumaa ('flashback') [Brunelli et al. 1999, 85]. Tiettyjä ohjelmatyyppejä varten suunniteltujen mallien avulla voidaan saavuttaa korkea tunnistustarkkuus, mutta ennen jäsentämistä on rakennettava malli, mikä vaatii paljon aikaa ja tietämystä kyseessä olevasta alasta. Rui ja muut [1999, 361] mainitsevat esimerkkejä mallipohjaisista menetelmistä, joita on esitetty TV-uutisia ja jalkapallolähetyksiä varten. Toinen vaihtoehto videoiden jäsentämiseen kohtausten tasolla on käyttää multimodaalista sääntöpohjaista lähestymistapaa, jossa ensin tunnistetaan ajallisia paikallisia sääntöjä, jotka saadaan välineen sisällöstä. Tämän jälkeen kohtauksia tuotetaan yhdistelemällä sääntöjä. [Rui et al. 1999, 361.] Suosituimmat kirjallisuudessa käsitellyt lähestymistavat videoiden hierarkkisten rakenteiden laskemiseen ja esittämiseen käyttävät aikarajoitettua klusterointia ('time constrained clustering') ja kohtaustensiirtymiskaaviota ('scene transition graph'). Ideana näissä lähestymistavoissa on se, että otoksia ryppäytetään niiden avainkehyksien ja ajallisten suhteiden vastaavuuden perusteella. Käsitellyt menetelmät eivät rajoitu tietyn tyyppisiin videoihin ja toimivat pakatulla ja pakkaamattomalla videodatalla. [Antani et al. 2002, 957; Bolle et al. 1998; Brunelli et al. 1999, 102; Del Bimbo 1999, 226–227; Kender & Yeo 1998, 7; Rui et al. 1999, 361; Yeo & Yeung 1997, 48.]

Aikarajoitetussa klusteroinnissa otokset ryppäytetään niiden visuaalisen sisällön ja ajallisen sijainnin ('temporal localities') perusteella. Otosten visuaalista samankaltaisuutta voidaan tarkastella muun muassa avainkehysten histogrammien leikkauspisteiden ('intersection'), pikseleiden korrelaation tai näiden yhdistelmän avulla. Avainkehyksistä voidaan valita myös vaihtoehtoisia piirteitä mitattavaksi: erilaisten piirteiden kuten esimerkiksi otosten keston, paikkasidonnaisen värien jakautumisen, hallitsevien liikeominaisuuksien ja tekstuurien ominaisuuksien sekä äänen piirteiden käyttäminen mahdollistavat erilaisten ryhmien muodostamisen. Aikarajoitettu klusterointi estää kahden toisistaan ajallisesti kaukana olevan samankaltaisen otoksen ryppäyttämisen: kaksi otosta, jotka ovat kaukana toisistaan ajallisesti, vaikka ne näyttäisivätkin samalta, edustavat eri kontekseja ja kuuluvat eri kohtauksiin. Ryppäyttäminen aloitetaan tunnistamalla otosrajat kuten normaalistikin kehysten välisiä eroja tarkkailemalla. Aluksi jokainen ryväs muodostuu yhdestä otoksesta. Jokaisella uudella askeleella algoritmi yhdistää kaksi samankaltaisinta ryvästä aikaikkunan ('time window')

sisällä uudeksi ryppääksi. Otosten etäisyys otetaan huomioon erilaisuuskynnyksen ('dissimilarity threshold') avulla, jolla mitataan suurinta sallittua erilaisuutta ryppään kahden otoksen välillä. Yksinkertaisesti esitettynä aikarajoitetussa klusteroinnissa visuaalisesti samankaltaiset otokset ryhmitellään samaan ryppääseen ja ne merkitään samoilla koodeilla olettaen, että ne eivät ole ajallisesti liian kaukana toisistaan. [Antani et al. 2002, 957; Bolle et al. 1998; Brunelli et al. 1999, 102; Del Bimbo 1999, 226–227; Kender & Yeo 1998, 7; Rui et al. 1999, 361; Yeo & Yeung 1997, 48.]

Otossekvenssi mallinnetaan muuttamalla se kaksiulotteiseen muotoon kohtauksensiirtymiskaavion avulla [ks. liite 1]. Otoksen käsitteen taustalla on tässä erityisen vahvasti se, että otos on kuvattu yhdellä kameralla ja yhdestä kameran sijainnista. Kaaviossa on solmuja ('nodes'), jotka koostuvat ryhmistä samankaltaisia otoksia eli kameran sijainteja, ja suunnattuja yhteyksiä ('directed edges') eli linkkejä ('arcs'), jotka esittävät tarinan ajallista kulkua eli ajassa eteenpäin suuntautuvia siirtymiä näiden otosten välillä. Yhden ryhmän (eli ryppään) sisäisten otosten oletetaan olevan keskenään samankaltaisia, ja jokainen ryhmä vastaa solmua kaaviossa. Tarinan ajallista virtaa eli siirtymiä otosten välillä kuvataan suunnatulla linkillä eli yhteydellä: ryhmästä A ryhmään B liitetään aina linkki, kun otos ryhmässä B on välittömästi seuraava ryhmän A otokseen nähden. Koska samankaltaisia otoksia (eli kameran sijainteja) toistetaan eri järjestyksessä, kaavioista tulee tiiviitä, jos otosryhmien välillä on paljon yhteyksiä ('edge') eli vuorovaikutusta. Kohtauksen raja tulee vastaan, kun kaavio on ohut eli otoksesta on yhteys vain yhteen sellaiseen otokseen, jota ei ole aiemmin käytetty; tällaista yksisuuntaista linkkiä ('directed edge') kutsutaan leikkausyhteydeksi ('cut edge'). Leikkausyhteys osittaa kohtauksensiirtymiskaavion alikaavioiksi eli tarinan yksiköiksi. Mikäli leikkausyhteyttä ei löydy, kaavio on syklinen ja kohtauksen vaihtumista ei tunnisteta. Tätä varten tarvitaan aikarajattua klusterointia, jossa aikaikkunan ulkopuolisten otosten välinen etäisyys asetetaan mahdollisimman suureksi, vaikka otokset olisivatkin visuaalisesti samankaltaisia: samankaltaistenkaan otosten välille ei merkitä enää ryhmän sisäistä vuorovaikutusta, koska ne on pakotettu eri ryhmiin. [Kender & Yeo 1998; Bolle et al. 1998; Brunelli et al. 1999, 102.] Kohtauksensiirtymiskaavio voi esittää tarinan etenemisen, jossa jokaista tarinan yksikköä edustaa kytketty alikaavio, joka on kytketty seuraavaan tarinan yksikköön leikkausyhteydellä; lisäksi jokainen tarinan yksikkö on itsessään kytketty alikaavio. Jokaisen alikaavion sisällä kuvasisältö ja ajallinen rakenne esitetään suppeassa muodossa solmujen ja yhteyksien avulla. Segmentointi tarinan yksikköihin toteutetaan tarkastelemalla koko kohtauksensiirtymiskaavion suljettuja alikaavioita. Otosten väliset vuorovaikutukset heijastuvat kohtauksensiirtymiskaaviossa läpi jaksojen ('cycles') ja kaavion solmuista saapuvien tai menevien leikkausyhteyksien ('cut edge') määrässä. Tarinan yksiköt poimitaan etsimällä kohtauksensiirtymiskaaviosta leikkausyhteyk-

siä, joiden poistaminen osittaa ('disconnect') kaavion. [Antani et al. 2002, 957; Bolle et al. 1998; Brunelli et al. 1999, 102; Del Bimbo 1999, 226–227; Kender & Yeo 1998, 6–8; Rui et al. 1999, 361; Yeo & Yeung 1997, 48.]

Bollen ja muiden [1998] mukaan useimmissa tarinan yksikössä esiintyy useita objekteja samanaikaisesti, ja kutakin objektia koskevat useat eri otokset on ketjutettu ('concatenate') toisiinsa. Eri tarinan yksiköihin kuuluvat solmut, jotka edustavat otosryhmiä kaaviossa, eivät ole ketjutettuja tai yhteydessä toisiinsa paitsi siirryttäessä tarinan yksiköistä toiseen leikkausyhteyden välityksellä. Tarinan yksikön otosten intensiivisestä vuorovaikutuksen takia otokset voidaan merkitä ('label') niiden sisällöllisen samankaltaisuuden perusteella: samaan ryhmään eli solmuun kuuluvilla otoksilla on siten sama symboli. Kun kaksi eri otosta on merkitty samalla tavalla, on hyvin todennäköistä, että ne esittävät samoja asioita. Merkeistä muodostettuja sekvenssejä voidaan käyttää paitsi tarinan yksiköiden segmentointiin myös yleisten ajallisten tapahtumien tunnistamiseen videoista. Kuhunkin tarinan yksikköön lasketaan kuuluvaksi joukko sisällöltään samankaltaisia otoksia. [Bolle et al. 1998.]

Oletetaan, että yksinkertaisen videosekvenssin otokset on merkitty joukkona {A-B-C-D-C---A-B}, missä “-” tarkoittaa leikkausta ja “---” asteittaista siirtymää. Tässä tapauksessa tarinan yksikkö muodostuisi siis yhteensä seitsemästä otoksesta ja neljästä erilaisesta kameran sijainnista. Aliryhmä {CDC}muodostaisi oman tarinan yksikkönsä elleivät otokset A ja B sulkisi yksikköä. Koska tarinan yksikkö suljetaan, siinä ei ole leikkausyhteyttä. Sen sijaan joukossa {A-B-C-A---D-E-D-E-F-E-G} on kolme aliryhmää {A,B,C,A}, {D,E,D,E,F,E} ja yksittäinen otos G. Tässä tarinan yksikössä on kaksi leikkausyhteyttä: A:n ja D:n välillä sekä E:n ja G:n välillä. Vastaavasti Bollen ja muiden [1998] mukaan yleisiä ajallisia tapahtumia voidaan tunnistaa käyttämällä merkkisekvenssejä ja ottamalla huomioon sekvenssien merkkien toisteisuuden tai sen puuttumisen. Esimerkiksi otoksien “D” ja “E” vaihtelu saattaa hyvinkin olla “puhuvien päiden” vuoropuhelua.

Brunelli ja muut [1999, 103] sekä Del Bimbo [1999, 225–226] käsittelevät myös mediapohjaisiin sääntöihin ('media-based rules') pohjautuvaa lähestymistapaa, joka perustuu siihen, että ajallisessa mediassa katsojalle täytyy antaa vihjeitä, jotta tämä voisi tunnistaa makroskooppisia muutoksia tarinassa. Videoissa tämänkaltaisia vihjeitä on monia: on erityisiä siirtymätehosteita otosten välillä – joilla voidaan viestittää esimerkiksi muutosta paikassa – ja muutoksia leikkausrytmissä ja niin edelleen. Malli perustuu joukkoon sääntöjä, jotka on muodostettu analysoimalla videoita, tutustumalla elokuvateoriaan ja keskustelemalla esimerkiksi tuottajien, kriitikoiden ja muiden asiantuntijoiden kanssa. Sääntöjä tuotettaessa otettava huomioon, kuinka (1) asteittaiset siirtymät sijoitetaan leikkausten väliin, (2) välimatka, jolla kaksi samankaltaista otosta toistetaan kuvavirrassa, (3) vierekkäisten otosten samankaltaisuus, (4)

leikkausrytmi, (5) musiikin läsnäolo hiljaisuuden jälkeen ja (6) kameran liikkeen samankaltaisuus. [Brunelli et al. 1999, 103; Del Bimbo 1999, 225–226.] Näin esimerkiksi kaksi samankaltaista otosta tunnistetaan samaan makrosegmenttiin (eli ryhmään), mikäli ne löydetään kahden tai kolmen otoksen välimatkan sisällä: tällä tavalla tunnistetaan haastatteluiden "puhuvat päät". Kehyksien välistä samankaltaisuutta tarkastellaan vertailemalla pikseleiden eroja ('pointwise differences') matalan resoluution valotiheyskuvien ('luminance images') välillä. [Del Bimbo 1999, 226.] Brunelli ja muut [1999, 103–104] käsittelevät mallia tarkemmin. Rui ja muut [1999, 361] mainitsevat, että sääntöpohjaiset menetelmät eivät ole vielä täysin kypsiä: uusien sääntöjen tuottaminen ja testaaminen saattavat olla aivan yhtä työläitä kuin sovelluskohtaisten mallien tuottaminen.

Rui ja muut [1999, 361–367] tuovat esille menetelmän, jossa käytetään älykästä valvomatonta ('unsupervised') klusterointitekniikkaa ja aikamukautuvaa ryhmittelyä kohtaustason sisällysluettelon rakentamiseksi (sisällysluetteloista lisää luvussa 5). [Rui et al. 1999, 361.] Aikamukautuvan ryhmittelyn avulla luodaan ryhmiä, jotka toimivat välittäjänä otosten ja kohtausten välillä. Tarkoitus on järjestää samankaltaiset otokset ryhmiin, sillä mitä enemmän otokset muistuttavat toisiaan, sitä suuremmalla todennäköisyydellä ne kuuluvat samaan kohtaukseen. Otosten samankaltaisuuden määrittelyssä otetaan huomioon, kuten edellä käsitellyissä menetelmissä, että (1) otosten pitäisi olla visuaalisesti samankaltaisia eli niillä pitäisi olla samantapaiset paikkasidonnaiset ja ajalliset piirteet; (2) samankaltaisten otosten pitäisi olla ajalliselta sijainniltaan ('time locality') lähellä toisiaan. Ajallispaikallisten piirteiden poimimisessa otetaan huomioon otosten aktiivisuus. Kehysten tasolla poimitaan piirteitä (tässä tapauksessa värihistogrammeja) paikkasidonnaisen informaation kuvailemiseksi. Videon sekventiaalisuuden takia visuaalisesti samankaltaisten otosten lisäksi pitää ryhmitellä myös otoksia, jotka ovat semanttisesti yhteydessä toisiinsa, vaikka ne eivät muistuttaisi toisiaan: esimerkiksi puhuvat päät erilaista taustaa vasten. Aikarajatun klusteroinnin sijaan Rui ja muut [1999] käyttävät yleisempää aikamukautuvaa ryhmittelyä, joka perustuu edellä mainituille samankaltaisille otoksille asetettuihin ehtoihin. Tässä lähestymistavassa kahden otoksen välinen samankaltaisuus on visuaalisen samankaltaisuuden lisääntyvä funktio ja kehysten erojen vähentyvä funktio. [Rui et al. 1999, 362.] Mitä enemmän otokset muistuttavat toisiaan ja mitä lähempänä ne ovat toisiaan, sitä suuremmalla todennäköisyydellä ne kuuluvat samaan ryhmään.

Ruin ja muiden [1999, 363] mukaan kohtausten rakenteen konstruoimiseksi tarkoituksenmukaisella semanttisella tasolla tarvitaan aikamukautuvan ryhmittelyn lisäksi älykästä valvomatonta klusterointitekniikkaa, joka toimii kahdella askeleella: (1) kerätään samankaltaiset otokset ryhmiin käyttämällä aikamukautuvaa ryhmittelyä ja (2) yhdistetään se-

manttisesti toisiinsa yhteydessä olevat ryhmät kohtauksiksi. Menetelmä ratkaisee aikarajoitetun klusteroinnin epäjatkuvuuden ongelman, joka johtuu aikaikkunan käyttämisestä otosten samankaltaisuuden laskemiseen. Aikaikkunan tarkoitus on varmistaa, että liian kaukana toisistaan olevat samankaltaiset otokset eivät sijoitu samaan ryhmään. ”Ikkunaefektiksi” kutsuttu ongelma tulee esiin, kun samankaltaiset otokset ovat hieman kauempana toisistaan kuin mitä ikkunan pituudeksi on säädetty – esimerkiksi tilanteessa, jossa aikaikkuna on säädetty kahdeksaan otokseen ja samankaltaiset otokset ovat yhdeksän otoksen päässä toisistaan. Tämänkaltainen epäjatkuvuus saattaa aiheuttaa väärän ryppäytyksen ja tekee ryppäytysmenetelmästä herkän ikkunoiden koolle. Ongelman ratkaisemiseksi Rui ja muut [1999, 363] ehdottavat käsitettä nimeltään ajallinen houkutus ('attraction'), joka on jatkuva ja vähenevä kehysten erojen funktio. Monissa tapauksissa otos ei ole tarpeeksi samankaltainen kuin muut sopiakseen mihinkään jaksoon ('scene'). Otos voi kuitenkin olla tarpeeksi samankaltainen tietyssä määrin useimpien jakson ryhmien kanssa. Osa lähestymistavoista vertaa vain yksittäistä otosta yksittäisiin ryhmiin eikä koko jakson kaikkiin ryhmiin. [Rui et al. 1999, 363.] Rui ja muut [1999, 363–366] käsittelevät järjestelmän toimintaa yksityiskohtaisemmin. Menetelmässä kohtausten tunnistaminen toimii kohtuullisen hyvin useimmilla videotyypeillä, mutta se on kuitenkin parempi hidastempoisissa videoissa kuin nopeatempoisissa, koska viimeksi mainituissa visuaalinen sisältö on yleensä monimutkaisempaa ja vaikeampaa tunnistaa ('capture'). Kuten monet muutkin menetelmät, lähestymistapa on taipuvainen ylisegmentointiin eli merkitsemään kohtauksen rajan, vaikka sellaista ei olisikaan. [Rui et al. 1999, 366.]

Brunelli ja muut [1999, 104], Bolle ja muut [1998] sekä Del Bimbo [1999, 227–228] käsittelevät TV-uutisten jäsentämiseen tarkoitettua mallia, joka perustuu aprioriseen tietoon mallinnettavasta alueesta. Esimerkiksi uutisten säädelty rakenne mahdollistaa yksinkertaisen ja tarkan mallintamisen. Mallit tuotetaan tilaa koskevien siirtymien avulla, jossa jokainen tila vastaa uutislähetyksen vaihetta, kuten uutisankkurin puhetta uutisstudiossa. Indeksointi aloitetaan otosten segmentoinnilla, minkä jälkeen otoksista poimitaan avainkehyksiä, jotka luokitellaan uutisankkurin sisältäviin otoksiin (eli uutisstudioon) ja varsinaisiin uutisotoksiin (eli uutisjuttuihin ja -sähkeisiin). Ankkuriotokset muodostuvat kolmesta alamallista: (1) aluemallit käsittävät uutisankkurin, uutisikonin, uutisohjelman otsikkopalkin, uutisankkurin nimipalkin ja taustan; (2) kehysmallit muodostuvat paikkasidonnaisista asetelmista edellä mainittuja aluemalleja; (3) otosmallit muodostuvat luetteloista edellä mainittuja kehysmalleja, joiden avulla kukin otostyyppi on mallinnettu. Ankkuriotokset tunnistetaan ennalta määriteltyjen mallikuvien ja ajallisten piirteiden avulla: ensin paikallistetaan potentiaalisia ankkuriotoksia, ja ajallisesti täsmäävistä otoksista valitaan kehyksiä kehysmallien kanssa täsmäytettäväksi. Del Bimbo [mt.] mainitsee vertailuun käytettävään esimerkiksi kaavio- ja histogrammi-

täsmäytyksen yhdistelmää. Mahdollisesti uutisankkurin sisältävien avainkehysten hahmoja ('pattern') verrataan kunkin ankkuriotoksen kehysmallin hahmoihin. Kehysmallin avulla kehykset ositetaan alikehyksiin eli alueisiin, joiden sijainnin kehysmalli osoittaa, ja kutakin alikehystä (eli aluetta) verrataan sitä vastaavaan aluemalliin. Lopulta jokaisen uutisjutun indeksi sisältää otosten määrän, aloitusajan, keston ja joukon avainkehyksiä, jotka edustavat otosten visuaalista sisältöä. Menetelmässä käytetään siis videon visuaalisia ja ajallisia elementtejä yhdessä objektien paikkasidonnaista sommittelua koskevan aihetiedon kanssa. Koska malli perustuu aprioriseen tietoon, sitä voidaan soveltaa vain tietyissä konteksteissa. [Bolle et al. 1998; Brunelli et al. 1999, 104; Del Bimbo 1999, 227–228.]

Brunelli ja muut [1999, 105], tiivistäessään digitaalisen videon automaattisten indeksointimenetelmien nykytilaa, toteavat, että videon automaattinen tiivistäminen ei ole vielä tuottanut korkealaatuisia tuloksia, koska semanttisten käsitteiden johtaminen on vaikeaa. Videon (ajallisen) rakenteen tunnistaminen on yhä alkutekijöissään ja hyviä tuloksia on saavutettu vain rakenteisissa videoissa. Nähtävästi kuvankäsittelymenetelmät eivät yksistään riitä vaan vaaditaan integroidumpia ratkaisuja, joissa otetaan videokuvan lisäksi huomioon viestinnän muut kanavat. [Brunelli et al. 1999, 105.] Lisää menetelmiä käsittelevät muun muassa Antani ja muut [2002, 957–958].

4.1.2.3 Segmentointi multimodaalisesti

Ääniraitaa voidaan käyttää apuna makrosegmenttien ja otosryhmien tunnistamisessa, etsittäessä kuvavirrasta mielenkiintoisia kohtia ja tarkistettaessa, että videodata sopii ennalta määriteltyyn malliin kohtauksista. Äänen analyysi perustuu esimerkiksi urheilutapahtumissa uutisreportterin tai juontajan avainsanoihin ja yleisön hurraukseen. [Del Bimbo 1999, 229.] Esimerkiksi Chang, Zeng, Kamel ja Alonso [1996] ovat kehittäneet urheiluvideoita varten automaattisen indeksointijärjestelmän, jossa ääniraitaa käytetään tiettyjen avainsanojen tunnistamiseen videosta – urheilulähetyksissähän puhe liittyy läheisesti kuvassa näkyviin tapahtumiin, ja sillä on keskeinen tehtävä informaation välittäjänä. Niinpä Chang ja muut [1996] päätyivät paikantamaan tärkeät kohdat videosta puheanalyysin perusteella ja vasta sitten käyttämään kuvananalyysimenetelmiä. [Chang et al. 1996; Del Bimbo 1999, 225, 229.] Ääniraitaa voidaan käyttää apuna paitsi kuvassa näkyvän henkilön tunnistamisessa myös loogisen segmentin (vrt. fyysisen segmentin eli otosrajan) vaihtumisen havaitsemisessa. Tästä lisää myöhemmin.

Millsin, Pyen, Hollinghurstin ja Woodin [2000, 4] lähestymistapa ottaa huo-

mioon ääniraidan akustisten rajojen ja videokuvan otosten rajojen samanaikaisuuden. Heidän algoritminsa jakaa äänivirran niistä kohdin, joissa se muuttuu merkittävästi, kuten puhujan muuttuessa; kuvavirralle tehdään samalla tavalla. Ääni- ja kuvavirtojen rajat yhdistetään käyttämällä iteratiivista algoritmia, joka valitsee akustisten rajojen muodostamiseksi sopivan kokoisia segmenttejä; algoritmi suosii niitä rajoja, jotka ovat lähellä kuvavirran rajoja. Millsin ja muiden [mt.] mukaan heidän lähestymistapansa tuottaa uutisvideoista käyttökelpoisempia segmenttejä kuin pelkkä ääni- tai kuvapohjainen segmentointi. [Mills et al. 1999, 4.]

Myös Lienhart ja muut [1997, 57] käyttävät ääniraitaa apuna kohtausten tunnistamisessa. Heidän menetelmässään kuvasekvenssistä tunnistettu muutos ei tarkoita kohtauksen (tai muun otosta korkeamman ajallisen yksikön) vaihtumista, mikäli ääniraidassa ei tapahdu merkittävää muutosta. Leikkauksia ääniraidassa kutsutaan aikaesiintymiksi ('time instances'). Aikaesiintymät rajoittavat ('delimit') ääneltään samankaltaisia aikajaksoja ('time periods'), ja niitä käytetään tutkimaan ääniraidan samankaltaisuutta eri otoksissa. Ääniraidan leikkaukset määritetään laskemalla ääniraidan jokaisen aikaikkunan taajuus ja intensiteettispektri ja ennustamalla seuraavaa aikaikkunaa varten arvoja. Leikkaus ääniraidassa tunnistetaan, jos ääniraidan taajuus ja intensiteettispektri poikkeavat huomattavasti ennustuksesta. [Lienhart et al. 1997, 57.]

4.2 Objektityyppien jäsennys ja objektien tunnistaminen

> Järjestelmän on eroteltava objektit ja tausta toisistaan, jäsennettävä erityyppiset objektit, tunnistettava kasvoja ja kuvatekstejä (eli tärkeimmät objektit uutisvideoissa) sekä tunnistettava suhteellinen kuvakulma ja -etäisyys.

Del Bimbon [1999, 203] mukaan videosisältöjen indeksointi perustuu kahteen peruskäsitteeseen: (1) ajalliseen segmentointiin eli videon merkityksellisten segmenttien kuten otosten ja kohtausten tunnistamiseen sekä (2) sisällön analyysiin eli alueiden, objektien ja liikkeen ominaisuuksien tunnistamiseen segmenteissä. Edellä käsiteltiin otosten ja uutisjuttujen jäsentämistä. Tässä luvussa käsitellään sisällön kuvailuun eli kommentointiin käytettäviä menetelmiä niiltä osin kuin ne ovat aiheen rajauksen kannalta oleellisia. Yeon ja Yeungin [1997, 49] mukaan videon kommentoinnissa havaittaviin piirteisiin liitetään sisältöä koskevia kommentteja, jotka voivat liittyä yksittäisten kehysten paikkasidonnaisiin piirteisiin, liikettä koskeviin piirteisiin, kokonaisiin tunnistettuihin objekteihin ja niiden liikkeisiin. Auditiivisten piirteiden tapauksessa kommentointi voi tarkoittaa puheen litterointia ja puhujan tunnistusta. [Ks. Yeo &

Yeung 1997, 49.]

Petkovićin ja Jonkerin [2002] mukaan objektien ja tapahtumien tunnistaminen rajatuissa konteksteissa on mahdollista. Alueet, jotka koostuvat joukosta vierekkäisiä pikseleitä ja jotka ovat homogeenisia piirteidensä puolesta, voidaan purkaa ja jäljittää automaattisesti. Video-objekti on kokoelma alueita, jotka on yhdistetty kontekstista riippuvan kriteerin mukaisesti. Tunnistetut objektit voidaan ryhmitellä korkeammiksi semanttisiksi ryhmiksi käyttäen puurakenteita ja alue-tietämystä ('domain knowledge'). [Petković & Jonker 2000.] Lisäksi objekteilla on suhteita toisiin objekteihin ja ne voivat liikkua ja suorittaa tehtäviä; objektien liikettä ja niiden toimia kutsutaan tapahtumiksi. Petkovićin ja Jonkerin [2002] mukaan täysin automaattinen kartoitus piirteistä semantiikaksi rajoittamattomissa konteksteissa on äärimmäisen vaikeaa ja on hyvin epätodennäköistä, että se koskaan tulisi olemaan mahdollista. Myös Brunelli ja muut [1999, 94] esittävät, että yleisten objektien tunnistaminen – varsinkin rajoittamattomissa konteksteissa – on yhä mahdotonta nykyisille algoritmeille. Objektien tunnistamisessa on ainakin kaksi ongelmaa: (1) alueita ei pystytä erottamaan toisistaan tai (2) segmentoitua aluetta ei pystytä tunnistamaan objektiksi, koska siitä ei ole mallia eli etukäteen määriteltyä kuvausta siitä, millaisista piirteistä muodostuvilla alueilla on minkäkinlainen semanttinen merkitys. Del Bimbo [1999, 26] käyttää pisteistä, viivoista ja objekteista yleisnimitystä olio. Tällä hetkellä tehokkaita algoritmeja on kehitetty kahdelle luokalle objekteja, tekstille ja kasvoille [Brunelli et al. 1999, 94–95]. Ne ovat myös keskeisessä osassa uutisvideoissa, joten tässä luvussa keskitytään niihin.

4.2.1 Objektien sijainti ja tyypittely

Prabhakaran [1997, 71] luettelee seuraavat vaiheet piirteiden poimimisessa kuvista:

1. *Objektien paikallistaminen kuvasta*: Kuva segmentoidaan alueiksi tai objekteiksi. Segmentointia varten vaaditaan algoritmi, joka osaa eristää yksittäisiä objekteja.
2. *Objekteja esittävien piirteiden valitseminen*: Valitaan ne piirteet, joiden katsotaan edustavan kutakin objektityyppiä.
3. *Matemaattisen perustan muodostaminen*: Tarvitaan periaatteita, joilla objektit erotetaan toisistaan niiden piirteiden avulla (vrt. etäisyysmitat).
4. *Säädettävien parametrien harjoituttaminen*: Kynnykset ja muut säädettävät arvot asetetaan sellaisiksi, että niiden avulla voidaan luokitella objekteja. [Prabhakaran 1997, 71.]

Kuvien segmentointi voi perustua objektien välisten rajojen tunnistamiseen tai pikselien määräämiseen objektien sisä- tai ulkopuolelle kuuluviksi. *Alueen kasvattamistekniikka* ('region growing technique') kasvattaa aluetta potentiaalisen objektin sisäpuolella, kunnes sen reunat tulevat vastaan. Menetelmä toimii niin, että kuva jaetaan pieniin alueisiin eli pikselijoukkoihin tai jopa yksittäisiin pikseleihin. Tämän jälkeen tunnistetaan objekteja erottavat ominaisuudet, kuten harmaatasot, värit tai tekstuurit, joille annetaan ('assign') arvo jokaista aluetta varten. Vierekkäisten alueiden välisiä rajoja tutkitaan vertaamalla jokaiselle ominaisuudelle määrättyjä arvoja. Jos ero on alle tietyn arvon, kahden alueen välinen ero liuotetaan. Tätä jatketaan niin kauan, kunnes ei enää löydetä rajoja liuotettavaksi. *Kynnystämistekniikassa* ('thresholding technique') kaikki pikselit, joiden harmaatasot vastaavat tai ylittävät objektille asetetun kynnyksen, katsotaan kuuluviksi kyseiseen objektiin. Kynnykset on asetettava huolellisesti, koska ne vaikuttavat rajojen sijaintiin ja tunnistettavan objektin kokoon. [Prabhakaran 1997, 71–72.]

Visuaalisten piirteiden jakautumista kuvassa voidaan käyttää luokiteltaessa (ja haettaessa) kuvia sisällön perusteella [Prabhakaran 1997, 99–102]. Esimerkiksi valokuvan rakenne voidaan johtaa kuva-avaruudessa sijaitsevasta joukosta, joka muodostuu reunoista ja kulmista, ja jossa sijaitsevat pisteet, viivat, alueet ja objektit muodostavat paikkasidonnaisia olioita ('entities'). Kun visuaaliset oliot on paikallistettu, niiden väliset suhteet lasketaan automaattisesti paikkajäsentimellä ('spatial parser'). Paikkasidonnaiset suhteet olioiden välillä käsittävät yleensä kuvan relevanteimmat ja säännöllisimmät informaation osat, mutta ne ovat yleensä epämääräisiä ja vaikeasti määriteltävissä. [Del Bimbo 1999, 26–27, 161.] Kun kuvat on segmentoitu eli jäsennetty paikkasidonnaisesti, kuvan objektit (eli oliot) luokitellaan haluttujen piirteiden perusteella. Tähän liittyvät edellisen luettelon kohdat 2–4: piirteiden valitseminen, matemaattisen perustan muodostaminen ja säädettävien parametrien harjoituttaminen. Näiden kohtien soveltaminen riippuu tietysti siitä, minkä tyyppisiä objekteja halutaan luokitella. Esimerkiksi kasvojen tunnistuksessa tunnistettavat objektit ovat vasen silmä, oikea silmä, nenä, suu, korvat ja muut kasvojen silmiinpistävät piirteet. Kasvojen aliobjektien väliset sijainnit ja suhteet tiedetään ennalta, kun kyseessä on apriorinen malli. Kasvontunnistus etenee niin, että ensin paikallistetaan kasvojen ääriviivat ja sitten silmät; tätä jatketaan kunnes kaikki kasvojen aliobjektit on tunnistettu ennalta oletetuista sijainneista. [Prabhakaran 1997, 72–73.] Olioiden väliset suhteet voidaan luokitella niiden sisältämien geometristen piirteiden avulla (1) johdattaviin ('directional') suhteisiin (joita kutsutaan myös projektiivisiksi suhteiksi) ja (2) topologisiin suhteisiin. Johdattaviin suhteisiin kuuluvat suunnat kuten vasemmalle (jostain), oikealle, ylös ja alas. Suunnat päätellään suhteessa johonkin kuvan olioon tai ulkoiseen viitekehykseen ('reference frame'). Suhteisiin kuuluvat myös etäisyys ja kulma. Topolo-

gisiin suhteisiin ei kuulu etäisyyden käsitettä vaan läheisten olioiden erotus ('disjunction'), läheisyys ('adjacency'), sisältyminen ('containment'), kierto ('rotation') ja limittäisyys. [Del Bimbo 1999, 26–27, 161.]

Esittävien rakenteiden, joita kutsutaan usein paikkasidonnaisiksi indekseiksi, monimutkaisuus riippuu kyselyjen tyypistä ja tarvittavasta päättelystä [Del Bimbo 1999, 27]. Esittävät rakenteet voidaan jakaa (1) objektipohjaisiin rakenteisiin, jotka käsittelevät paikkasidonnaisia suhteita ja visuaalista informaatiota yhtenä erottamattomana kokonaisuutena ('entity'), sekä (2) relationaalisiin eli suhdepohjaisiin rakenteisiin, jotka erottavat paikkasidonnaiset suhteet ja visuaalisen informaation. Objektipohjaisissa menetelmissä paikkaa koskevia käsitteellisiä suhteita ei tallenneta eksplisiittisesti, sillä niissä visuaalinen informaatio sisältyy itse esitykseen. Algoritmit hakevat paikkasidonnaiset suhteet tutkimalla objektien koordinaatteja. Objektipohjaiset rakenteet perustuvat avaruuden ositustekniikkaan ('space partitioning technique'), joka mahdollistaa objektin paikallistamisen avaruudessa, jota se pitää hallussaan ('occupy'). Relaatiopohjaiset rakenteet eivät säilytä visuaalista informaatiota vaan ainoastaan paikkasidonnaiset suhteet, joihin viitataan myös mallinnusavaruutena ('modelling space'); turhat suhteet jätetään huomiotta. Relaatiopohjaisissa rakenteissa objekteja edustetaan symbolisesti ja niiden paikkasidonnaiset suhteet ilmaistaan eksplisiittisesti. [Del Bimbo 1999, 161.] Relaatiopohjaiset rakenteet ovat suositeltavia niitä sovelluksia varten, jotka joutuvat käsittelemään epätarkkaa dataa esimerkiksi, jos samassa avaruudessa on eritasoisia yksityiskohtia [Del Bimbo 1999, 162].

Idrisin ja Panchanathanin [1997, 152] käsittelemässä tekniikassa kuva muutetaan symboliseksi ja siinä esiintyviä objekteja edustetaan symbolisin kuvin. Symboliset kuvat saadaan raakakuvista soveltamalla reunojen ('edge') tunnistamista ja alueiden segmentaatiota sekä luokittelemalla kuvan alueet objekteiksi. Jokainen objekti korvataan sitten symbolisella merkillä ('label'). [Idris & Panchanathan 1997, 152; Del Bimbo 1999, 165.] Symboleja voidaan käsitellä kaksiulotteisina merkkijonoina indeksissä ja kyselyissä, jolloin objektien ja kuvien hakemisen problematiikka vähennetään kaksiulotteisten merkkijonojen täsmäyttämiseksi [Idris & Panchanathan 1997, 152]. Idris ja Panchanathan [1997, 152–154] sekä Prabhakaran [1997, 99–100] käsittelevät algoritmeja kuvien indeksoimiseen kehysten paikallisten suhteiden perusteella.

4.2.2 Kasvojen tunnistaminen

Kasvontunnistus ('face recognition') on useampivaiheinen prosessi, jonka päävaiheita ovat

kasvojen havaitseminen ('detection') ja niiden nimeäminen. [7] Grossin, Shin ja Cohnin [2001, 1] mukaan kasvontunnistusalgoritmit voidaan jakaa kuvakaavioihin ('image template') perustuviin ja piirteisiin perustuviin menetelmiin. Kaavioihin perustuvat menetelmät käyttävät tunnistamisessa kasvokuvien globaaleja ominaisuuksia ja laskevat korrelaation kasvojen ja yhden tai useamman mallikaavion välillä. Kirjallisuudessa on käsitelty useita eri menetelmiä mallien tuottamiseksi, näitä ovat muun muassa tukivektorikoneet ('Support Vector Machines'), pääkomponenttianalyysi ('Principal Component Analysis'), todennäköisyyden tiheyden arviointi ('Probability Density Estimations') ja useat neuroverkkoihin pohjautuvat menetelmät. (Lisää menetelmiä mainitsevat Brunelli et al. 1999, 91–92 ja Gross et al. 2001, 1.) Piirrepohjaiset menetelmät perustuvat paikallisiin kasvopiirteisiin ja niiden geometrisiin suhteisiin: näistä menetelmistä voidaan erikseen mainita paikallinen piirteiden analyysi ('Local Feature Analysis'), josta kertovat lisää Gross ja muut [2001, 1]. [Gross et al. 2001, 1–2, 4; Brunelli et al. 1999, 91–92.]

Kasvontunnistusalgoritmia sovelletaan normalisoituihin kasvokuviin, joista paikannetaan kasvot ja niiden silmiä koskevat alueet; ne kohdistetaan ('register') sisäiseen malliin. Useimmat kasvontunnistusalgoritmit keskittyvät suoraan edestä otettuihin kuviin, ja asennon ('pose') muutokset laskevat tunnistustarkkuutta. Viimeaikaiset testit ovat kuitenkin osoittaneet, että kasvontunnistuksessa on tapahtunut merkittävää edistymistä ja että ongelmat on hyvin ymmärretty, vaikka kaupalliset järjestelmät laahaavat hieman jäljessä. [Gross et al. 2001, 1–2.] Tunnistettujen kasvojen kokoa voidaan käyttää kuvausetäisyyden päättelemiseen [Brunelli et al. 1999, 91].

Gross ja muut [2001] evaluoivat kahta johtavaa kasvontunnistusalgoritmia, jotka ovat *MIT*:n *Bayesian Eigenface*, joka perustuu kaavioihin, ja *Visionicsin FaceIt*, joka perustuu piirteisiin [Gross et al. 2001, 3–4]. Kasvojen asennon muutos on yhä haaste kasvontunnistuksessa, mutta parhaimmillaan saavutettiin kohtuullinen 70–80 %:n tarkkuus aina 45 asteen kulmaan asti. Valaistuksen muutokset kasvoissa eivät tuottaneet ongelmia. Mitä tulee ilmeisiin, pois lukien äärimmäiset variaatiot kuten huutaminen, evaluoidut algoritmit suoriutuivat varsin hyvin. Kasvojen peittyminen esimerkiksi aurinkolaseilla vaikutti algoritmien suoritustasoon eri tavalla: FaceIt oli herkempi kasvojen yläosan peittymiselle kuin MIT, mutta selviytyi kasvojen alaosan peittymisestä paremmin. [Gross et al. 2001, 9.] Lienhart ja muut [1997, 57] mainitsevat käyttämänsä neuroverkkopohjaisen järjestelmän saavuttavan 90 %:n tarkkuuden edestä kuvattujen kasvojen tunnistamisessa.

7 Terminologia on epäselvää verrattaessa Grossia ja muita [2001] ja Del Bimboa [1999, 28]: 'recognition' tarkoittaa tunnistamista kasvontunnistuksen yhteydessä, vaikka Del Bimbon tapa käyttää termiä (ks. mts. 28) viittaa havaitsemiseen, josta Gross ja muut [2001] käyttävät termiä 'detection'. Joka tapauksessa kasvot on ensin havaittava ja sitten tunnistettava eli nimettävä.

Satoh ja muut [1999] käsittelevät *Name-It*-järjestelmää, joka ei ole tavanomainen kaavioihin perustuva kasvontunnistusalgoritmi: se osaa yhdistää uutisvideoissa esiintyviä nimiä ja kasvoja ilman apriorista kasvo–nimi-assosiaatiojoukkoa, vaikka se pystyykin myös tuottamaan automaattisesti assosiaatiojoukkoja. Name-It perustuu multimodaaliseen lähestymistapaan: järjestelmä tunnistaa kuvavirrasta kasvosekvenssejä ja poimii nimikandidaatteja tuottamistaan ääniraidan ja kuvatekstien transkriptioista. Name-It yhdistää poimittuihin kasvoihin ääniraidalta poimimiaan nimikandidaatteja ajoituksen vastaavuuden ja kasvojen samankaltaisuuden perusteella. Kuvatekstit otetaan huomioon tukevana informaationa. Järjestelmän perimmäisenä ideana on, että eri menetelmien ja informaatiokanavien (ks. luku 2.1) integroiminen parantaa yksittäisten menetelmien luotettavuutta. [Satoh et al. 1999, 22–23.]

Satoh ja muut [1999, 24] ottavat huomioon, että ei ole yhtä täydellistä menetelmää, jolla mielenkiintoisten henkilöiden kasvot voidaan poimia yksistään kuvasekvenssien analyysimenetelmien avulla. Myös mielenkiintoisten henkilöiden nimien poimiminen ääniraidan transkriptiosta vaatii syvällistä semanttista analyysia: vaikka nimet saadaan erotettua muista sanoista riittävän tarkasti, mielenkiintoisten henkilöiden nimien valitseminen niistä on huomattavasti hankalampaa. Name-It poimii kasvoja ja nimiä, jotka todennäköisesti vastaavat mielenkiintoisia henkilöitä. Tätä varten vaaditaan kasvojen havaitsemista ('detection') ja niiden jäljittämistä kasvosekvenssien poimimiseksi. Lisäksi tarvitaan luonnollisen kielen käsittelymenetelmiä, jotka käyttävät sanastoa, tesaurusta ja jäsennintä nimien paikallistamiseksi transkriptiossa. [Satoh et al. 1999, 24.] Kasvojen ja nimien yhdistämisessä on se ongelma, että transkriptiot eivät välttämättä selitä videoiden sisältöä. Koska yleensä tärkeisiin kasvokuviin liitetään henkilön nimi kuvatekstinä – joskaan ei aina eikä kaikista henkilöistä – kuvatekstejä voidaan pitää videota selittävinä elementteinä. [Satoh et al. 1999, 24–25.]

Poimittaessa nimi-informaatiota transkriptioista aihe ('topic) on uutisvideon tärkein ja hierarkkisesti korkein komponentti. Jokainen uutisaihe sisältää yhden tai useamman kappaleen ('paragraph'), jotka karkeasti vastaavat kohtauksia. Ankkuria koskeva kappale ilmestyy yleensä uutisaiheen alussa, jossa ankkuri antaa yleiskuvan aiheesta. Tämän jälkeen alkaa varsinaista uutisjuttua koskeva kappale. Uutisjuttuja koskeva kappale on se osa transkriptiosta, joka on kiinnostava Name-It menetelmän kannalta. Name-It arvioi poimittujen nimikandidaattien yhteensattumista kasvosekvenssien kanssa ja tuottaa näin kasvo–nimi-assosiaation ilman etukäteen määriteltyjä malleja. Nimikandidaateille asetetaan seuraavia ehtoja: niiden on oltava (1) substantiiveja, (2) toiminnan agentteja, (3) kandidaatti pitää mainita ennen muita ja (4) kandidaatti mainitaan yleensä hieman ennen kuin uutisjuttu näytetään, sillä haastateltava ei yleensä mainitse omaa nimeään. [Satoh et al. 1999, 29.]

Multimodaalisessa analyysissa yhdistettynä ('unified') integroivana mittana käy-

tetään yhteisesiintymiskerrointa ('co-occurence factor'), joka edustaa todennäköisyyskerrointa, jolla kuva ja nimi vastaavat toisiaan. Ongelmina tällaisessa analyysissa on kasvon tai nimen puuttuminen tai moninkertainen vastaavuus näiden välillä, jolloin yksi kuva saa useamman nimen tai päinvastoin; tästä syystä myös samalla nimellä nimettyjen kasvojen samankaltaisuutta verrataan keskenään. [Satoh et al. 1999, 25.] Kuten ääniraidan transkriptiota ja kasvoja vertailtaessa, kuvatekstien ja kuvasekvenssien vertailussa käytetään yhteisesiintymiskerrointa. Lopullista kerrointa laskettaessa otetaan huomioon kasvosekvenssien poimiminen, kasvojen täsmäytys, nimikandidaattien poimiminen ääniradasta ja kuvatekstien poimiminen. Yksittäisten menetelmien epätarkkuutta kompensoidaan siis integroimalla. [Satoh et al. 1999, 31.] Koska uutisankkuri esiintyy melkein jokaisen nimen kohdalla, sen yhteisesiintymiskerrointa lasketaan. [Satoh et al. 1999, 32.] Satoh ja muut [1999, 25–32] kuvailevat järjestelmänsä komponentteja yksityiskohtaisesti.

Satohin ja muiden [1999, 28] mukaan Name-It havaitsee kasvosekvenssit parhaimmillaan 90 %:n tarkkuudella. Uutisvideosta löydettiin 65 kasvosekvenssiä ja vain neljä jäi huomaamatta; näistä kahdessa valo heijastui silmälaseista ja kahdessa kasvot olivat varjon peittämät. Lisäksi löydettiin yksi sekvenssi, jossa ei ollut kasvoja ollenkaan, ja eräässä tapauksessa yhdistettiin kaksi kasvosekvenssiä yhdeksi. Vaikka nimikandidaattien poimimisen tarkkuus oli vain 13 % ja saanti 91 %, huonosta tuloksesta huolimatta, eri menetelmiä integroimalla järjestelmällä saavutettiin lopulta hyviä tuloksia kasvo–kuva-assosiaatiossa. [Satoh et al. 2001, 28.]

4.2.3 Kuvatekstien tunnistaminen

Kuvatekstien tunnistamiseksi on useita lähestymistapoja. Jotkut niistä keskittyvät kuvatekstitapahtumien tunnistamiseen ja jotkut tekstin tunnistamiseen yksittäisessä kehyksessä [Brunelli et al. 1999, 92]. Kuvatekstin tunnistaminen aloitetaan etsimällä tekstialueita, jotka erotetaan taustasta. Tekstiä esikäsitellään kuvanlaadun parantamiseksi, jonka jälkeen teksti tunnistetaan OCR-ohjelmalla ('Optical Chraracter Recognizion'). Satoh ja muut [1999] käyttävät kaaviopohjaista ('template based') merkkientunnistusalgoritmia, joka pystyy parhaimmillaan 76 %:n tarkkuuteen. [Satoh et al. 1999, 30–31; Brunelli et al. 1999, 92.]

Brunellin ja muiden [1999, 93] esittelemässä kuvatekstitapahtumia tunnistavassa menetelmässä otetaan huomioon, että kuvateksti ilmestyy tai katoaa otoksen keskivaiheilla. Sulautettujen kuvatekstien poimiminen perustuu kuvatekstitapahtuman ('caption event') eli tekstin välittömän tai asteittaisen ilmestymisen ja katoamisen tunnistamiseen. Välittömät ku-

vatekstitapahtumat tunnistaa suurista kehyksien välisistä eroista ('interframe') pikselien välillä, jotka usein paikallistuvat kuvan alaosaan. [Brunelli et al. 1999, 93.] Asteittaisten kuvatekstitapahtumien tunnistaminen on vaikeaa varsinkin, jos kuvassa on nopeasti liikkuvia objekteja [Brunelli et al. 1999, 95]. Toisessa lähestymistavassa, joka perustuu kehyksien väliseen ja kehyksien sisäiseen lukemiseen, pääasialliset oletukset kuvatekstin suhteen ovat: yhtenäinen väri ja kirkkaus, selvät merkkien väliset reunat, irrallaan olevat ('disjoint') merkit ja paikallaan pysyvä horisontaalisesti järjestetty ('align') teksti. Tekstitettyjä kehyksiä paikallistettaessa vertaillaan peräkkäisten kehyksien eroja pikseleiden harmaatasojen avulla ja etsitään keskimääräistä kehystä paikallistetun ajanjakson sisällä. Tämän jälkeen merkit tunnistetaan OCR-menetelmällä, ja morfologisen analyysin jälkeen substantiivit puretaan kuvateksteistä. [Brunelli et al. 1999, 93.] Kolmannessa Brunellin ja muiden [1999, 93] esittelemässä lähestymistavassa kuvatekstialuetta pidetään horisontaalisena, neliskulmaisena, ryppäytetyistä terävistä reunoista muodostuvana rakenteena, joka pääasiassa muodostuu kontrastiltaan voimakkaista merkeistä. Tässäkin menetelmässä tunnistettuja kuvatekstejä käsitellään kuvankäsittelymenetelmillä, muun muassa interpoloimalla pikseleitä ja lisäämällä kontrastia. Parhaimmillaan kuvatekstit löydetään 91.8 %:n varmuudella tekstin tunnistamisen tarkkuuden ollessa 70.1 %. [Brunelli et al. 1999, 93.]

Liikkuva teksti tunnistetaan etsimällä jokaisesta kehyksestä alueita, jotka vastaavat keinotekoisen tekstin piirteitä, ja erottamalla ne muista alueista; jokainen kehys segmentoidaan paikkasidonnaisesti käyttämällä erota-ja-yhdistä-algoritmia. Ne alueet hylätään, joiden kontrasti on heikko ja joiden geometria ei ole hyväksyttävissä. Kun ehdokasalueet on tunnistettu, ne jäljitetään kehysten yli piirrevertailujen avulla rakentamalla ketjuja merkeistä ('character') niiden ilmestymisen ja katoamisen perusteella videosekvenssissä. Tunnistetut merkit ryppäytetään ja syötetään OCR-ohjelmalle. Uutislähetysten liikkuvista teksteistä tunnistettiin 41 % ja elokuvien otsikoista 76 %. [Brunelli et al. 93–94.]

4.3 Liikkeen ja tapahtumien havaitseminen ja tunnistaminen

> Järjestelmän on havaittava liike otoksissa eli luokiteltava otokset staattisiksi tai dynaamisiksi ja tunnistettava liikkeen tyyppi (kameran liike vaiko havaittujen objektien liike) sekä liikkeen suunta.

Vaikka liike ei välitäkään TV-uutisissa keskeistä informaatiota, liikepohjaiset kyselyt voivat muodostaa hyödyllisen tukitoiminnon videotiedonhaussa. Aiemmin esitettiin, että semanttis-

ten kyselyiden rajoittuneisuuden takia hakijoiden olisi järkevintä miettiä millaisia havaittavia piirteitä ja muita ominaisuuksia heidän etsimissään objekteissa saattaisi esiintyä ja sitten hakea objekteja näiden ominaisuuksien perusteella. Liike on yksi tämänkaltainen ominaisuus, jota voidaan käyttää hakualueen rajaamisessa. Objektien liikettä voidaan käyttää myös niiden luokittelemiseksi ja järjestämiseksi hierarkkisesti.

Gupta & Jain [1997, 75] erottavat videoista kolmenlaista liikeinformaatiota: objektien liikettä, kameran liikettä ja erikoistehosteiden tuottamaa liikettä [ks. liite 2]. Objektien liike tarkoittaa suoraa muutosta paikkasidonnaisten olioiden suhteellisessa asemassa [Del Bimbo 1999, 27–28; Idris & Panchanathan 1997, 159]. Liikkeenanalyysin tarkoituksena on tuottaa kaksoishierarkia, joka muodostuu videon paikkasidonnaisista ja ajallisista osista eli objekteista ja liikkeestä. Videosekvensseissä esiintyvät objektit luokitellaan niiden piirteiden (kuten muodon, värien ja liikkeen) perusteella, ja liikkeenanalyysin avulla johdetut liikekuviot kuvaillaan. [Brunelli et al. 1999, 95.] Liikkeenanalyysi liittyy läheisesti objektien tunnistamiseen ja luokitteluun, jota käsiteltiin edellisessä luvussa, sillä moniin objektityyppeihin liittyy niille ominaisia liikekuvioita.

Bollen ja muiden [1998] mukaan liikkeen tunnistaminen aloitetaan kameran liikkeen poimimisella, josta käytetään termiä ”ego-motion problem” konenäön kirjallisuudessa. Kameran liikkeen poimimisen jälkeen voidaan tunnistaa yksittäiset liikkuvat objektit ja lopulta poimitaan objektien hahmot. [Bolle et al. 1998.] Lopulta tarjotaan kerrostettu esitys videosta. Kerroksia käytetään tunnistamaan kohtausten merkittäviä objekteja piirteiden laskemista ja kyselyjä varten. Objektien lentoratojen tunnistamisen lisäksi liikkeen analyysin avulla pystytään tunnistamaan objekteja, kameran liikkeitä ja luomaan otoksista staattisia kuvia mosaiikkistamisen avulla. [Brunelli et al. 1999, 95.]

4.3.1 Objektien liikkeen analyysi ja objektien kerrostaminen

Objektin liikkeen poimiminen edellyttää yhtenäistä liikettä sisältävien alueiden jäljittämistä peräkkäisistä kehyksistä ja objektien segmentointia eli jäsentämistä. Lähestymistavat perustuvat muutoksiin kirkkaudessa, liikkeeseen ja liikevektoreihin sekä väri-informaatioon. [Del Bimbo 1999, 238.]

Brunelli ja muut [1999, 95–96] sekä Del Bimbo [1999, 238] esittelevät lähestymistapoja kerrostetun videoesityksen tuottamiseksi liikkeenanalyysin avulla. Esitellyt menetelmät perustuvat liikemallien käyttämiseen ja kuvasekvenssien liikealueiden sovittamiseen niihin. Yhtenäiset liikealueet ('motion region') tunnistetaan iteratiivisesti muodostamalla hy-

poteeseja liikkeestä ja luokittelemalla jokainen alue kuvassa johonkin näistä hypoteeseista. Jokaista hypoteesia (eli liikepaikkaa) määritellään parametreilla, joilla kuvataan muodot säilyttävää ('affine') liikemallia, jolla voidaan kuvailla videoista tyypillisesti tavattavat liikkeet kuten pyöriminen ja tarkennus (ks. liite 2). Jos staattisia taustakohtia, missä hallitsevana ominaisuutena on liikkeen puuttuminen, kuvataan yksittäisen liikemallin läpi, niin liikkuvat objektit tunnistetaan poikkeuksina tästä liikemallista. Koska videoissa esiintyy useita samanaikaisesti liikkuvia objekteja, tarvitaan useampi liikemalli, jotka kilpailevat aluetuesta eli siitä löytyykö mallia vastaavaa liikettä kullakin toistokerralla. Ne alueet, joista löytyy samankaltaista liikettä, ryhmitellään yhdeksi kerrokseksi. Jokainen alue voidaan näin jäljittää useamman kehyksen matkalle käyttäen sitä kuvailevaa liikemallia. Lopputuloksena on joukko kerroksia, joita voidaan käyttää samankaltaisten objektien jäljittämiseen. [Brunelli et al. 1999, 95; Del Bimbo 1999, 238.]

Ne alueet, jotka esiintyvät johdonmukaisesti useiden kehysten ajan vähintään yhden havaittavan piirteen suhteen, tunnistetaan video-objektiksi [Brunelli et al. 1999, 96]. Del Bimbo [1999, 238] kutsuu yhtenäisesti liikkuvia alueita avainobjekteiksi, joiden visuaalisten ominaisuuksien avulla voidaan tehdä kyselyjä. Brunelli ja muut [1999, 96] esittelevät VideoQ-hakujärjestelmän, jossa kyselyt annetaan animoituina hahmotelmina, jotka liittävät ('assign') liikettä ja muita ominaisuuksia otoksen mihin tahansa osaan. Idris ja Panchanathan [1997, 159–160] sekä Brunelli ja muut [1999, 97] käsittelevät objektien liikkeen tunnistamista liikepohjaisten kyselyiden näkökulmasta, joiden tarkoituksena on noutaa järjestetty joukko objektisekvenssejä, joissa on samankaltaista objektien liikettä kuin mitä kyselyssä on määritelty. He esittelevät kirjallisuudessa käsiteltyä menetelmää, joka käyttää liikeinformaatiota hakuavaimena. Menetelmässä indeksointi aloitetaan osittamalla jokainen kehys neliskulmaisiin lohkoihin. Liikevektorit johdetaan kuvasarjoista lohkojen täsmäytyksen avulla ('block matching') ja kartoitetaan ('map') ajallispaikalliseen avaruuteen; jokaisen lohkon liike esitetään yksittäisenä vektorina vektoriavaruudessa. Vektorit ryppäytetään; jokaiselle vektoriryhmälle tuotetaan niitä esittävä lentorata, joka on lähinnä ryppään keskivertoa vektoria, jolla on pisin elinaika ryppäässä. Kyselyt annetaan liikeratojen muodossa ja täsmäytetään tietokantaan tallennettujen lentoratojen kanssa käyttämällä etäisyysmittaa. Hakujärjestelmän on otettava huomioon tarkkojen liikekyselyjen epävarmuus ja käytettävä joustavaa täsmäytysmenetelmää. [Brunelli et al. 1999, 97; Idris & Panchanathan 1997, 160.] Liikepohjaisiin kyselyihin sopivia hakuvälineitä käsitellään luvussa 5. Idris ja Panchanathan [1997, 160] esittelevät myös kirjallisuudessa käsiteltyä menetelmää, joka perustuu MPEG-koodausmenetelmän liikettä kompensoivaan komponenttiin. Menetelmä perustuu makrolohkoihin (eli ryppäytettyihin vektoreihin eli vektoriryhmiin) ja niistä poimittaviin liikeratoihin: hierarkkinen klusterointialgoritmi aloit-

taa niistä ryppäistä, joilla on vain yksi lentorata ja askelittain yhdistää vierekkäisiä ryppäitä keskiarvottamalla niiden vektoreita. [Idris & Panchanathan 1996, 160.]

Eräs Idrisin ja Panchanathanin [1997, 160] sekä Brunellin ja muiden [1999, 97] käsittelemä menetelmä perustuu liikkuvan objektin reitin ('track') esittämiseen seuraavien primitiivisten liiketyyppien avulla:

1. Käännös ('translation'): perusilmansuunnat (pohjoinen, koillinen, itä, kaakko jne.)
2. Käännös syvyyssuunnassa ('translation in depth'): kameraan päin ja poispäin kamerasta
3. Kierto ('rotation'): myötäpäivään ja vastapäivään
4. Kierto syvyyssuunnassa ('rotation in depth'): kierto vasemmalle ja oikealle, kierto ylöspäin ja alaspäin.

Lopuksi on otettava huomioon, että on yleisesti ottaen vaikeaa poimia liikkuvia objekteja videoista ja vielä vaikeampaa tunnistaa niitä. On kuitenkin mahdollista arvioida objektien liikettä, vaikka niitä ei tarkasti tunnistettaisikaan. [Brunelli et al. 1999, 97.]

4.3.2 Kameran liikkeen analyysi ja erityiset tapahtumat

Brunelli ja muut [1999, 96] toteavat, että kameran käytön tunnistaminen on erittäin tärkeää otosten analyysin ja luokittelun kannalta, sillä se usein eksplisiittisesti heijastaa ohjaajan tarkoituksia kommunikoinnissa. Myös Del Bimbo [1999, 10–11] on samoilla linjoilla: kameran liike ja sijainti ovat tärkeitä analysoitaessa ohjaajan tyyliä, ja hän mainitsee myös liikkeellä olevan usein semanttista merkitystä. Idris ja Panchanathan [1997, 160] sekä Brunelli ja muut [1999, 96] erittelevät kuusi perustavaa kameran liikkeen tapaa, joista ja joiden yhdistelmistä kameran käyttötavat muodostuvat:

1. Panorointi ('panning') eli horisontaalinen pyörimisliike ('rotation')
2. Kiepautus ('tilting') eli vertikaalinen pyörimisliike
3. Jäljitys ('tracking') eli horisontaalinen poikittainen liike
4. Nostaminen ('booming') eli vertikaalinen liike ('transverse')
5. Siirrot ('dollying') eli horisontaalinen sivuttainen liike
6. Zoomaus eli tarkennus.

Jokainen näistä käyttötavoista saa aikaan ('induce') erityisen kuvion liikevektoreiden kentässä kehyksestä toiseen.

Kameran liike voidaan poimia sekä pakkaamattomasta että pakatusta videovirrasta. Ensiksi mainittu lähestymistapa perustuu kahden peräkkäisen kehyksen pikselien liikekentän ja virtauskentän ('flow field') analysointiin. Jälkimmäinen lähestymistapa perustuu pakattuun videodataan koodattujen liikevektorien analysointiin. [Del Bimbo 1999, 232– 233.] Del Bimbo [1999, 233–236] käsittelee liikevektoreihin perustuvia menetelmiä, joilla poimitaan erilaisia kameran liikkeitä; optiseen virtauskenttään perustuvia menetelmiä käsitellään tarkemmin sivuilla 236–237. Brunelli ja muut [1999, 96–97] esittelevät kirjallisuudessa käsiteltyjä menetelmiä kameran liikkeiden tunnistamiseksi.

Panorointia, kiepautusta ja zoomausta varten on esitetty yksinkertaisia menetelmiä, jotka perustuvat joko optiseen virtauskenttään tai MPEG:n kaltaisiin pakkausalgoritmeihin. Kameran liikkeen tunnistamisessa ensimmäinen vaihe on erottaa staattiset ja liikettä sisältävät kohtaukset, mikä voidaan tehdä tarkastelemalla liikevektorien keskivertoa kokoa. Liikevektorikenttä minkä tahansa panoroinnin ja kiepautuksen yhdistelmän kanssa näyttää yhden vahvan modaalisen vektoriarvon, jolla on selvä suunta, joka vastaa kameran liikkeen suuntaa. Useimmat liikevektorit ovat rinnakkaisia tälle vektorille. [Brunelli et al. 1999, 96–97; Del Bimbo 1999, 233.]

Idris ja Panchanathan [1997, 160] sekä Antani ja muut [2002, 956] esittelevät kirjallisuudessa mainittua menetelmää, jossa liikevektoreita ja niiden Hough-muunnoksia käytetään edellä mainittujen kamerankäyttötapojen tunnistamiseen. Menetelmä perustuu siihen, että liikevektorien jälkeensä jättämää kuviota kuvaa fyysisesti ja paikkasidonnaisesti (1) liikevektorien suuruus ('magnitude') sekä (2) niiden eriäväisyyden ('divergence') ja yhdentymisen ('convergence') kohta. Kyseinen liikevektoreihin perustuva menetelmä on kuitenkin herkkä hälylle ja se vaatii paljon laskenta-aikaa. [Idris & Panchanathan 1997, 160–161; Antani et al. 2002, 956.] Idrisin ja Panchanathanin [1997, 160–161] käsittelemässä vaihtoehtoisessa menetelmässä ensin tunnistetaan otoksen kaikista kehyksistä reunat, minkä jälkeen otetaan horisontaalisessa suunnassa painotettu integraali reunakehyksistä ja saadaan horisontaalinen “röntgenkuva” ('X-Ray image'); sama tehdään reunakehyksille vertikaalisessa suunnassa. Kameran liikkeet saadaan arvioimalla horisontaalisten ja vertikaalisten “röntgenkuvien” reunojen kulmien ('angle') paikkasidonnaista jakaumaa. Reunojen etsiminen kaikista kehyksistä on kuitenkin laskennallisesti raskasta. [Idris & Panchanathan 1997, 160–161.] Idrisin ja Panchanathanin [1997, 161–162] mukaan näiden menetelmien ongelmana on, että ne eivät erota jäljittämistä ja panorointia tai nostamista ja kallistelua. Ylipäätään kameran liikkeiden tunnistamisen onnistumiseksi kuvassa ei saisi olla suuria liikkuvia objekteja, jotka hallitsevat näkö-

kenttää, sillä ne aiheuttavat väärän tunnistuksen; objektit on tunnistettava ensin, ja niiden liikkeet on otettava huomioon kameran liikkeitä tunnistettaessa. [Idris & Panchanathan 1997, 161–162.] Myös Brunelli ja muut [1999, 97] ottavat huomioon isojen objektien liikkeiden vaikutuksen kameran liikkeiden tunnistamiseen. Antani ja muut [2002, 956–957] käsittelevät lisää muun muassa kameran liikkeen tunnistukseen sopivia menetelmiä.

Varsinkin uutisvideoissa salamavalojen kanssa samaan aikaan kuvassa esiintyy tärkeitä henkilöitä [Brunelli et al. 1999, 94]. Brunelli et al. [1999, 94] mainitsevat kirjallisuudessa käsitellyn algoritmin, joka tunnistaa salamavaloja. Salamavalot ilmenevät algoritmin kannalta kahtena terävänä piikkinä, jotka ovat jotakuinkin arvoltaan samankaltaisia kehyksien välisten erojen esityksessä, arviolta puolen sekunnin aikaikkunassa. Algoritmin saanti on 40 % ja tarkkuus 100 %. [Brunelli et al. 1999, 94.]

4.4 Ääniraidan jäsennys ja tunnistaminen

> Järjestelmän on jäsennettävä ääniraita tyyppeihin (eli puheeseen, taustaääniin ja musiikkiin) ja tunnistettava puhetta, puhuja ja käytetty kieli.

Televisiossa ja erityisesti uutislähetyksissä äänellä on keskeinen osa informaation välittäjänä. Ääni ei pelkästään välitä informaatiota, vaan ääniraitaa voidaan käyttää apuna kontekstin määrittelemisessä, kun yritetään johtaa semantiikkaa videokuvan havaittavista piirteistä; yhdessä ääni ja kuva välittävät enemmän informaatiota kuin mikään media yksin [Petković & Jonker 2000]. Ääniraidan indeksoiminen on siis olennainen tukitoiminto videon automaattisessa indeksoinnissa. Koska puhe välittää suurimman osan uutislähetyksien informaatiosta, ja äänitehosteiden ja musiikin merkitys TV-uutisissa on vähäinen, tarkastelu tullaan rajoittamaan puheeseen. Sheridanin ja muiden [1997, 100] mukaan puhehaun tutkimus on pitkään keskittynyt siihen, kuinka automaattisesti tunnistaa ('identify') ja indeksoida informaatiota puhesignaalista. Tässä tehtävässä vaaditaan puheentunnistusteknologiaa esimerkiksi sanojen tunnistamista varten. [Sheridan et al. 1997, 100.]

4.4.1 Puheentunnistus

Puhetta varten tuotettava metadata voi Prabhakaranin [1997, 62] mukaan koskea

1. puhuttujen sanojen tunnistamista ('identification'), jota kutsutaan puheentunnistukseksi

2. puhujan tunnistamista, joka tarkoittaa puhuvan henkilön identiteetin valitsemista joukosta tunnettuja puhujia
3. prosodisen [8] informaation tunnistamista, jota voidaan käyttää huomion kiinnittämiseksi johonkin fraasiin tai lauseeseen.

Prabhakaranin [1997, 63] mukaan puheentunnistuksen yleisin muoto, jossa ei ole rajoituksia käytetyn sanaston tai puhujien määrän suhteen, on vielä epätarkka. Tunnistamistarkkuutta voidaan parantaa ottamalla huomioon, että eristyksissä olevat sanat on helpompi tunnistaa kuin jatkuva puhe, jossa sanojen välisten rajojen tunnistaminen voi tuottaa vaikeuksia. Mahdollisimman pieni puhujien määrä ja sanasto laskevat todennäköisyyttä sille, että sanastossa on useampia samalta kuulostavia sanoja. Lisäksi kielioppi rajoittaa sitä, missä järjestyksessä sanat voivat esiintyä; sanajärjestystä voidaan hyödyntää varsinkin epäselvien tapauksien kohdalla. Myös ympäristö, jossa tunnistettava puhe on tuotettu, vaikuttaa puheentunnistuksen tarkkuuteen; ympäristöön kuuluvat muun muassa taustamelu ja mikrofonin sijainnin muutokset. [Prabhakaran 1997, 63–64.]

4.4.1.1 Puheentunnistusjärjestelmä

Prabhakaranin [1997, 64] mukaan tyypillisessä äänentunnistusjärjestelmässä on on kaksi komponenttia: signaalinkäsittely ('signal processing module') ja hahmontunnistus ('pattern matching module'). Signaalinkäsittelykomponentti muuttaa analogisen signaalin digitaaliseksi. Digitaalista näytettä prosessoidaan etsimällä siitä hiljaisia kohtia, erottamalla puhe muista äänistä ja muuttamalla raaka aaltomuoto taajuusalue-esitykseksi ('frequency domain representation') sekä pakkaamalla näytettä. Ääninäyte ryhmitellään sanavartaloiksi ('frame'), jotka ovat yleensä pituudeltaan 10-30 millisekuntia. Tarkoituksena on säilyttää vain ne osat, jotka ovat hyödyllisiä puheentunnistuksen kannalta. [Prabhakaran 1997, 64.] Signaalinkäsittelyvaiheessa mallinnetaan foneettisia yksiköitä (eli foneemeita, sanoja yms.). Näytteistetty puhesignaali muutetaan sekvenssiksi koodeja, jotka esittävät prototyyppivektoreita eli sentroideja. Vaiheen lopussa dokumentti esitetään vektorikoodien sekvenssinä, jotka karkeasti ottaen edustavat äänielimistön ('vocal tract') erilaisia muotoja. [Glavitsch & Schäuble 1992, 171.]

Prabhakaranin [1997, 65] mukaan prosessoitua puhenäytettä käytetään puhuttujen sanojen, puhujan tai prosodisten painotusten tunnistamiseen. Sanat ja muut foneettiset yksiköt ovat indeksointipiirteitä, joita puheentunnistuskomponentti tunnistaa [Glavitsch &

8 Prosodinen tarkoittaa puheessa ilmeneviä korostuksia ja äänenvärin muutoksia.

Schäuble 1992, 171]. Prabhakaranin [1997, 65] mukaan puhetta tunnistettaessa prosessoitu puhe täsmäytetään tallennettuihin hahmoihin. Hahmontunnistuskomponenttiin kuuluu varasto viitekuvioita ('reference patterns'), jotka muodostuvat

- saman sanajoukon erilaisista ääntämyksistä ('utterance') puheentunnistusta varten
- saman puhujan erilaisia ääntämyksiä puhujan tunnistamista varten
- saman sanan erilaisia muunnoksia prosodisen informaation tunnistamiseksi. [Prabhakaran 1997, 65.]

Puheen tunnistamiseksi puhedataa verrataan tallennettuihin harjoituskaavioihin ('template') tai -malleihin. Tätä varten algoritmien on laskettava samankaltaisuusmitta kaavioiden ja näytteiden välille. Seuraavaksi käsitellään algoritmeja, jotka laskevat näytteiden ja mallien välisen samankaltaisuuden. [Prabhakaran 1997, 65.]

4.4.1.2 Hahmontunnistusalgoritmit

Suosittuja puheentunnistusalgoritmeja ovat Prabhakaranin [1997, 65] mukaan dynaaminen ajan liittäminen ('Dynamic Time Warping'), kätketyt Markovin mallit ('Hidden Markov Models') ja keinotekoiset neuroverkot ('Artificial Neural Networks').

Dynaamisessa ajan liittämisessä lähdetään siitä, että puhenäytteen vertaaminen mallikaavioon on käsitteellisesti yksinkertaista, jos esikäsiteltyä puheaaltomuotoa ('speech waveform') verrataan suoraan viitekaavioon ('a reference template') laskemalla yhteen asianomaisten puhevartaloiden ('speech frame') väliset etäisyydet. Kyseinen etäisyyksien summa tarjoaa yleisen etäisyysmitan samankaltaisuuden laskemista varten. Kuitenkin epälineaariset ajoituksen vaihtelut lausumien välillä aiheuttavat virheen puhuttujen sanojen vartaloiden kohdistamisessa ('alignment') viitekaavioon. Kaaviota voidaan kuitenkin venyttää tai tiivistää soveliaissa kohdissa optimaalisen täsmäytyksen löytämiseksi. Kyseisstä ajan poimuttamisen ('warping') ja liittämisen prosessia kaaviolla kutsutaan dynaamiseksi ajan liittämiseksi. Tarkoituksena on löytää liitos ('warp'), joka minimoi etäisyyksien summan kaavioiden täsmäytyksessä. [Prabhakaran 1997, 65.]

Kätketyt Markovin mallit ovat stokastisia eli ne perustuvat arvauksiin ja todennäköisyyksiin. Kätkettyihin Markovin malleihin kuuluu stokastisia, perustavia äärellisiä tilakoneita ('finite state machines'), joita määrittää joukko tiloja, tulostusaakkosto ja joukko siirtymä- ja tulostustodennäköisyyksiä. [Prabhakaran 1997, 65–66; Glavitsch & Schäuble 1992,

171.] Sanan kätketty Markovin malli rakennetaan kaaviolla, jolla on joukko tiloja, joiden väliset linkit ('arcs') esittävät positiivista todennäköisyyttä siirtymälle ('transition'). Jos {s1, s2, s3, s4} on joukko tiloja ja {h,e,l,o} on tulostettavien aakkosten joukko, kätketty Markovin malli on suunniteltu tunnistamaan sana “hello”. Siirtymän todennäköisyys määritellään jokaisen tilaparin välillä (s1 ja s2, s2 ja s3, jne.). Tulostustodennäköisyydet liitetään ('associate') jokaiselle siirtymälle (s1 --> s2, s2 --> s3, jne.), ja ne määrittävät todennäköisyyden kunkin tulostusaakkosen tuottamiselle ('emit') siirtymässä. [Prabhakaran 1997, 65–66.] Todennäköisyys tietyn sanan tulostamiselle kasvaa jokaisella siirtymällä tilojen välillä: jos tulostettuna on tilat “h”, “e” ja “l”, vaihtoehtoisten sanojen joukko on paljon pienempi ja todennäköisyys se tulostamiselle on suurempi kuin ensimmäisen tilan (“h”) jälkeen. Prabhakaranin [1997, 65–66] mukaan menetelmän nimessä “kätketty” viittaa siihen, että äärellisen tilakoneen tosiasiallista tilaa ei voida tarkastella suoraan vaan ainoastaan niiden tuottamien aakkosten läpi. Kätketty Markovin malli luo satunnaisia sekvenssejä siirtymä- ja tulostustodennäköisyyksien määrittämän jakauman mukaan. Sanaston jokaisella sanalla on kätketty Markovin malli. Kätkettyjä Markovin malleja pitää harjoituttaa, jotta ne tunnistaisivat irrallisia sanoja tai jatkuvaa puhetta. Harjoittamisen tarkoitus on lisätä todennäköisyyttä, jolla kätketty Markovin malli luo halutun tulostussekvenssin. Harjoitusdata muodostuu annetusta joukosta tulostussekvenssejä eli niistä sanoista, jotka mallin halutaan tunnistavan. [Prabhakaran 1997, 65–66.] Harjoittamista varten kätkettyihin Markovin malleihin yhdistetään seuraavat komponentit:

1. Eteenpäinsyöttöalgoritmi ('forward algorithm'), jota tarvitaan irrallisten sanojen tunnistamiseksi. Algoritmin tarkoitus on laskea todennäköisyys sille, että kätketty Markovin malli luo halutun tulostussekvenssin. Sekvenssi tunnistetaan tietyksi sanaksi, jos todennäköisyys sille, että vastaava kätketty Markovin malli tulostaa sanan, on maksimaalinen.
2. Baum–Welch-algoritmi, jota käytetään foneemimallien harjoittamiseen.
3. Viterbi-algoritmi, jolla tunnistetaan jatkuvaa puhetta. Algoritmi määrittää tilojen siirtymisen polun, joka perustuu tunnistettavan jatkuvan puheen kielioppimalliin. [Prabhakaran 1997, 67; Glavitsch & Schäuble 1992, 171.]

Kätketyt Markovin mallit ovat yleisin lähestymistapa foneettisten yksikköjen mallintamiseen [Glavitsch & Schäuble 1992, 171].

Keinotekoiset neuroverkot ovat informaation käsittelyjärjestelmiä, jotka simuloivat ihmisaivojen kognitiivisia prosesseja. Perustavana ideana niissä on rakentaa neurorakenne, jota voidaan harjoituttaa suorittamaan syötesignaalien kognitiivinen tehtävä. Neuroverkko

muodostuu lukuisista yksinkertaisista ja keskenään liitetyistä suorittimista eli neurodeista ('neudodes'), jotka vastaavat aivojen neuroneita. Neuroverkko on jaettu kerroksiin, joiden neurodit ottavat vastaan ja lähettävät eteenpäin päätöksiään. Neurodit on yhdistetty lukuisilla linkeillä, joihin on yhdistetty painotettuja tehtäviä. Neurodit viestivät päätöksistään painotettujen linkkien avulla; päätökset saattavat saada eri painoja eri linkkejä varten. Linkkien painot määritetään harjoittamalla neuroverkkoa. Harjoitusvaihe muodostuu syötedatasta, kuten puhekaavioista ja halutun tulostuksen kuvaamisesta: neuroverkko oppii tunnistamaan syötedataa ja määräämään linkeille painot. [Prabhakaran 1997, 67–68.]

4.4.1.3 Indeksointipiirteet

Puhetta indeksoidaan kääntämällä alkuperäinen puhedata tekstuaaliseen muotoon eli transkriptioksi [Abberley, Kirby, Renals & Robinson 1999]. Tunnistettua ja litteroitua puhedataa voidaan siten pienin varauksin käsitellä kuin mitä tahansa tekstiä [Prabhakaran 1997, 95]. Esimerkiksi Glavitsch ja Schäuble [1992, 168] esittelevät multimediahakumallia, jossa puhetta ja tekstiä indeksoidaan yhtenäisesti. Heidän hakumallinsa on yhdistelmä tekstihakumallia (vektoriavaruusmalli) ja puheentunnistusmallia (kätketty Markovin malli). Kätkettyjä Markovin malleja käytetään mallintamaan foneemien, sanojen tai fraasien erilaisia ääntämyksiä. [Glavitsch & Schäuble 1992, 168.] Puhetta indeksoitaessa indeksointipiirteidessä on kuitenkin joitain rajoituksia [Prabhakaran 1997, 95].

Glavitschin ja Schäublen [1992, 168–169] mukaan indeksointisanasto ('indexing vocabulary') koostuu indeksointipiirteistä ('indexing features'). Perinteisessä tekstihaussa indeksointipiirteet koostuvat sanavartaloista. Puhetiedonhaussa tarvitaan myös tarkoituksenmukaisia indeksointipiirteitä, jotka puheentunnistusjärjestelmä voi tunnistaa. [Glavitsch & Schäuble 1992, 168–169.] Indeksointipiirteet ovat tekstuaalisia yksikköjä, jotka on mallinnettu foneettisesti niin, että ne voidaan tunnistaa puhedokumenteista. Tunnistamisessa vaaditaan puheentunnistuskomponenttia, aivan kuten tekstidokumenttien indeksointipiirteiden tunnistamisessa vaaditaan soveliasta jäsennintä. [Glavitsch & Schäuble 1992, 170.] Glavitsch ja Schäuble [1992, 168–169] sekä Prabhakaran [1997, 96] esittävät seuraavia vaatimuksia puhedokumenttien indeksointipiirteille:

1. Indeksointipiirteiden on oltava foneettisia yksikköjä, jotka ovat puheentunnistusmenetelmien helposti tunnistettavissa.
2. Erilaisten indeksointipiirteiden määrän on oltava niin pieni, että kätkettyjen Markovin

mallien harjoituttamiseen vaaditaan vain kohtuullinen määrä dataa; määrä riippuu jokaisen piirteen mallintamiseen vaaditun harjoitusdatan saatavuudesta.

3. Indeksointipiirteiden pitäisi olla riittävän erottelukykyisiä eli niiden pitäisi auttaa erottamaan dokumentit toisistaan, jolloin kyseisiin indeksointipiirteisiin perustuvan hakumenetelmän pitäisi saavuttaa hyväksyttävä hakutehokkuus.
4. Indeksointipiirteet pitäisi voida tunnistaa myös tekstidokumenteissa, jotta puhe- ja tekstidokumentteja voidaan hakea samanaikaisesti.
5. Indeksointipiirteiden keräämistiheys ei saa olla liian pieni. [Glavitsch & Schäuble 1992, 168–169; Prabhakaran 1997, 96.]

Mahdollisia indeksointipiirteitä ovat foneemit ('phonemes'), foniparit ('biphones'), fonikolmikot ('triphones'), sanat tai fraasit [Glavitsch & Schäuble 1992, 169; Abberley et al. 1999]. Sanastopohjaisessa puheentunnistuksessa Srinivasanin ja Petkovicin [2000, 81] mukaan sanastolla tarkoitetaan joukkoa sanoja, joita puheentunnistusmoottori tunnistaa kääntäessään puhetta tekstiksi. Dekoodausprosessissa moottori täsmäyttää puhesignaalia sanaston sanoihin. Vain sanastossa esiintyvät sanat voidaan tunnistaa. [Srinivasan & Petkovic 2000, 81.] Sheridan ja muut [1997, 100] antavat esimerkkejä sanojen tunnistamiseen perustuvien puhehakujärjestelmien parissa tehdystä tutkimuksesta ja järjestelmien ongelmista. Glavitschin ja Schäublen [1992, 169] sekä Prabhakaranin [1997, 96] mukaan sanat ja fraasit vaikuttavat liian laajoilta indeksointipiirteiksi. Sanoja esiintyy dokumenteissa liikaa, ja osa niistä ei ole erottelukykyisiä. Koska sanasto pitää määritellä etukäteen, se on rajoittunut laajuudeltaan; harvemmin käytetyt sanat, kuten nimet, saattavat hyvinkin puuttua sanastosta eikä eksoottisia sanoja voida harjoituttaa tarpeeksi tai välttämättä ei ollenkaan. [Glavitsch & Schäuble 1992, 169; Sheridan et al. 1997, 100.] Usein sanastosta löytymättömät sanat tunnistetaan virheellisesti joksikin toisiksi sanoiksi, jotka ovat foneettisesti samankaltaisia [Srinivasan & Petkovic 2000, 81].

Srinivasanin ja Petkovicin [2000, 81] mukaan sanastopohjaisten lähestymistapojen ongelmia on yritetty ratkaista tutkimalla foneemeihin perustuvien osasanaesitysten ('subword') käyttöä indeksointitermeinä. Foneemilla tarkoitetaan mitä tahansa kielen foneettisen järjestelmän abstraktia yksikköä, joka vastaa joukkoa ('correspond') kielestä havaittavia yksittäisiä ja erillisiä puheääniä. Foneemit ovat ihmisen havainnointikyvyn määrittelemiä; foneiksi ('phone') kutsutaan vastaavia puheentunnistusjärjestelmän määrittelemiä ja datasta johtamia alisanojen yksikköjä. [Srinivasan & Petkovic 2000, 81.] Glavitschin ja Schäublen [1992, 169] mukaan foneemien etuna on niiden suhteellisen pieni määrä ja suuri keräämistiheys, mutta ne ovat kuitenkin liian pieniä yksiköitä indeksointipiirteiksi: foneemien tunnistaminen on vaikeaa ja ne esiintyvät liian usein ollakseen erottelukykyisiä. [Glavitsch & Schäub-

le 1992, 169.] Srinivasanin ja Petkovicin [2000, 81] mukaan foneettisten menetelmien ongelmana on niiden rajoittunut tarkkuus verrattuna sanastopohjaisiin menetelmiin, varsinkin lyhyiden sanojen kohdalla. Sanaston ulkopuolisten sanojen ('out-of-vocabulary') hakemista varten hyödytään foneettisen ja sana-tason informaation yhdistämisestä. [Srinivasan & Petkovic 2000, 81.] Vaikka foneettinen haku sopiikin joihinkin tarkoituksiin, kuten Srinivasanin ja Petkovicin [2000] menetelmässä, ne eivät välttämättä sovi kovin hyvin uutislähetyksiin [mts. 81]. Koska hyvät indeksointipiirteet löytyvät sanojen ja foneemien välistä, Glavitsch ja Schäuble [1992, 169] ehdottavat trigrammeja hyvin erotteleviksi piirteiksi.

4.4.1.4 Puheentunnistuksen tarkkuus ja yhdistetyt menetelmät

Srinivasan ja Petkovic [2000, 81–82] viittaavat useisiin tutkimuksiin, joissa käsitellään eri puheentunnistusjärjestelmien tarkkuutta sanojen tunnistamisessa. Heidän mukaansa pelkällä foneemientunnistusjärjestelmällä on saavutettu 60–70 %:n tarkkuus uutislähetyksessä. Laajaan sanastoon perustuvalla tunnistusjärjestelmällä saavutetaan kuitenkin vielä parempia tuloksia kuin pelkkiin foneihin (eli foneemeihin) perustuvilla järjestelmillä, ja yhdistetyt menetelmät, joissa käytetään sekä sanastoa että staattisesti laskettua foniristikkoa ('phone lattice'), parantavat tuloksia entisestään: Srinivasanin ja Petkovicin [2000, 81] mainitsemassa esimerkissä saavutettiin 82–85 %:n suhteellinen tarkkuus videopostin ('videomail') yhteydessä verrattuna tekstihakuun. Yhdistetyt menetelmät toimivat paremmin kuin mikään menetelmä yksin [Srinivasan & Petkovic 2000, 81]. Myös Informedia-projektin [9] yhteydessä on tutkittu sana- ja foneemi-pohjaisten menetelmien yhdistämistä uutisjuttujen kohdalla ja saavutettu 84.6 %:n suorituskyky verrattuna tekstihakujärjestelmään [Srinivasan & Petkovic 2000, 81–82]. Myös Sheridan ja muut [1997, 101] viittaavat tutkimuksiin, joissa sanastopohjaista menetelmää on käytetty yhdessä foneemeihin perustuvan järjestelmän kanssa ja saatu tarkkuudeksi 85 % tekstihakuun verrattuna. Yhdistetyt menetelmät vaativat kuitenkin paljon tehoa tietokoneelta tunnistamis- ja harjoittamisvaiheessa. [Sheridan et al. 1997, 101.]

Srinivasan ja Petkovic [2000, 82] viittaavat kirjallisuudessa käsiteltyyn foni-pohjaiseen osasanayksikköjen indeksointiin. Käyttämällä vektoriavaruusmallia on osoitettu, että tarkoituksenmukaiset osasanayksiköt täysin virheettömien foneettisten transkriptioiden kanssa saavuttavat sanastopohjaisten järjestelmien tasoisen suorituskyvyn. [Srinivasan & Petkovic 2000, 82]. Ongelmana vain on se, että automaattisesti tuotetut foneettiset transkriptiot eivät ole yleensä lähellekään täysin virheettömiä. Sheridan ja muut [1997, 101] viittaavat Glavit-

9 Informedia Project: <URL: http://www.informedia.cs.cmu.edu/ >

schin ja Schäublen [1992] tutkimukseen vaihtoehtoisena lähestymistapana; tuossa tutkimuksessahan tavoitteena oli määritellä pieni joukko alisanoja, jotka voisivat olla tarpeeksi voimakkaita hakuja ja tunnistamista varten. Sheridanin ja muiden [1997, 106] johtopäätösten pohjalta voidaan johtaa, että järjestelmät vaativat säätämistä ja sopeuttamista kuhunkin ympäristöön, jotta niillä voitaisiin saavuttaa niiden maksimaalinen suorituskyky. Parhaimmat tulokset puheentunnistuksessa saavutetaan yhdistämällä sanasto- ja foneemipohjaisia menetelmiä – näin voidaan päätellä Srinivasanin ja Petkovicin [2000] ja Sheridanin ja muiden [1997] perusteella.

4.4.2 Kielen ja puhujan tunnistaminen

Zissmanin ja Berklingin [2001, 115] mukaan automaattinen kielen tunnistaminen on prosessi, jossa tietokone tunnistaa digitoidun puhelausuman ('speech utterance') kielen. Se on yksi niistä prosesseista, joissa puhesignaalista poimitaan tietoa – kieli voidaan tunnistaa myös tekstistä [Zissman & Berkling 2001, 115]. [10] Zissmanin ja Berklingin [2001, 115] mukaan kieli voidaan tunnistaa hakujärjestelmän puhekomentojen tunnistamisen yhteydessä tai sitä ennen. Kielen ja puheen tunnistaminen samaan aikaan vaatii jokaista kieltä varten oman puheentunnistusmoduulin. Vaihtoehtoisesti kielentunnistusjärjestelmää voitaisiin käyttää ennen puheentunnistusta poimimaan todennäköisimmät kielet, jonka jälkeen sopivin kielestä riippuvainen puheentunnistusmalli otettaisiin käyttöön. Lopullinen kielentunnistuspäätös tehtäisiin kuitenkin vasta puheentunnistuksen jälkeen. [Zissman & Berkling 2001, 115.]

Zissman ja Berkling [2001, 116] tekevät yhteenvedon niistä piirteistä, joiden avulla ihmiset ja tietokoneet voivat erottaa kielet toisistaan:

1. *Fonologia*: Foneemit ovat perustavia kielen fonologisten yksikköjen mentaalisia esityksiä. Fonit ('phone') ovat akustis–foneettisten yksikköjen tai segmenttien realisaatioita, niitä tosiasiallisia ääniä, joita puhuja tuottaa ajatellessaan tai puhuessaan foneemeista. Foneemit ja foneemijoukot ovat eri kielissä erilaisia, vaikkakin monissa kielissä yksittäiset foneemit ovatkin samanlaisia.
2. *Morfologia*: Sanavartalot ('word roots') ja leksikot ovat yleensä erilaisia eri kielissä. Jokaisella kielellä on oma sanavarastonsa ja tapansa muodostaa sanoja.

10 Ks. esim. TextCat : <URL: http://odur.let.rug.nl/~vannoord/TextCat/Demo/textcat.html>

3. *Syntaksi*: Lausekuviot ('sentence patterns') ovat erilaisia eri kielissä. Vaikka jotkin sanat saattavatkin olla eri kielissä samanlaisia, niitä edeltävät ja niitä seuraavat sanat ovat sitten erilaisia.
4. *Prosodia*: Äänen kesto ('duration'), äänenkorkeuden ('pitch') kontuurit ('contours') ja painotuskuviot ('stress patterns') ovat erilaisia kielestä toiseen. [Zissman & Berkling 2001, 116.]

Leavers ja Burley [2001, 641] luettelevat myös keinoja, joilla ihmiset erottelevat itselleen tuntemattomia kieliä toisistaan. Näihin kuuluvat: (1) suprasegmentaalinen ('suprasegmental') strategia, jossa hyödynnetään eroja rytmissä, äänenpainossa ja intonaatiossa; (2) segmentaalinen strategia, jossa hyödynnetään kielen foneettisia ominaisuuksia; (3) leksikaalinen ('lexical') strategia, jossa yksittäiset sanat tunnistetaan johonkin kieleen kuuluvaksi. [Leavers & Burley 2001, 641.] Leaversin ja Burleyn [2001, 641] mukaan automaattista kielentunnistusta voidaan tehostaa ottamalla huomioon, kuinka ihmiset käyttävät edellä mainittuja strategioita ja vihjeitä kielten tunnistamisessa. Lingvistiset vihjeet parametrisoidaan kääntämällä ne vihjeitä parhaiten kuvaavan akustisen signaalin piirteiksi. Luodut parametrit laitetaan järjestykseen edellä mainittujen kognitiivisten strategioiden perusteella. [Leavers & Burley 2001, 641–642.] Leavers ja Burley [2001, 642–648] käsittelevät lingvististen vihjeiden kääntämistä akustisen signaalin piirteiksi käyttämällä ääniväriä ('tone') globaalina prosodisena piirteenä sekä hahmontunnistamisen suprasegmentaalista, segmentaalista ja leksikaalista strategiaa.

Kielentunnistuksessa on Zissmanin ja Berklingin [2001, 116] mukaan kaksi vaihetta. Harjoitusvaiheessa järjestelmälle esitellään näytteitä eri kielistä, jotka lasketaan virraksi piirrevektoreita. Piirrevektorit lasketaan puhesignaalin aaltomuotojen lyhyistä aikaikkunoista, pituudeltaan noin 20 ms. Harjoitusalgoritmi analysoi vektorisekvenssit ja tuottaa yhden tai useamman mallin jokaista kieltä varten. Mallit edustavat joukkoa kielestä riippuvia ja perustavia harjoituspuheen ominaisuuksia, joita käytetään kielentunnistusprosessin seuraavassa vaiheessa. [Zissman & Berkling 2001, 116.] Tunnistusvaiheessa uudesta lausumasta ('utterance') laskettuja piirrevektoreita verrataan jokaiseen edellisessä vaiheessa tuotettuun kieliriippuvaan malliin. Todennäköisyys, jolla uusi lausuma on samalla kielellä kuin harjoitusdata, lasketaan ja maksimaalinen todennäköisyysmalli ('maximum-likelihood model') tunnistetaan ('found'). Sen lausuman kieli, jota käytettiin harjoituttamaan maksimaalisen todennäköisyyden tuottanutta mallia, on hypoteesi uuden lausuman kieleksi. [Zissman & Berkling 2001, 116–117.]

Zissmanin ja Berklingin [2001, 117] mukaan kielentunnistusjärjestelmät eroavat pääosin siinä suhteessa, että mitä ja kuinka monimutkaista menetelmää ne käyttävät kielten mallintamiseen. Näitä menetelmiä käsittelevät Zissman ja Berkling [2001, 117–120]:

1. *Spektraalinen samankaltaisuus-lähestymistapa* ('Spectral-similarity approaches'): Varhaiset järjestelmät korostivat kielten välisiä eroja spektraalisessa sisällössä ja hyödynsivät sitä tosiasiaa, että puhe eri kielillä sisältää eri foneemeja ja foneja. [Zissman & Berkling 2001, 117.]
2. *Prosodiapohjaiset lähestymistavat*: Prosodista informaatiota kantavia piirteitä, esimerkiksi äänenkorkeuden ('pitch') ja amplitudin korkeuskäyriä ('contour'), on käytetty syötteenä automaattisessa kielentunnistamisessa, sillä on osoitettu, että ihmiset voivat käyttää prosodisia piirteitä kielten tunnistamiseen. Prosodisen informaation hyödyllisyys ja sen tarjoama erottelukyky kielten välillä foneettisiin järjestelmiin verrattuna riippuu paljon vertailtavista kielistä. [Zissman & Berkling 2001, 118–119.]
3. *Fonin tunnistamiseen perustuvat lähestymistavat*: Koska eri kielissä on erilaisia fonivarastoja, kielentunnistusjärjestelmät voivat hypotisoida tarkasti, mitkä fonit puhutaan ajan funktiona, ja määrittää käytettävän kielen kyseessä olevan fonisekvenssin tilastojen ('statistics') perusteella. [Zissman & Berkling 2001, 119.]
4. *Monikieliset puheyksiköt*: Kielestä riippuvien foneemintunnistajien sijaan voidaan rakentaa monikielellisiä puheyksiköitä. Tutkimuksessa on etsitty myös erottelukykyisimpiä foneja, joita kutsutaan avainfoneiksi. [Zissman & Berkling 2001, 119–120.]
5. *Sana-tason lähestymistavat*: Sanoihin perustuvat lähestymistavat käyttävät hienostuneempia sekvenssinmallintamistapoja kuin foni-tason järjestelmien fonotaktiset ('phonotactic') mallit, vaikkakaan sanoihin perustuvat lähestymistavat eivät käytä täysimittaisia puheesta-tekstiin järjestelmiä. Kielessä alhaalta ylöspäin liikuttaessa ensin tunnistetaan fonit, sitten sanat ja lopulta kieli. [Zissman & Berkling 2001, 120.]
6. *Laajaan sanastoon perustuvat jatkuvan puheen tunnistusjärjestelmät*: Harjoitusvaiheessa jokaista kieltä varten luodaan oma puheentunnistaja. Testausvaiheessa jokaista tunnistajaa ajetaan rinnakkain: suurimman todennäköisyyden tuottaneen tunnistajan harjoittamiseen käytetty kieli otaksutaan syötteenä annetun lausuman kieleksi. Tämänkaltaiset järjestelmät ovat lupaavimpia, mitä tulee korkealaatuiseen kielentunnistamiseen, sillä ne ovat huomattavasti kehittyneempiä kuin foneja ja fonisekvenssejä kielentunnistuksessa käyttävät järjestelmät. [Zissman & Berkling 2001, 120.]

Zissman ja Berkling [2001, 121] käsittelevät kielentunnistusjärjestelmien suorituskykyä vuosina 1993, 1994 ja 1995 tehtyjen testien avulla. Käytettäessä lausumia pituudeltaan 45 sekuntia tai 10 sekuntia, parhaat järjestelmät tunnistivat kieliä neljän ja kahden prosentin virhetasolla ('error rate'). Parhaat tulokset on saavutettu tavupiirrejärjestelmillä ('syllabic-feature sys-

tem') sekä useita fonin tunnistajia ja fonotaktista ('phonotactic') kielen mallinnusta käyttävillä järjestelmillä. [Zissman & Berkling 2001, 121.] Zissmanin ja Berklingin [mts. 122] mukaan on hyviä syitä uskoa, että järjestelmien suorituskyky paranee otettaessa käyttöön korkeamman tason lingvististä informaatiota hyödyntäviä järjestelmiä, jotka mallintavat foneja, niiden frekvenssejä ja fonotaktisuutta. Kyseisten ominaisuuksien käyttäminen tosin edellyttää kielentunnistusjärjestelmien harjoittamista etukäteen, mikä vie aikaa. [Zissman & Berkling 2001, 122.]

Leaversin ja Burleyn [2001, 639–640] mukaan automaattinen kielen tunnistaminen liittyy puhujasta riippumattomaan puheentunnistukseen ja puhujan tunnistamiseen, joista nimenomaan puhujan tunnistamisen keinot suoriutuvat tällä hetkellä parhaiten. Kyseiset algoritmit perustuvat akustisten piirteiden purkamiseen ja erilaisten hahmontunnistusmenetelmien käyttämiseen. Leaversin ja Burleyn [2001, 640] mukaan tällä hetkellä trendit tutkimuksessa keskittyvät vektorikvantittumisen ('Vector Quantisation'), kätkettyjen Markovin mallien ja neuroverkkojen käyttämiseen. Heidän mukaansa tunnistamistarkkuus ei ole tarpeeksi hyvä monia käytännöllisiä tilanteita varten (vrt. Zissman & Berkling 2001). Lisäksi Leaversin ja Burleyn [2001, 640] mukaan tunnistamismenetelmät tarvitsevat pitkiä näytteitä ja algoritmien ajoajat ovat pitkiä. Heidän mukaansa standardit puheen- ja hahmontunnistusmenetelmät eivät sovi automaattiseen kielen tunnistamiseen, sillä ne eivät ota huomioon kognitiivisia prosessointistrategioita ja lingvistisiä vihjeitä ('cue'), joita ihmiset käyttävät tehokkaasti tunnistaessaan kieliä. [Leavers & Burley 2001, 639–640.]

5 VIDEOTIEDONHAUN KÄYTTÖLIITTYMÄT JA VIDEODATAN VISUALISOINTI

Tiedonhakujärjestelmän käyttöliittymä on alue, jossa käyttäjä ja järjestelmä kohtaavat [Lu 1999, 185]. Käyttöliittymä yhdistää järjestelmän toiminnallisuuden ja käyttäjän järjestelmälle osoittamat vaatimukset [Lee & Smeaton 1999, 1]. Käyttöliittymien suunnittelussa pitäisi ottaa huomioon, että käyttäjät eivät ole ensisijaisesti kiinnostuneita raa'asta videodatasta ja videodokumenteista vaan niiden sisällöstä eli siitä, mitä niillä pyritään viestimään. Toisin sanottuna, vaikka käyttäjä etsisikin tiettyä videonauhaa, jolle on tallennettu tietty uutislähetys, hän ei ole niinkään kiinnostunut tuosta nauhasta itsestään tai edes sen sisällöstä muuten kuin hyvin rajatulta osin; juuri videodokumenttien sisältö pitäisi saada käyttäjien ulottuville, heidän ha-

luamiltaan osin, ei fyysinen videonauha tai binaarimuotoinen videosekvenssi. Tästä syystä hakujärjestelmän pitäisi pystyä esittämään videon sisältöä rakenteisessa ja helposti haettavassa muodossa [Ks. Petković & Jonker 2000]. Käyttöliittymän pitäisi mahdollistaa tehokas ja vaivaton tapa videosisältöjen hakemiseen [Lee & Smeaton 1999, 1–2]. Del Bimbon [1999, 16, 244] mukaan, jotta tämä olisi mahdollista, on tarpeellista tiivistää videoiden informaatiosisältöä helpommin selattavaan muotoon. Myös informaation visualisointivälineet ovat tärkeitä selailun tehokkuuden parantamiseksi. [Del Bimbo 1999, 16, 244.] Tässä luvussa tarkastelu keskittyy käyttöliittymiin ja hakuvälineisiin, jotka sopivat erityisesti TV-uutisten hakuun. Esimerkkinä käyttäjäryhmästä käytetään TV-toimitusta, jonka työkäytäntöjä tarkastelemalla on tehty arvioita heidän tarpeistaan videotiedonhakujärjestelmien suhteen. Olemassa olevia hakuvälineitä arvioidaan kriteereillä, joista lisää myöhemmin.

5.1 Käyttöliittymien periaatteet

Ennen käyttöliittymien hakuvälineiden tarkastelua käsitellään niiden toiminnan kannalta keskeisiä periaatteita: käyttöliittymien tehtäviä, videoinformaation esittämistä uudelleen ja informaation tiivistämistä.

5.1.1 Tehtävät ja datatyypit informaation visualisoinnissa

Shneiderman [1998, 512] käsittelee tiedonhaun käyttöliittymäympäristöön liittyviä elementtejä, joita ovat (1) tehtäväobjektit ('task objects'), joihin lasketaan esimerkiksi videosekvenssit, (2) käyttöliittymäobjektit ('interface objects'), jotka edustavat kyseisiä tehtäväobjekteja, (3) tehtävätoimenpiteet ('task actions'), esimerkiksi faktatietojen etsiminen ja siitä edelleen eriteltävät selailu ja hakeminen, joita edustetaan (4) käyttöliittymän toimenpiteillä ('interface actions'), kuten vierittämisellä, zoomaamisella, läheisyysperiaatteella tai linkittämisellä. Käyttäjät aloittavat tiedonhaun muotoilemalla tiedontarpeensa tehtäväalueella. [Shneiderman 1998, 512.] Tehtävät voidaan Shneidermanin [1998, 512–513] mukaan jakaa rakenteisiin ja rakenteettomiin:

1. Faktatietojen etsiminen ('specific fact finding') eli tunnettujen nimekkeiden hakeminen on rakenteinen tehtävä, jossa on yksi selvästi tunnistettavissa oleva lopputulos, esimerkiksi henkilön ikä.
2. Laajennettu haku on rakenteinen tehtävä, jossa ei ole varmuutta lopputuloksista.

3. Tunnettujen kokoelmien avoin ('open-ended') selailu on rakenteeton tehtävä.
4. Aiheesta saatavilla olevan informaation tutkiminen on rakenteeton tehtävä. [Shneiderman 1998, 512–513; vrt. Antani et al. 2002, 950.]

Selvitettyään tiedontarpeensa käyttäjien on päätettävä mistä etsiä. Tiedontarpeiden selventämiseksi voidaan käyttää apuvälineitä ('finding aids'). Esimerkkinä näistä etsimisen apuvälineistä voidaan ottaa sisällysluettelot ja hakemistot kirjoissa. Tehtäväalueella ilmaistujen tiedontarpeiden muuttaminen käyttöliittymän toimenpiteiksi on suuri kognitiivinen askel, joka on tehtävä ennen hakuprosessissa jatkamista. [Shneiderman 1998, 513.] Kuten edellisissä luvuista on käynyt ilmi, videoiden sisällönkuvailu ei ole sillä tasolla, että käyttäjät voisivat ratkaista useimmat hakuongelmansa rakenteisilla ja analyyttisillä kyselyillä. Jos käyttäjä joutuu semanttisten kyselyiden sijaan miettimään, että millaisia havaittavia piirteitä hänen etsimäänsä sisältöön liittyy, kognitiivinen askel tehtäväalueen tiedontarpeiden ja käyttöliittymätoimenpiteiden välillä on normaaliakin suurempi. Käytännön videotiedonhaku perustuu oleellisesti selailuun, ja kyselyt ovat enemmänkin työkalu aiheen rajaamiseen. Niin kauan, kun automaattinen sisällönkuvailu on nykytilassaan, videotiedonhaun käyttöliittymien suunnittelussa keskeisin ongelma on videosisältöjen visualisointi selaamista varten.

Shneidermanin [1998, 523] mukaan visuaalisen suunnittelun keskeisin periaate voidaan tiivistää periaatteeseen “*yleiskuva ensin, zoomaus ja suodin, sitten yksityiskohdat kysynnän mukaan*”. Leen ja Smeatonin [2002] mukaan Shneidermanin [1998, 523] ohje on hyvä mihin tahansa informaation esittämiseen. Kun tätä periaatetta tarkennetaan, seitsemän mahdollista ja abstraktiotasoltaan korkeaa informaation visualisoinnin tehtävää ovat: yleiskuva, tarkennus, suodin, yksityiskohdat pyydettäessä ('details-on-demand'), läheisyysperiaate, historia ja poimiminen ('extract'). [Shneiderman 1998, 523.] Luettelo etenee yleisimmästä yksityiskohtaisimpaan.

1. *Yleiskuva*: Informaation visualisoinnissa yleiskuvalla ('overview') pyritään antamaan käyttäjälle näkemys kokoelman sisällöstä koko hakualueella. Yleiskuvaan kuuluvia strategioita ovat erilaiset tavat tarkentaa visualisoidun kokoelman yksityiskohtia kohti haettua informaatiota, jolloin yleiskuva muuttuu kuvaksi tarkennuskohdasta ja sen ympäristöstä. [Shneiderman 1998, 535–536.]
2. *Tarkennus***:** Käyttäjä voi tarkentaa hakualueessa toivomiansa nimekkeitä. Tarkennus on kuitenkin toteutettava siten, että käyttäjälle jää käsitys sijainnista hakuavaruudessa. Järjestelmän on tuettava käyttäjän käsitystä haettavan tiedon kontekstista. [Shneiderman 1998, 536–537.]

3. *Suodatus*: Suodatuksessa ('filter') käyttäjälle annetaan keinoja suodattaa pois sisältöä, jota ei haluta sisällytettäväksi hakuun. Suodattaminen voidaan toteuttaa esimerkiksi antamalla tietyntyyppisille dokumenteille eri painoarvoja tai yksinkertaisella visualisoidulla Boolen-logiikalla (eli tässä tapauksessa NOT-operaattorilla). [Shneiderman 1998, 538.]
4. *Yksityiskohtia pyynnöstä* ('details-on-demand'): Käyttäjä saa haluamastaan hakutuloksena saadusta dokumentista lisätietoja, esimerkiksi metatiedot, napauttamalla hiirellä dokumentin otsikkoa tai viemällä hiiren osoittimen sen päälle. [Shneiderman 1998, 538–-539.]
5. *Läheisyysperiaate*: Hakutulosten ominaisuuksien liittymistä ('related') toisiinsa voidaan myös hyödyntää. Aineistoa voidaan rajata haettavaksi esimerkiksi jo tulokseksi saatujen nimekkeiden tekijöiden avulla, jolloin seuraava hakutulos sisältää vain tietyn tekijän aineiston. Läheisyysperiaate sopii erityisen hyvin esimerkiksi elokuvien hakuun, jolloin haku voidaan rajata koskemaan tiettyä näyttelijää tai ohjaajaa. [Shneiderman 1998, 539.]
6. *Historia*: Hyvän hakujärjestelmän pitäisi tukea käyttäjää tallentamalla tietoa suoritetuista toimenpiteistä. Koska haut eivät läheskään aina ole onnistuneita, on tarjottava mahdollisuus palata aiempaan tilanteeseen. Lisäksi järjestelmässä pitäisi olla ominaisuus, joka mahdollistaa hakemisen edellisen haun tuloksilla, jotta koko hakua ei tarvitsisi aina aloittaa alusta. [Shneiderman 1998, 540.]
7. *Tallennus*: Hakutulosten ja -lauseiden poimiminen ('extract'), muiden ominaisuuksien ohella, mahdollistaisivat paremmin käyttäjää tukevien järjestelmien rakentamisen. Käyttäjälle annettaisiin järjestelmässä mahdollisuus palata hakuun myöhemmin, joko sen jatkamiseksi, sen käyttämiseksi avuksi uudessa haussa tai sen tallentamiseksi hakutuloksen myöhempää jatkokäsittelyä varten. [Shneiderman 1998, 540–541.]

Lee ja Smeaton [2002] mainitsevat esimerkkinä perinteiset bibliografisen informaation hakemisen apuvälineet: hakutuloksessa näytetään yhdellä rivillä pitkälle tiivistetty otsikko ja dokumentin päiväys (eli yleiskuva), otsikkoa napsauttamalla järjestelmä näyttää tiivistelmän ruudun alaosassa (eli tarkennus), ja lopulta tarkempi perehtyminen avaa kokotekstidokumentin (eli yksityiskohtia pyynnöstä). [Lee & Smeaton 2002.] Näin ollen tyypillisessä hakuprosessissa informaatiota visualisoitaessa ensin tarjotaan käyttäjälle yleiskuva tarjolla olevasta informaatiosta, minkä jälkeen hän voi tarkentaa näkymää haluamiinsa nimekkeisiin ja suodattaa pois niitä nimekkeitä, joita hän ei halua. Lopulta hän voi keskittyä vain nimekkeiden niihin ominaisuuksiin, joista hän on kiinnostunut. Tähän palataan myöhemmin videoinformaation

tiivistämisen yhteydessä. Mikäli käyttöliittymä muistuttaisi kaksiulotteista kohtauksensiirtymiskaaviota, yleiskuva esittäisi videodokumentin kaikki tarinan yksiköt eli alikaaviot. Käyttäjä voisi sitten tarkentaa näkymää johonkin tiettyyn tarinan yksikköön ja tarkastella sen otosryhmien eli solmujen välistä vuorovaikutusta. Läheisyysperiaatteen avulla voitaisiin tarkastella kaikkia solmuja, joilla on vuorovaikutusta tietyn solmun kanssa. Halutessaan käyttäjä voisi pyytää vielä yksityiskohtia jonkun solmun yksittäisistä otoksista.

Shneidermanin [1998, 523] mukaan tehtäväalueen informaatio-objekteja luonnehditaan informaation visualisoinnin datatyyppitaksonomian datatyypeillä, joille suoritettavia tehtäviä kuten yleiskuvaa ja tarkennusta käsiteltiin edellä. Shneiderman [1998, 524–535] luettelee datatyyppejä, joista käsitellään niitä, jotka liittyvät videoiden visualisointiin.

1. *Yksiulotteinen lineaarinen data*: Lineaarisiin datatyyppeihin kuuluvat peräkkäin järjestetyt tekstuaaliset dokumentit. Jokainen nimeke kokoelmassa on rivi tekstiä, joka muodostuu jonosta merkkejä. Käyttöliittymän ('interface') suunnittelussa on otettava huomioon, mitä kirjasimia, värejä ja kokoja käytetään sekä mitä yleiskuvan, vierittämisen ja valinnan keinoja käytetään. Käyttäjän tehtäviä voivat olla nimekejoukkojen löytäminen tai tietyt ominaisuudet käsittävien nimekkeiden löytäminen. [Shneiderman 1998, 524–-526.]
2. *Kaksiulotteinen karttadata*: Jokainen nimeke kokoelmassa kattaa osan kokonaisessa alueessa. Jokaisella nimekkeellä on tehtävä–alue-ominaisuuksia kuten nimi ja arvo sekä käyttöliittymä-alueen ('interface-domain') piirteitä kuten koko ja väri. Käyttäjätehtäviä ovat vierekkäisten nimekkeiden löytäminen ja seitsemän perustehtävän suorittaminen. [Shneiderman 1998, 526–528.]
3. *Kolmiulotteinen maailma*: Tosimaailman objekteilla ('real-world objects') kuten molekyyleillä ja rakennuksilla on nimekkeitä, joilla on ominaisuutenaan tilavuus ('volume') ja potentiaalisesti monimutkaiset suhteet toisten nimekkeiden kanssa. Käyttäjän tehtäviin kuuluvat läheisyyden ja perustehtävien lisäksi läheisyys ja yläpuolella-alapuolella sekä sisäpuolella-ulkopuolella suhteet. [Shneiderman 1998, 528.]
4. *Ajallinen data*: Aikajanoja ('time line') käytetään laajalti. Ajallisen datan erityispiirteitä ovat nimekkeet, joilla on alku- ja loppuaika, sekä nimekkeiden mahdollinen limittäisyys. Tavallisia tehtäviä ovat kaikki tapahtumat ennen, jälkeen tai jonkin ajanjakson aikana sekä perustehtävät. [Shneiderman 1998, 528.]
5. *Puudata*: Hierarkiat tai puurakenteet ovat kokoelmia nimekkeitä, joissa jokaisella nimekkeellä (paitsi juurella) on linkki yhteen kantanimekkeeseen ('parent item'). Jokaisella linkillä ja nimekkeellä voi olla useampia ominaisuuksia. Perustehtäviä voidaan sovel-

taa nimekkeisiin ja linkkeihin. [Shneiderman 1998, 531–533.]

Esimerkiksi Rui ja muut [1999, 361] käsittelevät mallia, jossa videon rakenne esitetään niin, että kohtaukset näytetään vertikaalisesti ja avainkehykset horisontaalisesti. Kyseessä on kaksiulotteinen ajallinen datatyyppi.

5.1.2 Videoinformaation esittäminen

Käyttöliittymän on mahdollistettava alkuperäisen visuaalisen informaation esittäminen uudelleen tiivistetyssä muodossa tiedonhaun ja halutun informaation tunnistamisen helpottamiseksi [ks. Yeo & Yeung 1997, 49]. Ensisijainen tavoite videotiedonhaussa on mahdollistaa videoinformaation esittäminen mahdollisimman kattavasti mutta mahdollisimman lyhyesti. Erilaisia korvikkeita, joita voidaan kutsua muun muassa representaatioiksi, abstraktioiksi ja tiivistelmiksi, on käytetty tiedonhaun apuna jo kauan, sillä niiden selaaminen on helpompaa kuin alkuperäisten informaatio-objektien. Videokorvikkeet voidaan luokitella staattisiin ('still-image') ja liikkuviin. Staattisia korvikkeita ovat avainkuvat ja mosaiikit, joita voidaan esittää esimerkiksi diaesitysten tavoin. Liikkuvaan kuvaan perustuvia korvikkeita ovat esimerkiksi videotrailerit ja kohokohdat ('skims'). [Ks. Geisler et al. 2001, 68.] Seuraavaksi käsitellään informaation tiivistämisen keinoja, minkä jälkeen tarkastellaan näiden esitystenmuotojen järjestämistä tasoiksi. Visuaalisia selaimia tarkastellaan myöhemmin.

Aiemmin avainkehyksiä käsiteltiin videon paikkasidonnaisten ominaisuuksien tunnistamisen yhteydessä, mutta niitä käytetään myös videosisällön esittämiseksi: Del Bimbon [1999, 244] mukaan eräs tapa visualisoida videosisältöjä on poimia videosta joukko yksittäisiä kuvia, pienentää niitä ja esittää ne ruudulla. Leen ja Smeatonin [2002] mukaan, koska kehysten analysointi on niin keskeisessä osassa videoiden indeksoinnissa, on luonnollista, että avainkehysten näyttäminen on yleinen lähestymistapa videoiden selailemiseksi. Avainkehyksien avulla esitetään ajan etenemistä muuttamalla ajallinen sisältö staattisiksi kuviksi – se, että avainkehykset ovat ajallisesti järjestetty, erottaa ne tavanomaisista peukalonpääkuvista. [Lee & Smeaton 2002.] Avainkehyksiä valittaessa on tärkeää päättää, mitkä kehykset valitaan ja kuinka monta [Lee & Smeaton 2002]. Yksinkertaisimmat menetelmät valitsevat avainkehyksiä videoista säännöllisin väliajoin, mutta kehittyneemmät menetelmät ottavat huomioon myös videosisällön [ks. liite 4]. Leen ja Smeatonin [1999, 14] mukaan älykäs avainkehyksen valitseminen viittaa mihin tahansa sisältöpohjaiseen menetelmään, joka valitsee yhden tai useamman kehyksen videosekvenssistä, jolloin valituilla kehyksillä voidaan edustaa koko

otosta, eräänlaisena videotiivistelmänä. Näiden kehysten tarkoitus on antaa mahdollisimman kattava kuva siitä, mistä kokonaisissa otoksissa on kysymys. [Lee & Smeaton 1999, 14.] Bollen ja muiden [1998] mukaan avainkehyksiä voidaan valita otoksesta useampia kuin yksi, jos otoksessa on paljon liikettä ja toimintaa.

Millä tahansa menetelmällä ne valitaankin (tasaisin väliajoin, yksi tai useampi otosta kohden, otoksen ensimmäinen, keskimmäinen ja viimeinen, tai jollain vielä älykkäämmällä menetelmällä), tuloksena on joukko avainkehyksiä, jotka pitää sellaisessa muodossa, että niiden selaileminen on mahdollisimman helppoa [Lee & Smeaton 2002]. Koska avainkehyksiä valitaan otosten suuren määrän vuoksi paljon – Bollen ja muiden [1998] mukaan tunnissa on muutama sata otosta – visualisointi globaalilla eli kokonaisten videodokumenttien tasolla on vaikeaa. Avainkehykset sopinevat parhaiten kyselyjen ja hakutulosten visualisointiin. [Del Bimbo 1999, 245; Bolle et al. 1998]. Avainkehysten kaltainen esitystapa on leikekartta ('clipmap'), jolla tarkoitetaan näytöllistä mikoneiksi ('micons') kutsuttavia kolmiulotteisia ikoneja, joista jokainen edustaa yhtä otosta ja joiden kolmas ulottuvuus esittää otoksen pituutta. Mikonit sopivat kyselyjen visualisointiin. [Del Bimbo 1999, 247.]

Brunelli ja muut [1999, 97–98] sekä Bolle ja muut [1998] huomauttavat, että kokonaisten otosten esittäminen staattisina kuvina on olut huomattava edistysaskel videoiden esittämisessä. Brunellin ja muiden [1999, 98] mukaan videoiden esittämistä varten kirjallisuudessa on ehdotettu uutta luokkaa kuvia: silmiinpistäviä kehyksiä ('salient stills'). Kyseiset kuvatyypit esittävät videosekvenssissä tapahtuvien ajallisten muutosten kasautumista sekvenssien silmiinpistävien piirteiden avulla. Kaikki videoleikkeen kehyksissä tapahtuvat muutokset yhdistetään kuvankäsittelymenetelmillä yhteen ainoaan kuvaan, josta saatetaan korostaa sen silmiinpistäviä eli muuttuneita ominaisuuksia. [Brunelli et al. 1999, 98; Bolle et al. 1998.] Brunelli ja muut [1999, 98] kirjoittavat, että yleensä yleensä videoissa on paikallaan pysyvä tausta ja mahdollisesti itsenäisesti liikkuvia objekteja; hallitsevat muutokset taustassa johtuvat yleensä liikkeestä ja kameran käytöstä. Kun kameratyö on tunnistettu, valitaan viitekehys ('a reference frame'). Seuraavat kehykset väännetään ('warp') viitekehyksen koordinaattijärjestelmään, mistä syntyvää kuvaa kutsutaan mosaiikiksi [liite 3]. Bolle ja muut [1998] kirjoittavat mosaiikkistamisesta, että otoksessa, jossa ei ole kuin kameran liikettä, voidaan peräkkäisten kehyksien muutos määrittää ja kehykset liittää päällekkäin yhdeksi isommaksi kuvaksi. Jos otoksessa on itsenäistä liikettä, liikkuvat objektit erotetaan taustasta, jolloin taustasta muodostetaan erillinen kuva ja liikkuvista objekteista ja niiden lentoradoista omat kuvansa. [Bolle et al. 1998.] Del Bimbon [1999, 248] mukaan mosaiikkistamisessa koko otos ajallisine muutoksineen esitetään yhdessä panoraamakuvassa, kun kaikki otoksen kehykset asetetaan päällekkäin ajallisten suotimien läpi. Mosaiikkistamisen kanssa samankaltaista menetelmää kutsutaan

videoavaruusikoneiksi ('video space icons') [Bolle et al. 1998; Del Bimbo 1999, 248].

5.1.3 Videoinformaation tiivistäminen

Rui ja muut [1999, 359] kirjoittavat, että selailun ja haun näkökulmasta video on verrattavissa kirjaan: kirjan käyttöä helpottaa sisällysluettelo, joka esittää sen semanttisen rakenteen. Videotiivistelmä ('video abstract') on sarja pysäytyskuvia tai liikkuvia kuvasarjoja, jotka tiivistävät videosisällön niin, että alkuperäisestä videosta ei menetetä mitään keskeistä [Brunelli et al. 1999, 100]. Leen ja Smeatonin [1999, 5] sekä Lienhartin ja muiden [1997, 55] mukaan hyvä tiivistelmä esittää selailun helpottamiseksi alkuperäisen nimekkeen uudessa muodossa yksinkertaisesti ja lyhyesti, mutta silti niin kattavassa muodossa, että alkuperäisen nimekkeen keskeinen sisältö säilyy. On kuitenkin otettava huomioon, että kovin tiivis esitys ei välttämättä välitä riittävästi informaatiota nimekkeistä; jos taas esitys on yksityiskohtainen, se saattaa välittää liian paljon informaatiota, jolloin selailu vie enemmän aikaa ja nimekkeiden välisten suhteiden tutkiminen ei ole niin yksinkertaista. [Lee & Smeaton 1999, 5.] Esimerkiksi katsottaessa videota läpi videonauhuria muistuttavalla käyttöliittymällä, vaikka kuvaa hieman nopeuttaenkin, hakija saa huomattavasti enemmän informaatiota nähtäväkseen kuin jos hänellä olisi selattavanaan vain videodokumentin otosten avainkuvat. Toisaalta perinteinen kuvanauhuria matkiva käyttöliittymä – joka on varsin yleinen myös digitaalisissa videohakujärjestelmissä – on liian kömpelö haluttujen kohtauksien paikallistamiseen [Lee & Smeaton 1999, 1; Rui et al. 1999, 359].

Koska selailussa hakuongelman kannalta ylimääräinen informaatio on haitaksi, kaikki kunnolliset videoiden selailun mahdollistavat käyttöliittymät joutuvat tiivistämään informaatiota. Jotta käyttöliittymä voisi toteuttaa Shneidermanin [1998, 523] periaatteen, ”yleiskuva ensin, zoomaus ja suodin, sitten yksityiskohdat kysynnän mukaan”, on sen mahdollistettava videoiden selaileminen eri tasoilla esitettyihin yksityiskohtiin nähden, mitä kutsutaan videon abstraktioimiseksi eli tiivistämiseksi [Lee & Smeaton 2002]. Lee ja Smeaton [1999, 6–7] huomauttavat, että mitä enemmän vaihtoehtoisia abstraktiomenetelmiä ja -tasoja on käytettävissä, sen parempi, kun käyttäjä voi valita haluamansa esitystavan yleiskuvaa varten ja siirtyä halutessaan tiettyä nimekettä varten saatavilla oleviin yksityiskohtaisempiin esitystapoihin. Käyttöliittymän on siis mahdollistettava siirtyminen tiivistelmän tasolta toiselle (eli tarkennus ja yksityiskohtia pyynnöstä).

Yksinkertaisen eteen- ja taaksepäinkelaamisen ongelmat selailuvälineenä on otettu havaittu myös kirjallisuudessa. Selaaminen on aina ajan rajoittamaa: Toistoa nopeutet-

taessa sisällön tunnistaminen käy vaikeammaksi ja tietyn kohdan paikallistaminen rasittavaksi. Lisäksi ajallisiin rajoituksiin kuuluvat, että käyttäjä voi katsoa vain yhtä kohtaa videosta kerrallaan ja että liikkuminen ajassa eteen- ja taaksepäin on vaivalloista. Alkuperäisestä videodatasta tiivistetyn esityksen selaileminen on vähemmän aikaa kuluttavaa ja ponnisteluja vaativaa kuin alkuperäisen videon katsominen kokonaan läpi. [Lee & Smeaton 2002.] Samat ongelmat, jotka koskevat videokuvaa, koskevat myös äänidokumenttien selaamista: käyttäjä päätyy helposti selailemaan ääniraitaa eteen- ja taaksepäin ilman mitään aavistustakaan kontekstista. Tämä keskeinen ongelma johtuu siitä, että video ja ääni ovat aikapohjaisia medioita. Nykyään, kun tarvittava prosessointiteho ei enää ole utopiaa, haaste on siirtynyt niille, jotka pyrkivät kehittämään ideoita videosisällön selailua varten. Tällä hetkellä on kuitenkin käytössä vain kourallinen ideoita videoiden selaamiseen eri abstraktiotasoilla. [Lee & Smeaton 2002.]

Kuinka sekventiaalinen aikaan pohjautuva videodokumentti saadaan muutettua (siis tiivistettyä) helpommin selattavaan muotoon, on Leen ja Smeatonin [2002] mukaan avoin kysymys, ja uusia ideoita tarvitaan. Sopivaa visualisointikeinoa mietittäessä on päätettävä, käytetäänkö staattisia vaiko dynaamisia esityksiä, millä perusteella avainkehykset tai videoleikkeet valitaan, miten esitykset järjestetään näytölle ja kuinka siirrytään tiivistämisen eri tasojen välillä.

Ruin ja muiden [1999, 360] mukaan yksinkertainen yksiulotteinen sekventiaalinen videosisällön esitys avainkehyksillä on lähes hyödytön avainkehyksien suuresta määrästä johtuen – selattavaa olisi aivan liian paljon. Koska videodata koostuu joukoista kehyksiä ajallisessa järjestyksessä, on luonnollista tiivistää sitä suodattamalla epäolennaisia kehyksiä pois esityksestä [ks. Lee & Smeaton 2002]. Videoselaimen suunnittelussa pitäisi selailun helpottamiseksi ottaa huomioon otoksista poimittujen avainkehyksien suodattaminen ja niiden määrän vähentäminen esimerkiksi (1) ryhmitteleмällä kaikki avainkehykset tietyn ajanjakson sisällä ja valitsemalla niistä joku; vaihtoehtoisesti voidaan (2) käyttää sisältöpohjaista menetelmää, jossa otetaan huomioon esimerkiksi objektien vaihtuminen, tai sitten voidaan (3) ryhmitellä avainkehyksiä semanttisesti. [Lee & Smeaton 2002.] Muutenkin on muistettava, että videoita katsellaan semanttisen sisällön eikä fyysisten otosten tai avainkehyksien takia. Sisällysluettelon tuottaminen videosta kohtausten tasolla on olennainen vaatimus, koska kohtaukset kantavat semanttisen merkityksen. [Rui et al. 1999, 360.]

Videon tiivistämisessä pitäisi Brunellin ja muiden [1999, 101] käsittelemän kirjallisuuden mukaan ottaa huomioon, että (1) toimintaa sisältävät kohtaukset ovat tärkeitä (ja ne pitäisi ottaa mukaan tiivistelmään), (2) ihminen kiinnittää huomiota kontrastiin, (3) värit välittävät tunteita, (4) alkuperäisen videon kohtausten järjestys pitää säilyttää kontekstin

vuoksi ja (5) esitettävien kohtausten pitää olla tarpeeksi pitkiä (eli vähintään 3 sekuntia). Värit kannattaa ottaa huomioon siitäkin syystä, että Markkulan [2002] mukaan toimittajat ovat niistä kiinnostuneita. Leen ja Smeatonin [2002] mukaan eräs idea on näyttää avainkehyksiä aikajanoilla kolmiulotteisessa muodossa. Tavat ilmaista aikaa käsittävät esimerkiksi (1) suoraviivaisen aikajanapalkin, jota suurin osa videontoisto-ohjelmista käyttää, (2) videotiivistelmistä löytyvien laskettavien objektien ilmaisemisen, ja (3) tavan ilmaista aikaa selailtavien yksikköjen (eli korvikkeiden) syvyytenä. [Lee & Smeaton 2002.] Syvyyttä voidaan ilmaista pinoamalla vierekkäisiä kehyksiä päällekkäin ja näyttämällä kehysten reunat päällimmäisen alla [ks. ”mikonit”, Del Bimbo 1999, 247]. Näin on mahdollista nähdä kehysten välinen jatkuvuus näkemättä suurinta osaa niiden sisällöstä. Toinen idea on järjestää avainkehykset yhdelle ruudulle sarjakuvan tavoin. [Brunelli et al. 1999, 101–102.] Brunelli ja muut [1999, 101–102] käsittelevät menetelmiä tiivistelmien automaattiseen muodostamiseen.

Lee ja Smeaton [1999, 5] esittävät, että käyttöliittymän pitäisi tarjota käyttäjälle erilaisia tasoja tiivistetyn informaation tarkastelemiseen ja äärimmäisten abstraktiotasojen välille vaihtoehtoja. Mikä on järkevin abstraktiotaso esitykselle kussakin tilanteessa ja kuinka monta tasoa erilaiset tarpeet vaativat, on avoin kysymys. Tällä hetkellä mahdollisia vaihtoehtoja ei ole kuitenkaan kovinkaan runsaasti. [Lee & Smeaton 1999, 5.] Kerrostuneisuuden ääripäissä ovat tapa, jossa yhdellä avainkehyksellä edustetaan koko videonimekettä, ja tapa, jossa näytetään kaikki videosta poimitut avainkehykset. Näiden kahden ääripään väliin voidaan sijoittaa vaihtoehtoja, joiden tarve loppujen lopuksi riippuu käyttäjästä ja tilanteesta. Tärkeää olisi, että käytettävissä olisi yksityiskohtaisuuksiltaan erilaisia kerroksia ja niiden välillä olisi linkkejä, joiden avulla käyttäjä voisi halutessaan siirtyä yksityiskohtaisempaan tarkasteluun, mutta silti palata ylemmälle tasolle siihen kohtaan josta hän lähtikin. Käyttäjä voisi siis tarkastella samaa kohtaa videossa eri tasoilla. Abstraktiotasot olisi käytännöllistä kuvailla niiden yksityiskohtaisuuden mukaan. Lee ja Smeaton [2002; 1999, 6, 12–13] mainitsevat löytäneensä käsittelemistään järjestelmistä seuraavat menetelmät ja tasot videon tiivistämiseksi, jotka luetellaan tiivistetyimmästä (eli yleiskuvasta) yksityiskohtaisimpaan:

1. *Avainkehyksistä* tuotettavat (a) pienoiskuvat ('thumbnails') ja (b) syntetisoidut pysäytyskuvat. Syntetisoituja kuvia ovat esimerkiksi mosaiikit ja muut avainkehyksistä edelleen tuotetut esitykset.
2. *Kronologisesti tai hierarkkisesti järjestetyt avainkehysluettelot*: (a) Kronologisesti järjestetyt luettelot (ns. “storyboard”) ovat yleinen abstraktiomenetelmä, joka näyttää joukon avainkehyksiä otoksesta, kohtauksesta tai ohjelmasta ajallisesti järjestettynä. (b) Interaktiivisessa hierarkkisessa avainkehysselaimessa näytetään kaikki videon avainke-

hykset hierarkkisesti. (c) Verkkoympäristössä tarjotaan yleensä mahdollisuus eritasoiseen tiivistykseen avainkehysluettelossa, millä autetaan käyttäjää saamaan yleiskuva videosisällöstä ennen tilaa vievän datan lataamista.

3. (a) *Dynaamiset pienoiskuvat* esittävät kaikki avainkehykset yhdestä kohtauksesta ('slot'), ja (b) *ajastetulla avainkehyksien toistolla* voidaan näyttää sekventiaalisesti joukko avainkehyksiä yhdessä paikassa. Tässä menetelmässä näyttö ('screen real estate') voidaan tallentaa ('save'), koska kaikkia avainkehyksiä varten vaaditaan vain yksi neliskulmainen alue. Käyttäjä tosin joutuu katsomaan näitä dynaamisia pienoiskuvia kauemmin nähdäkseen avainkehykset diaesityksen ('slide show') tavoin.
4. *Kohokohtien* ('skims') näyttäminen on ajastetun avainkehysten toiston liikkuva laajennos. Kohokohdat ovat elokuvatrailerin kaltainen tiivistetty versio videosekvenssistä, jossa näytetään videon tärkeimmät kohdat. Kohokohdat tuotetaan alinäytteistämällä ('sub-sampling') tai jollain kehittyneemmällä menetelmällä. Siinä missä avainkehykset ovat yksittäisiä, staattisia kuvia, kohokohdilla tarkoitetaan lyhyitä videoleikkeitä.
5. *Toisto* ('playback') on tavallinen videonauhuria muistuttava väline, joka toistaa videosekvenssejä. [Lee & Smeaton 1999, 6, 12–13.]

Leen ja Smeatonin [2002] mukaan kirjallisuudessa käsitellään enemmänkin videon tiivistämisen tasoja ja esimerkkinä mainitaan esimerkiksi *Informedia Digital Video Library Systems*, jonka *käyttämillä* tasoilla

1. otsikko esittää videodokumentin nimen tekstuaalisessa muodossa.
2. ”julistekehys” ('poster frame') on videosekvenssistä otettu yksittäinen kehys
3. videoleike on sarja kehyksiä, jotka on otettu videodokumentin sisällöstä
4. kohokohdat ovat merkittäviä videosekvenssien palasia, jotka on otettu alkuperäisestä videodokumentista.

Lee ja Smeaton [2002] antavat esimerkkejä myös muista videon selailujärjestelmistä, jotka hyödyntävät Shneidermanin [1998] visuaalisen suunnittelun keskeisintä periaatetta, jota käsiteltiin aiemmin.

5.2 Käyttöliittymä tiedonhaun eri vaiheissa

Sekä Bolle ja muut [1998] että Lee ja Smeaton [2002] mallintavat kyselyn sarjana vaiheita,

joista jokainen suodattaa informaatiota. Leen ja Smeatonin [2002] mukaan videotiedonhakujärjestelmien käyttöliittymistä ajatteleminen on selkeämpää, mikäli käyttöliittymä jaetaan elementteihin, jotka tukevat käyttäjän toimia tiedonhaun eri vaiheissa. Tätä varten pitää tunnistaa eri vaiheet ('stage') käyttäjän tiedonhakukäyttäytymisessä alkaen päätöksestä käyttää tiettyä tiedonlähdettä, dokumentin etsimisestä valitusta kokoelmasta, dokumentin tietyn osan etsimisestä, kyseisen kohdan katsomisesta, mahdollisesta hakuun palaamisesta ja niin edelleen. On selkeää ajatella hakuprosessia yksittäisinä tasoina, sillä jokainen tiedonhaun taso vaatii käyttöliittymältä erilaisia tukiominaisuuksia. [Lee & Smeaton 2002.]

Geislerin ja muiden [2001, 68–69] mukaan, aivan kuten www-ympäristö on osallaan osoittanut, myös videotiedonhaussa tarvitaan kyselyjen lisäksi selailua. Selailun tarpeeseen vaikuttaa se, että kyselyt ovat sumeita: käyttäjät saattavat tietää, mitä he etsivät, mutta eivät pysty ilmaisemaan sitä selvästi tai sitten heidän tiedontarpeensa eivät ole selvästi määriteltyjä [Lu 1999, 185–186]. Bolle ja muut [1998] sekä Lee ja Smeaton [2002] käsittelevät videotiedonhaun vaiheita ja niitä tukevia käyttöliittymiä, mitkä voidaan jaotella karkeasti seuraavalla tavalla kyselyyn ja selailuun alivaiheineen.

5.2.1 Kyselyt

Bollen ja muiden [1998] mukaan videotiedonhaun avainkysymyksiä on, kuinka kyselyssä otetaan huomioon videon moninaiset informaation modaliteetit. Tärkeintä on, että kysely on helppo muotoilla, käyttöliittymä avustaa kyselyn muodostamisessa ja että kyselyn tulokset voidaan esittää järkevässä, organisoidussa muodossa. Kyselyn suorittamisen pitäisi olla vielä nopeaa. [Bolle et al. 1998.] Kyselyt voivat yleensä perustua (1) sisältöön eli havaittaviin piirteisiin, joita haetaan visuaalisia piirteitä koskevien parametrien avulla, tai ne voivat perustua auditiivisiin piirteisiin ja olla haettavissa avainsanojen avulla; kyselyt voivat perustua myös (2) esimerkkikuviin tai ne voivat olla (3) paikkasidonnaisia kyselyitä, joissa määritellään haluttujen objektien sijainteja kuvissa. [Lu 1999, 186; Prabhakaran 1997, 18.]

Bollen ja muiden [1998] mukaan *navigointivaiheessa* käyttäjä päättää, minkä tyyppisistä videoista hän on kiinnostunut. Tässä vaiheessa kysely kohdistetaan metadataan, esimerkiksi johonkin ajanjaksoon, aiheeseen tai tietyn tyyppiseen videoaineistoon. Viimeksi mainittua huomiota on Bollen ja muiden [1998] mukaan käsitelty kirjallisuudessa: videoita on luokiteltu niiden sovellustarkoituksien perusteella viihteeseen, informaatioon, viestintään ja datan analyysiin. Koska tässä tutkielmassa keskitytään TV-uutisiin, voidaan navigointivaiheen aineistonrajaamista pitää esimerkiksi haun rajaamisena tiettyyn ajanjaksoon. Myös Lee

ja Smeaton [2002] huomauttavat, että laajoja videotietokantoja käytetään yleensä kohdistamalla kyselyitä metadataan, kuten videon otsikkoon ('title'), päiväykseen tai kuvailuun. Laajat videotietokannat tarvitsevat siis myös liittymän tekstuaalisia kyselyitä varten. [Lee & Smeaton 2002.]

Seuraavaksi siirrytään keskeiseen *hakuvaiheeseen* ('searching'). Haun tuloksena on ihanteellisesti mahdollisimman lyhyt mutta kattava luettelo relevanteista nimekkeistä, millä tarkoitetaan, että vain kaikki relevantit dokumentit löydetään. Vaikka navigointivaiheessa haettavia nimekkeitä onkin jo ehditty suodattaa, hakuvaiheessa on usein välttämätöntä suodattaa niitä vielä lisää. [Bolle et al. 1998.] Myös Lee ja Smeaton [1999, 3] sanovat, että kyselyjen avulla hakuavaruutta voidaan suodattaa jättämällä pois nimekkeitä ('items'), joita ei varmasti tarvita selattavaksi. Tässä vaiheessa suodattaminen voidaan tehdä esimerkiksi kohdistamalla kysely tietyntyyppisiin objekteihin, jotka on automaattisen semanttisen päättelyn avulla indeksoitu.

Lee ja Smeaton [2002] huomauttavat, että jos kysely tehdään suoraan videon sisällön ominaisuuksien perusteella, tarvitaan erikoistuneempia käyttöliittymiä kuin tekstitiedonhaussa. *Avainkuvapohjainen hahmonpiirtäminen* ('sketch-drawing') tarkoittaa sisältöpohjaista kyselyvälinettä, jossa käyttäjä määrittelee staattiset visuaaliset piirteet (esim. väriprosentin, tekstuurin, muodon yms.) piirtotyökalussa ja järjestelmä täsmäyttää ne tietokannassa [Lee & Smeaton 1999, 11]. *Histogramminkäsittely* on tekniikka, jossa käyttäjä voi muutella avainkehyksen visuaalisten piirteiden histogrammia ja sitten pyytää järjestelmältä avainkehyksiä, joilla on samanlainen histogrammi. Tekniikka mahdollistaa avainkehyksen matalan tason visuaalisten piirteiden määrittämisen ja hakemisen hyvin tarkasti. Koska histogramminkäsittely ja hahmojen piirtäminen voivat olla aluksi vaikeita oppia, kyselyn muotoileminen on usein aloitteleville käyttäjille hakuprosessin vaikein osa. Niiden sijaan voidaan käyttää *avainkuvapohjaista kysely esimerkillä* -menetelmää, jossa käyttäjä valitsee avainkehysten joukosta kehyksiä, joiden kaltaisia hän haluaa lisää. [Lee & Smeaton 1999, 11.] *Liikepohjainen hahmonpiirtäminen* laajentaa avainkuvapohjaisen hahmonpiirtämisen koskemaan myös objektien ja kameran liikettä objektien staattisten ominaisuuksien lisäksi. Tämä liikekyselyt mahdollistava väline perustuu idealle, että hyvän hakujärjestelmän pitäisi tarjota median ominaisuuksiin perustuvia hakuvälineitä. [Lee & Smeaton 1999, 11–12.] Toisaalta myös liikeominaisuuksien hahmotteleminen voi olla vaikeaa käyttäjälle. Tätä ongelmaa voidaan kiertää mahdollistamalla valmiiden kyselyjen tallentaminen ja käyttämällä *liikepohjaista kysely esimerkillä* -menetelmää, josta esimerkkinä MovEase. [Lee & Smeaton 1999, 12.]

Hakujärjestelmistä esimerkiksi *VideoQ*, *NeTra-V* ja *MovEas*e mahdollistavat visuaaliset ja liikepohjaiset kyselyt. [Lee & Smeaton 2002.] *QBIC* ('Query-By-Image-Content')

suodattaa ensin tekstuaalisella kyselyllä ei-toivottuja nimekkeitä pois ja sitten antaa mahdollisuuden selata tulosjoukkoa. [Lee & Smeaton 1999, 3–4]. VideoQ hyväksyy kyselynä haettavan objektin tai kameran liikkeen lentoradan ja siten mahdollistaa dokumenttien hakemisen niissä esiintyvien objektien tai niiden liikkeen ominaisuuksien perusteella. Vastaavanlainen järjestelmä on MovEase, jossa videon objekteja edustetaan ikoneilla. NeTra-V mahdollistaa myös liikekyselyn, mutta siinä käyttäjä valitsee esimerkkinä annetuista videoleikkeistä (ns. “Query-By-Example”-järjestelmä) sellaisen, jossa esiintyvä liike vastaa hänen etsimäänsä. [Lee & Smeaton 1999, 4.] Lee ja Smeaton [1999, 4–5] epäilevät kuitenkin edellä mainittujen järjestelmien riittävyyttä koko videojärjestelmän ainoaksi hakumenetelmäksi, vaikka heidän mukaansa ne saattavatkin tarjota hyvän aloituskohdan ja täydentää myöhemmin käytettäviä selailuvälineitä.

5.2.2 Selailu

Leen ja Smeatonin [2002] mukaan videoiden selaamista suoraan, ilman hakuavaruuden suodattamista kyselyillä, voidaan verrata tekstitiedonhaun käyttöliittymien kokotekstien tai tiivistelmien selaamiseen. Joka tapauksessa, kuten Bolle ja muut [1998] kirjoittavat, hakuvaiheen tuloksena voi olla varsin pitkä joukko nimekkeitä, joten oikeanlaisten etsiminen vaatii tulosjoukon selaamista. Yeon ja Yeuning [1997, 49] mukaan selailun tavoitteena on välttää sekventiaalinen ja aikaa vievä videoiden katsominen kokonaisuudessaan. Tätä varten vaaditaan välineitä, joilla voidaan navigoida videoissa otoksesta ja tarinasta seuraavaan. [Yeo & Yeung 1997, 49.] Leen ja Smeatonin [1999, 3] mukaan selailun merkitys hakustrategiana korostuu kuva- ja videotiedonhaussa, joissa eräänä ongelmista on indeksoinnin edustavuus indeksoitavaan dataan nähden – olipa kysymyksessä manuaalinen tai automaattinen indeksointi. Järjestelmien, jotka mahdollistavat videosekvenssien käyttämisen kyselyinä, ei tarvitse muodostaa mitään erityisempää tulkintaa datasta, vaan pelkkä piirteiden poimiminen kyselystä ja niiden täsmäyttäminen dokumenttien indeksiin riittää; jos ei haluta tuottaa tulkintoja videodatasta automaattisesti, visuaaliset ja liikepohjaiset kyselyt ovat hyödyllisiä. Selailun merkitystä korostavat visuaalisen tiedontarpeen ilmaisemisen ja määrittelyn ongelmat [Lu 1999, 184].

Lee ja Smeaton [2002] jakavat käyttöliittymien suunnittelun kannalta käyttöliittymät kolmeen ulottuvuuteen: (1) kerrostuneisuus ('layeredness') käsittelee avainkehysten määrää eli sitä näytetäänkö kaikki saatavilla olevat avainkehykset suoraan käyttäjille vaiko valikoidusti vain osa; (2) ajallisen suuntautumisen ('orientation') tarjonta ('provision') tarkoittaa sitä, kuinka aikainformaatio saadaan näkyville käyttäjää varten videosta; (3) avainkehys-

ten paikkasidonnainen tai ajallinen esitys käsittelee sitä, missä muodossa avainkehykset on järjestetty näytöllä.

Selailuvaihe aloitetaan korkean tason yhteenvedosta, joka käsittää selattavat videonimekkeet eli -dokumentit. Korkean tason yhteenvedon eli tiivistelmän avulla käyttäjä saa jonkinlaisen käsityksen kunkin haettavan videonimekkeen sisällöstä ja voi selata lyhyessä ajassa suuren määrän nimekkeitä. [Bolle et al. 1998.] Korkean tason yhteenveto voi muodostua esimerkiksi kyselyn tulosjoukosta. Visuaaliset yhteenvedot ovat Bollen ja muiden [1998] mukaan tärkeä osa hakuprosessia. Yhteenvedot voivat muodostua avainkehyksistä, mutta aivan yhtä hyvin myös esimerkiksi mosaiikeista. Lisäksi käyttäjä saa jo tässä vaiheessa käsityksen siitä, että onko hän kysynyt oikeat kysymykset edellisissä vaiheissa (ks. luku 5.2.3 kyselyn uudelleenmuotoilusta). Yeo ja Yeung [1997, 49] kutsuvat mikroskooppiseksi selailuksi selailua yksittäisten kehysten ja otosten tasolla; makroskooppinen selailu tapahtuu kohtausten ja tarinan tasolla. Selailuvaiheen käyttöliittymä voi esimerkiksi alkaa näytöstä, jolle on sovitettu mahdollisimman suuri määrä nimekkeitä kerralla. Kun nimekkeet esitetään mahdollisimman tiiviissä muodossa näytöllä, käyttäjä voi tutustua suureen joukkoon nimekkeitä kerralla ja siirtyä sitten halutessaan niiden yksityiskohtiin.

Leen ja Smeatonin [2002] mukaan on tärkeää tarjota käyttäjälle selailtaessa käsitys ajasta, koska alkuperäinen video on aikapohjainen ja sekventiaalinen. Useimmissa videotiedonhaun käyttöliittymissä ajallisuus tuodaan esille sekventiaalisen selaimen avulla, joka toimii kuten videonauhurin eteen- ja taaksepäin kelaaminen: selaimen aikajana osoittaa, missä kohden videota ollaan [liite 5]. Perinteisessä avainkehysten selailuideassa avainkehykset esitetään paikkasidonnaisesti, joukkona ruudulla. Ne voidaan kuitenkin esittää ruudulla yksitellen, diaesityksen tavoin. Avainkehysten esittäminen ajallisesti on järkevä vaihtoehto, vaikkakaan joukko avainkehyksiä ei yleensä välitä ajankulua muuten kuin karkeana muutoksena avainkehyksestä toiseen, mistä syystä monet käyttöliittymät kuvaavat ajan kulumista aikajanapalkilla. Lee ja Smeaton [2002] viittaavat tutkimukseen, jossa on havaittu, että paikkasidonnainen esitys toimii paremmin tiettyjen objektien tunnistamisessa ja paikallistamisessa videosta, ja ajallinen esitys taas tuo paremmin esiin videon olennaisimmat ('gist') asiat. [Lee & Smeaton 2002.]

Lee ja Smeaton [2002] tuovat esille erilaisia konkreettisia keinoja esittää kyselyjen tuloksia erityisesti avainkehyksien avulla. Nämä keinot voidaan esittää karkeasti seuraavalla tavalla:

1. *Aikajanaselaimeksi* kutsutussa menetelmässä, jota voidaan kutsua myös “tarinatauluksi” ('storyboard'), avainkehysluetteloksi ('keyframe list') tai filmileikkeeksi ('filmstrip'),

ideana on esittää ruudulla joukko miniatyrisoituja avainkehyksiä paikkasidonnaisesti ja kronologisessa järjestyksessä. Leen ja Smeatonin [2002] sovelluksessa on kaksi tasoa: (a) yleiskuvataso, jolla näytetään kolmekymmentäkaksi koko videosta valittua avainkehystä ja (b) yksityiskohtaisempi taso, jolla kaikki otosten avainkehykset ovat järjestettynä aikajanan mukaan.

2. *Diaesitys* ('slide show') näyttää kuvat yksitellen ajallisesti, ja avainkehyksen alapuolella oleva aikajana osoittaa avainkehyksen aseman suhteessa koko videon pituuteen.
3. *Dynaaminen yleiskuvaselain* on yhdistelmä paikkasidonnaista ja ajallista selainta. Siinä esitetään ruudulla joukko avainkehyksiä, jotka antavat yleiskuvan videon sisällöstä. Mikäli hiiren osoitin viedään jonkin kehyksen päälle, käyttöliittymä esittää ajallisesti yksityiskohtaisempia avainkehyksiä videon kyseisestä osasta. Aikajana näkyy osoitetun avainkehyksen alapuolella ja näyttää ajallisesti sen hetkisen kohdan selailussa.
4. *Hierarkkisesti järjestetty selain* näyttää yksityiskohtien eri tasoja hierarkkisesti. Käyttäjä selaa avainkehyksiä ”porautumalla yksityiskohtiin” ('in a drill-down manner'), mikä sopii erittäin hyvin uutisten kaltaiseen rakenteiseen videoon. [Lee & Smeaton 2002.]

Lee ja Smeaton [2002] huomauttavat, että eri käyttäjät ja tehtävät vaativat erilaisia käyttöliittymiä; yksi käyttöliittymä ei sovi kaikkiin käyttötarkoituksiin. Leen ja Smeatonin [2002] käyttäjätestien perusteella dynaaminen yleiskuvaselain oli pidetyin käyttöliittymä, vaikka käyttäjät toivoivatkin yleensä valinnan varaa selaimiin: jotkin käyttäjät inhosivat niitä ominaisuuksia, joista toiset pitivät. Esimerkkinä Lee ja Smeaton [2002] antavat Físchlár-järjestelmän, joka järjestelee videonimekkeitä eri hakemistoihin, jotta niitä voidaan helposti järjestää, suodattaa ja selailla hieman samaan tapaan kuin esimerkiksi sähköpostia sähköpostiohjelmissa. Minkä tahansa avainkehyksen napsauttaminen hiirellä avaa uuden ikkunan ja toistaa videon kyseisestä kohdasta eteenpäin. [Lee & Smeaton 2002.] Geisler ja muut [2001, 68–71] käsittelevät selainta, joka tukee muun muassa yleiskuvia kokoelmasta, näyttää hakutulokset videodokumenttien otsikkojen avulla, tukee siirtymistä selailussa kokoelman dokumenttien tasolta segmenttien tasolle ja mahdollistaa nopean esikatselun (avainkehyksillä) sekä tukee myös hakuhistoriaa eli mahdollisuutta palata takaisinpäin suoritetuissa toimenpiteissä.

Del Bimbon [1999, 249] esittelee selaimen, joka muistuttaa paljon kohtauksensiirtymiskaaviota. Selain mahdollistaa navigoinnin videoissa niiden ajallisen rakenteen perusteella [liite 6]. Menetelmässä videoiden rakennetta esitetään käsitteellisellä kaaviolla ('schema'). Kaavion avulla videosisältö on käytettävissä eri yksityiskohtatasoilla ja eri näkökulmista, ja sitä voidaan selata myös ei-sekventiaalisesti. Kaavion graafisen esityksen ('graph') avulla käyttäjä voi määritellä näkökulmia, jotka rajoittavat hänen keskittymistään

('focus') videoon – rajoittaminen tarkoittaa suodattamista. Näkökulma voidaan kohdistaa olennaiseen informaation alijoukkoon, jota edustetaan alikaaviolla ja niitä vastaavilla luokan ilmentymillä ('instance'). Rajoittamalla ilmentymien ominaisuuksien arvoja on mahdollista keskittyä vielä näiden ilmentymien alijoukkoihin. Seuraamalla luokkien ilmentymien joukossa olevia linkkejä on mahdollista navigoida videossa. [Del Bimbo 1999, 249.] Kohtauksensiirtymiskaavion tavoin selain esittää videon hierarkkiset ja ajalliset suhteet sekä mahdollistaa tarinan yksiköiden ja niiden välisen ajallisen virran ja siirtymien esittämisen [ks. Yeo & Yeung 1997, 50].

5.2.3 Videosisällön katsominen ja kyselyn uudelleenmuotoilu

Toisto on yleensä viimeinen vaihe käyttäjän vuorovaikutuksessa videohakujärjestelmän kanssa. Sitä varten on olemassa jo useita erilaisia sovelluksia, joita toimitetaan jo tietokoneiden käyttöjärjestelmienkin mukana [ks. liite 5]. Videoiden toistamiseen tarkoitettujen välineiden käyttöliittymät ovat hyvin samanlaisia videonauhurista peräisin olevine toimintoineen ja aikajanaa osoittavine palkkeineen. [Lee & Smeaton 2002.]

Koska videokyselyt ovat sumeita, käyttöliittymän pitäisi tukea kyselyn uudelleenmuotoilua. Uudelleenmuotoilu voidaan toteuttaa niin, että käyttäjän löytäessä nimekkeen, joka vastaa jollakin tavalla hänen tiedontarpeeseensa, nimekkeen olennaiset piirteet voidaan liittää uuteen kyselyyn. [Lu 1999, 187.] Myös Lee ja Smeaton [2002] muistuttavat, että olipa haettava nimeke yksittäinen kohtaus, videoleike tai kokonainen ohjelma, kyselyn tulosten visualisointi tiivistettynä esityksenä tai alkuperäisen kuvasekvenssin toistona on hyvä tapa informoida hakijaa löydetyn nimekkeen ominaisuuksista; tämän jälkeen hakijalle voidaan antaa vihjeitä kyselyn uudelleen muotoilemiseksi tai hakija voi siirtyä tarkastelemaan lähemmin jotain löydetyistä nimekkeistä. Leen ja Smeatonin [2002] mukaan *kyselyn uudelleenmuotoilusta* ('re-querying') on tullut tärkeä elementti käyttäjän ja järjestelmän välisessä vuorovaikutuksessa, kun käyttäjä muotoilee uudelleen tavoitteitaan ja tiedontarvettaan. Lee ja Smeaton [2002] mainitsevat järjestelmiä, joissa on hyödynnetty kyselyn uudelleenmuotoilua.

5.2.4 Puhedokumenttien visualisointi ja selaaminen

Erilaisten esitystapojen yhdistäminen, esimerkiksi puheen transkription käyttäminen yhdessä visuaalisten esitysten kanssa, helpottaa halutunkaltaisen informaation paikallistamista selailemalla. Multimodaalinen visuaalinen selailu voidaan toteuttaa niin, että transkriptiosta tunnis-

tettujen avainsanojen avulla paikannetaan haluttuja kohtia videossa, jonka jälkeen siirrytään kuvaraidan visuaaliseen selailuun. [Lee & Smeaton 1999, 6.] Lee ja Smeaton [1999, 13–14] mainitsevatkin esimerkkejä synkronoiduista videoabstraktioista, joissa on yhdistetty esimerkiksi avainkuvaluettelo ja toisto tai transkriptio ja toisto. Tämänkaltainen järjestelmä sopisi TV-uutisia varten erittäin hyvin, sillä juuri puhe välittää suurimman osan uutisten informaatiosta. Seuraavaksi käsitellään tarkemmin puhedokumenttien visualisointia.

Tekstitiedonhaun tutkimus on keskittynyt keinoihin relevanttien dokumenttien löytämiseksi käyttäjälle, mutta haetun informaation paikallistamiseen löydetyn dokumentin sisällä ei ole keskitytty läheskään yhtä huolellisesti. Tekstiä vastaavalla tavalla puhe on luonteeltaan sarjamainen ja sekventiaalinen media, joten pelkkä puhedokumenttien paikallistaminen ei riitä, koska dokumentit saattavat olla hyvinkin pitkiä. Whittaker ja muut [1999] käsittelevät puhedokumenttien visualisointiin tarkoitettua *SCAN*-järjestelmää, joka tukee paikallista navigaatiota puhedokumenttien sisällä halutun informaation paikallistamiseksi. [Whittaker et al. 1999, 26.] Järjestelmä jäsentää puheen paratoneihin ('paratone') eli “äänikappaleisiin” ('audio paragraph') akustista informaatiota käyttämällä. Jokaisesta paratonista tuotetaan puheentunnistusmenetelmällä tekstuaalinen transkriptio, joka sitten indeksoidaan. Hakutulokset esitetään luettelona puhedokumentteja muun informaation kanssa. SCAN-järjestelmän hakukomponentti mahdollistaa siis tekstimuotoisten puhedokumenttien käyttämisen. [Whittaker et al. 1999, 27.]

Järjestelmä visualisoi yleiskuvaa puhedokumentteihin näyttämällä värikoodien avulla, mitkä hakutermit esiintyvät missäkin dokumentin paratonissa. Jokaista paratonia esittää vertikaalinen palkki histogrammissa: mitä korkeampi palkki on, sitä suurempi osa hakutermeistä esiintyy paratonissa; mitä leveämpi palkki on, sitä ajallisesti pidempi paratoni on kyseessä. Käyttäjät voivat toistaa puhedokumentin napauttamalla hiirellä paratonia vastaavan histogrammin saraketta. Käyttöliittymä mahdollistaa myös dokumenttien vertailun globaalilla tasolla yleiskuvia vertaamalla. [Whittaker et al. 1999, 28.] SCAN-järjestelmän kolmas komponentti muodostuu hakukomponentin käyttämistä tekstuaalisista, jo aiemmin mainituista, paratonien transkriptioista. Hakutermit on korostettu transkriptiossakin samoilla väreillä kuin yleiskuvassakin. Transkription värikoodit mahdollistavat nopean selailun. [Whittaker et al. 1999, 28–29.] SCAN-järjestelmässä on myös oma komponentti puhedokumenttien toistamista varten ja siinä nauhurin ominaisuudet. Periaatteena edellä käsitellyssä lähestymistavassa on siis tarjota puheelle visuaalinen vastike ('analogue') muotoillun tekstin avulla [Whittaker et al. 1999, 29].

On kuitenkin otettava huomioon, että automaattisesti tuotetut transkriptiot sisältävät virheitä: sanoja puuttuu, niitä on liikaa tai tunnistamattomat sanat ovat korvautuneet toi-

silla. Virheetön automaattinen puheentunnistus on vaikeaa uutislähetyksissä, sillä ne sisältävät spontaania puhetta, odottamattomia äänitystilanteita ja paljon sanaston ulkopuolisia sanoja jatkuvasti vaihtuvan sisällön takia. Transkriptioiden virheiden ja intonaation puuttumisen takia alkuperäisiä puhedokumentteja ja niiden toistamista tarvitaan edelleen. [Whittaker et al. 1999, 29.] Whittaker ja muut [1999, 31] mainitsevat, että heidän kokeessaan maksimissaan tunnistettiin 88 % sanoista, minimissään 35 % ja keskimäärin 67 %. Lopulta Whittaker ja muut [1999, 32] huomauttavat, että SCAN sopii vain osaan tiedonhakutehtävistä, kuten faktojen etsimiseen ja relevanssin määrittämiseen: käyttäjätutkimus osoitti, että yhteenvetotehtävä ('summary task') vaati koko dokumentin tutkimista, sillä ilman kunnollista transkriptiota oli vaikea määritellä, mikä oli tärkeää informaatiota. [Whittaker et al. 1999, 32.]

5.3 Videotiedonhaun välineiden arviointi

Kirjallisuudessa käyttöliittymille asetetaan erilaisia vaatimuksia. Leen ja Smeatonin [1999, 1–2] mukaan käyttöliittymä on se osa järjestelmästä, jonka kanssa käyttäjät kommunikoivat. Käyttöliittymäsuunnittelussa on erityisen tärkeää tuntea käyttäjät ja heidän tarpeidensa [mts. 1–2]. Eri käyttäjäryhmät tarvitsevat erilaisia ominaisuuksia ja toimintoja, joiden perusteella järjestelmä pitäisi valita tai suunnitella [Lee & Smeaton 1999, 17; ks. Shneiderman 1998, 509–510]. Bolle ja muut [1998] mainitsevat käyttäjien erilaisia tarpeita: käyttäjät haluavat nähdä uudestaan ennen näkemänsä videon, he saattavat olla kiinnostuneita jostain videosta, jota he eivät ole koskaan nähneet tai heillä saattaa olla vain epämääräinen idea siitä, mitä he etsivät. [Bolle et al. 1998.]

Leen ja Smeatonin [1999, 1–2] mukaan kattavia tutkimuksia käyttöliittymäsuunnittelusta, koskien erityisesti videoiden sisältöpohjaisia hakujärjestelmiä, ei ole vielä eikä edes teoreettista pohjaa niiden tekemiseksi. Digitaalisen videon alueella ei ole tehty systemaattista tutkimusta edes käyttäjistä, heidän tarpeistaan eikä heidän käyttäytymisestään informaation etsinnässä. [Lee & Smeaton 1999, 1–2.] Ylipäätään käyttöliittymät eivät saa tietojärjestelmien suunnittelussa läheskään yhtä paljon huomiota kuin järjestelmien muut osat sen enempää tutkimuksessa kuin kaupallisissa sovelluksissakaan [Lee & Smeaton 2002]. Leen ja Smeatonin [2002] mukaan kaikki käyttöliittymän elementit vaativat enemmän tutkimusta kuin mitä tähän asti on tehty.

Tässä luvussa arvioidaan sitä, kuinka kirjallisuudessa esiteltyjen hakuvälineiden ominaisuudet vastaavat Markkulan [2002] tutkiman TV-toimituksen käytäntöihin ja tiedontarpeisiin. Vaikka nykyisten käytäntöjen ja niihin liittyvien ongelmien perusteella ei voida suun-

nitella uusia järjestelmiä, niihin liittyy seikkoja, jotka on kuitenkin otettava huomioon uusia järjestelmiä suunniteltaessa: jos hakijat eivät aina edes jaksa opetella hakuvälineiden käyttöä kunnolla, on epätodennäköistä, että he haluaisivat muuttaa kaikkia työskentelytapojaan hetkessä. Nykyiset käytännöt antavat käsityksen ympäristöstä, jossa potentiaalinen käyttäjäryhmä työskentelee, ja niiden pohjalta voidaan tehdä huomioita järjestelmiltä vaadittavista ominaisuuksista. Arvioinnin kohteena eivät ole kokonaiset hakujärjestelmät, joiden arvioinnissa pitäisi ottaa huomioon muun muassa käyttöliittymät sekä niiden käytettävyys ja sopivuus mainitulle käyttäjäryhmälle; lisäksi pitäisi huomioida näiden hakujärjestelmien tukemat indeksointimenetelmät ja niiden sopivuus toimituksen tarpeisiin. Asiantuntija-arviointi suoritetaan hakuvälineistä annettujen tietojen perusteella ja Markkulan [2002] tutkiman TV-toimituksen käytännöistä päättelemällä.

5.3.1 Käyttäjät ja vaaditut ominaisuudet

Markkulan [2002] mukaan TV-toimituksen informaatiokäyttäytyminen on kaksijakoista. Dokumentoijat tekevät laajoja kyselyitä ja käyttävät hakuihin paljon aikaa; he suosivat selaamista. Ajankohtaisten asioiden ('current affairs') toimittajat jättävät mielellään kiireidensä takia haut ja materiaalin valitsemisen apulaisilleen. Heidän tekemänsä kyselyt ovat niin yksityiskohtaisia kuin mahdollista – tulosjoukko halutaan pieneksi ja nopeasti selattavaksi. Eräs huomionarvoinen piirre on se, että käyttäjät (informaatikkoja lukuun ottamatta) eivät tunne hakujärjestelmää, jota he käyttävät satunnaisesti ja jota varten heille ei ole annettu koulutusta. [Markkula 2002.] Yleisellä tasolla tarkasteltaessa Markkulan [2002] esiraportista käy ilmi, että YLE:n TV-toimittajien tiedontarpeet jakautuvat aihetta ja taustaa koskeviin sekä visuaalisia piirteitä koskeviin:

- *Aiheet ja taustatutkimus*: Toimittajat käyttävät videoarkistoja muun muassa etsiäkseen ideoita, kehitelläkseen aiheita ja niihin sopivia näkökulmia. Ensimmäisessä vaiheessa videoarkistoja käytetään taustatutkimuksen tekemiseen ja sen selvittämiseen, mitä on tehty aiemmin ja mitä materiaalia on käytettävissä.

Visuaalisia piirteitä koskevat tiedontarpeet ovat konkreettisemmin määriteltävissä kuin aihetta ja taustatutkimusta koskevat:

- *Esteettiset tarkoitukset*: Toimittajat käyttävät videoarkistoja myös sopivan kuvamate-

riaalin (eli objektien, paikkojen, toimien, tunnelmien, kamerakulmien ja liikkeiden) sekä äänimateriaalin (eli musiikin, murteiden, lausuntojen, asenteiden ja tiettyjen sanontojen) etsimiseen.

Aiheeseen ja taustatutkimukseen liittyen toimituksessa ollaan kiinnostuneita videoista seuraavien ominaisuuksien kannalta, jotka antavat yleiskuvan käytettävissä olevasta materiaalista:

1. Tekijänoikeudet eli käyttöoikeus videomateriaaliin
2. Tallenteen formaatti
3. Tekninen laatu
4. Videonauhan fyysinen sijainti (irrelevantti digitaalisissa videoarkistoissa)
5. Potentiaalisen leikkeen pituus (pitäisi olla yli 10 sekuntia)
6. Kuvausvuosi
7. Värit videossa (kuvattu väreissä tai mustavalkoisena)
8. Ääniraidan olemassaolo
9. Ohjelman tyyppi
10. Käsiteltävä aihe (perustuen materiaalin sisällön- ja aiheenkuvailuun). [Markkula 2002.]

Indeksointijärjestelmän kannalta nämä ominaisuudet ovat sisältöriippumatonta metadataa. Toimituksen tehtäväalueelle sijoittuvat tiedontarpeet voidaan muuttaa käyttöliittymän toimenpiteiksi tarjoamalla tekstipohjaisia kyselyitä tukeva väline ja joku selain sisältöriippumattoman metadatan selaamiseen ja suodattamiseen. Suuri osa luetelluista videodokumenttien ominaisuuksista on kuitenkin luonteeltaan faktatietoja, joita voidaan kysellä rakenteisesti. Varsinaisia hakunimekkeitä aiheeseen perustuvissa hauissa ovat videodokumentit.

Koko järjestelmän näkökulmasta tarkasteltuna ongelmana vain on, että aiheeseen ja taustatutkimukseen liittyvää sisällöstä riippumatonta metadataa ei voida tuottaa automaattisilla indeksointimenetelmillä, esimerkkinä tekijänoikeuksiin liittyvät tiedot ja kuvausvuosi (ja -paikka). Jos kyseessä olisi alun perinkin digitaalisella järjestelmällä tuotettu materiaali, suuri osa tämänkaltaisista tiedoista saataisiin automaattisesti. Äänen ja värien olemassaolo videodokumentissa voidaan indeksoida automaattisesti ilman suurempia ongelmia eikä haettavan sekvenssin pituuden rajaaminen sekunneissa tuota ongelmia. Mitä tulee ohjelman tyyppiin ja käsiteltävään aiheeseen, jos ne halutaan indeksoida automaattisesti, indeksointijärjestelmän on tuettava automaattista semanttista päättelyä. Toisaalta, koska tässä tutkielmassa on keskitytty TV-uutisiin, jäisi ongelmaksi enää uutisjutun aiheen tunnistaminen, joka voitaisiin johtaa ääniraidan transkriptiosta.

Esteettisiä, visuaalisiin piirteisiin kohdistuvia tiedontarpeita voidaan myös täsmentää. Markkulan [2002] mukaan videoita selailtaessa toimittajat kiinnittävät huomiota kuvassa

1. tekstitykseen
2. ääniraitaan ja sen mielenkiintoisuuteen sekä käyttökelpoisuuteen (käsiteltävän aiheen kannalta)
3. esityksen havainnollisuuteen ('clarity')
4. kauneuteen (eli otosten “hyvyyteen”)
5. otosten erilaisuuteen (kuten kuvausetäisyyteen ja kuvakulmaan)
6. toimintaan tai sen puuttumiseen
7. kuvassa mahdollisesti esiintyviin henkilöihin
8. liikkeeseen
9. otosten väreihin (ja eri otosten värien täsmäävyyteen)
10. vuodenaikaan (ohjelmaan pyritään saamaan kuvamateriaalia siltä vuodenajalta, jolla ohjelma esitetään). [Markkula 2002.]

Indeksointijärjestelmän kannalta tämänkaltaiset ominaisuudet ovat sisältöä koskevaa metadataa. Hakunimekkeitä, joihin haut kohdistetaan, ovat tässä tapauksessa havaittavat piirteet eli objektit ja niiden liikkeet sekä muut vastaavat ominaisuudet. Esteettiset työvaiheet edellyttävät käytännössä pääosin selailua, vaikka myös (visuaalisia) kyselyitä voitaisiin käyttää. Sekä visuaalisen selailun että tekstuaalisten kyselyjen tukemista, edellä esitettyjen seikkojen lisäksi, puoltavat myös toimituksen ammattiryhmien väliset erot. Nykyisen käytännön mukaan journalistit materiaalia valitessaan selaavat ensin videomateriaalin tekstuaalisia kuvailuja; paperille tulostettuja kuvailuja vertaillaan ja sopiva videomateriaali tilataan, minkä jälkeen videoita katsotaan nauhalta ja niistä etsitään haluttuja kohtia [Markkula 2002]. Työtapa vastaa aiemmin käsiteltyä hakustrategiaa, jossa ensin käytetään tekstuaalisia kyselyjä aineiston suodattamiseksi ja vasta tämän jälkeen selataan. Videoiden katsominen nauhalta videonauhurin kankean käyttöliittymän avulla on juuri niitä työskentelytapoja, jotka pitäisi korvata jollakin toimivammalla ratkaisulla. Käyttöliittymän pitäisi tukea siis selailua ja olla sujuvampi kuin perinteinen videonauhurin käyttöliittymä. Lisäksi toimituksen ongelmiin nykyisten järjestelmien kanssa kuuluu Markkulan [2002] mukaan se, että tulosjoukot ovat liian isoja eikä ole keinoja kyselyn rajoittamiseen tai suodattamiseen.

Esteettisten, havaittaviin piirteisiin kohdistuvien, tiedontarpeiden määritteleminen toimituksen tarpeista on monimutkaista. Edellisen luettelon kohdista 2, 3 ja 4 ovat niin

subjektiivisia, että ne edellyttävät selailua: kukin hakija on paras arvioimaan, mikä hänen mielestään mielenkiintoista, käyttökelpoista, havainnollista ja kaunista. Ei pidä unohtaa käyttäjän kykyä tehdä arvioita videon sisällöstä ja sen semanttisista merkityksistä; hyvä käyttöliittymä ja -hakuväline tukee käyttäjää, mutta ei ajattele hänen puolestaan. Kohta 1 eli tekstitys ja sen läsnäolo kuvasekvenssissä voidaan havaita ja tunnistaa automaattisesti kuten myös kohdan 5 kriteerit eli kuvausetäisyys ja -kulma. Kohtaan 2 liittyen järjestelmän pitäisi tukea puheentunnistusta puheraidan muuttamiseksi tekstuaaliseen muotoon ja (tai) kuvatekstien tunnistamista, jotta olisi jotain, mihin tekstuaaliset kyselyt voi kohdistaa. Vastaavasti kohdat 6–9 voidaan käsitellä osittain automaattisesti: kuvasekvenssistä voidaan tunnistaa objektien läsnäolo ja niiden liikkuminen. Kohta 10 voidaan periaatteessa toteuttaa värijakaumia tarkastelemalla; tarkkuus voi olla kyseenalaista. Automaattisesti tunnistettavia osia voidaan kysellä myös tekstuaalisen liittymän kautta, vaikkakin käytännössä nämäkin kohdat vaativat selailua semanttisen päättelyn epätarkkuuden vuoksi.

5.3.2 Arviointikriteerit

Videohakujärjestelmiä arvioidaan tässä tutkielmassa Leen ja Smeatonin [1999 & 2002] kriteerein. Näihin lainattuihin kriteereihin päädyttiin käytännöllisistä syistä: kaikki järjestelmät eivät ole yleisessä käytössä eikä niistä ollut kaikista saatavilla toisiaan vastaavia tietoja yleisesti. Leen ja Smeatonin [1999 & 2002] kriteerit vastaavat kuitenkin hyvin pitkälle niitä tarpeita, joita edellä toimituksen käytännöistä johdettiin: vaikka tutkielmaa tehtäessä olisikin ollut mahdollista saada käyttöön esiteltävät järjestelmät, arvioitavat ominaisuudet olisivat olleet pitkälti samat. Käytettävät arviointikriteerit koskevat järjestelmien hakuominaisuuksia eli niiden tarjoamaa tukea analyyttisille kyselyille ja selailulle. Arvioinnissa kiinnitetään huomiota seuraavin ominaisuuksiin:

- Tukeeko järjestelmä (1) luonnollisella kielellä tehtyjä tai avainsanoihin perustuvia tekstuaalisia kyselyjä; käyttääkö se (2) ääniraidan informaatiota indeksoimiseen ja hakuihin; osaako se tuottaa (3) ääniraidasta transkriptioita.
- Tukeeko järjestelmä visuaalisia kyselyjä eli (4) histogrammien käsittelyä, (5) avainkehyspohjaista hahmonpiirtämistä, (6) avainkehyspohjaista kyselyä esimerkillä ('query-by-example'), (7) liikepohjaista hahmon piirtämistä tai (8) liikepohjaista kyselyä esimerkillä.
- Tukeeko järjestelmä videon sisällysluetteloiden ('video abstractions') tuottamista ja se-

laamista (9) avainsana- tai kategorialuettelon, (10) tekstuaalisen kuvailun, (11) transkription, (12) yksittäisen avainkehyksen, (13) kronologisen avainkehysluettelon (ns. ”storyboard”), (14) eri tasoilla tiivistetyn avainkehysluettelon, (15) vuorovaikutteisen hierarkkisen avainkehysselaimen, (16) ajastetun ('timed') avainkehyksien toiston, (17) kohokohtien ('skim') toistamisen tai (18) kokonaisten nimekkeiden toiston avulla; tukeeko järjestelmä eri menetelmien kuten (19) transkription ja toiston synkronisaation, (20) avainkehyksien ja toiston synkronisaation, (21) tekstuaalisen haun, toiston ja avainkehyksien synkronisaation yhdistelmiä tai (22) avainkehyksien älykästä valitsemista.

Järjestelmiä arvioitiin niin, että jokaisen edellä luetellun kohdan toteuttamisesta saa kaksi pistettä; jos kriteeriä ei täytetä pisteitä ei tule, ja kohtien 2–3, 12 ja 18 puuttumisesta menettää kaksi pistettä. Painotuksia perustellaan sillä, että kyseiset vaatimukset ovat osoittautuneet kirjallisuuden valossa erityisen tärkeiksi: ääniraidasta on paljon hyötyä kasvojen tunnistamisessa ja avainkehykset ovat selaamisen kannalta välttämättömiä.

5.3.3 Arvioitavat hakuvälineet

Arvioinnin tarkoitus on tutkia missä määrin Leen ja Smeatonin [1999, 7–10] luettelemat välineet, jotka mahdollistavat digitaalisen videon sisällön käytön ('access') ja selailun, vastaavat edellä määriteltyihin vaatimuksiin. Läheskään kaikki niistä eivät ole yleisesti käytettävissä, joten analyysi perustuu vain Leen ja Smeatonin [1999 & 2002] ilmoittamiin ominaisuuksiin. Luettelo ei ole missään tapauksessa kata kaikkia saatavilla olevia videotiedonhaun välineitä: lisää hakuvälineitä ja -järjestelmiä luettelevat esimerkiksi Lee ja Smeaton [1999], Del Bimbo [1999, 66–67] ja Antani ja muut [2002, 947–949]. Arvioitavat hakuvälineet on valittu havainnollistamissyistä: koska valitut välineet on suunniteltu eri tarkoituksiin, niiden väliset erot osoittavat, kuinka tärkeää on kuhunkin tarkoitukseen sopiva väline.

- *VISION* eli *Video Indexing for Searching Over Networks* [ks. Gauch 1998] on Kansasin yliopiston kehittämä järjestelmä, jonka tarkoitus on välittää uutislähetyksiä verkossa. Järjestelmää sovelletaan *Digital Jayhawk* -multimediatietokannassa [11], joka sisältää uutisartikkeleita ja -lähetyksiä.
- *VideoSTAR* eli *Video Storage And Retrieval* [12] [ks. Hjelsvold, Lagorgen, Midtstraum &

11 Digital Jayhawk: <URL: http://www.digitaljayhawk.org/news/>
12 VideoStar: <URL: http://www.idt.unit.no/~videodb/videostar/> ja
<URL: http://www.idt.unit.no/~videodb/videostar/interface-examples.html>

Sandsta 1995] on yleinen tietokanta-alusta ('platform'), jossa on kehitetty useita indeksointi- ja hakuvälineitä ammattimaisia kirjaston- ja arkistonhoitajia varten TV-uutislähetysten ja filmien dokumentointiin ja hakemiseen.

- *OLIVE* [13] on projekti, joka pyrkii kehittämään monikielisen, automaattisen alaotsikoihin ja puheentunnistukseen keskittyvän videoita arkistoivan järjestelmän.
- *VideoQ* [14] on kokeellinen visuaalisia kyselyvälineitä tarjoava järjestelmä, joka mahdollistaa väreihin, tekstuuriin, muotoihin ja liikkeeseen pohjautuvat kyselyt.
- *NeTra-V* [15] on kokeellinen matalan tason sisältöpohjainen kyselyväline ('query tool').
- *Screening Room* [16] on automaattinen videon arkistointi- ja hakujärjestelmä, joka on tarkoitettu lähetystoiminnan harjoittajille ('broadcasters'), videon tuottajille yms.
- *VideoLogger* [17] on kaupallisessa käytössä oleva automaattinen reaaliaikainen videon kommentoinnin ('logging') ja luetteloinnin järjestelmä.

Taulukko järjestelmille annetuista pisteistä on esitetty liitteessä 7.

5.3.4 Kommentteja hakuvälineiden ominaisuuksista

Jos kyseessä olisi yksinkertainen kilpailu tehtävään parhaiten sopivasta järjestelmästä, sen voittajat olisivat Viragen Videologger ja OLIVE, jotka keräsivät 18 pistettä. Asia ei ole kuitenkaan aivan näin yksinkertainen, sillä paremmuus riippuu ominaisuuksien painotuksesta eli myös käyttötarkoituksesta. Pisteytys olisi ollut erilainen, jos ominaisuuksissa olisi painotettu enemmän visuaalisia kyselyjä (kriteerit 4–8), joita voidaan käyttää semanttisten kyselyiden korvikkeena.

Leen ja Smeatonin [1999 & 2002] mukaan useimmat järjestelmät sisältävät tekstuaalisen välineen kyselyjä varten, vaikkakin jotkin järjestelmät mahdollistavat myös joitakin visuaalisia kyselyjä. Ensimmäisen kriteerin täyttävät kaikki muut järjestelmät paitsi NeTra-V, joka oli Leen ja Smeatonin [1999, 16] mukaan kesken. Parhaiten menestyneet järjestelmät, OLIVE, Screening Room ja VideoLogger, käyttivät jollain tasolla hyväkseen ääniraidan informaatiota – kuten evaluointikriteereissä painotettiinkin – OLIVE ainoana tuotti myös täyden transkription ääniraidasta. Visuaalisille kyselyille ei löytynyt kovinkaan kattavaa tukea käsitellyistä järjestelmistä: VideoQ tuki liikepohjaisia kyselyitä, mutta mikään hakuväline ei tuke-

13 OLIVE: <URL: http://twentyone.tpd.tno.nl/olive/>
14 VideoQ: <URL: http://www.ctr.columbia.edu/videoq/>
15 NeTra-V: <URL: http://vision.ece.ucsb.edu/netra/NetraV.html>
16 Screening Room: <URL: http://www.convera.com/Products/products_sr.asp>
17 VideoLogger: <URL: http://www.virage.com/products/details.cfm?productID=5&categoryID=1>

nut avainkehyspohjaisia kyselyitä. Tiivistelmien tuottaminen oli yleisesti ottaen visuaalisia kyselyjä kattavammin tuettu: monet järjestelmät tukivat transkriptioita ja lähes kaikki tuottivat yksittäisiä avainkehyksiä. Parhaimmissa järjestelmissä (OLIVE, Screening Room ja VideoLogger) oli myös tuki kronologiselle avainkehysluettelolle. VideoLogger tuki myös eri abstraktiomenetelmien yhdistämistä, kuten tekstuaalisen haun ja toiston synkronisointia.

Useimpien järjestelmien selailuvälineet ovat avainkehyspohjaisia. Kaikki järjestelmät mahdollistavat toiston, mikä Leen ja Smeatonin [2002] mukaan osoittaa, että sitä pidetään käyttöliittymän tärkeimpänä ominaisuutena. (Toiston yleisyys ominaisuutena voi johtua myös siitä, että se on helppo toteuttaa). Useimmat järjestelmät tarjoavat useamman kuin yhden videon selailumenetelmän, jotka usein koostuvat yhdistelmästä transkriptiota ja toistoa sekä avainkehyksiä ja toistoa. Yksikään Leen ja Smeatonin [2002] luettelema järjestelmä ei toteuta kaikkia heidän luettelemiaan ominaisuuksia, mikä heidän mukaansa tarkoittaa, että eri järjestelmät on tarkoitettu erilaisia käyttäjiä ja tarkoituksia varten. Kuitenkin joistakin järjestelmistä selvästi puuttuu niiden kohderyhmän tarpeiden kannalta hyödyllisiä ominaisuuksia. Varsinkin järjestelmien tuki visuaalisille kyselyille ja selaamisen kehittyneemmille menetelmille oli varsin rajoittunut: monissa kohdin kriteereille ei löytynyt tukea mistään järjestelmästä, ei edes eniten pisteitä saaneista. Monia aiemmin tässä luvussa käsiteltyjä hyödyllisiä selailuvälineitä ei tuettu laisinkaan. Vaikka järjestelmät olisikin suunniteltu jotain tiettyä tarkoitusta varten, ne hyötyisivät monista puuttuvista ominaisuuksista suunnitellun tarkoituksen toteuttamisessa.

Tässä luvussa käsiteltiin videotiedonhaun käyttöliittymiä, videosisällön visualisointia eri tasoilla sekä olemassa olevia videotiedonhaun hakuvälineitä. Seuraavassa luvussa pohditaan muun muassa sitä, miten näitä menetelmiä voidaan soveltaa TV-uutisten yhteydessä.

6 TV-UUTISLÄHETYKSEN JÄSENTÄMINEN

Tämän luvun tarkoitus on pohtia uutislähetysten mallintamista ja videosisällön visualisointia konkreettisessa Yleisradion uutislähetyksessä [liite 8]. Luvussa tarkastellaan elementtejä, joista uutislähetys on koottu eli ajallisia jaksoja ja paikkasidonnaisia sommitelmia, ja tämän jälkeen tarkastellaan uutislähetyksen indeksointia: ajallisen rakenteen jäsentämistä ja sisällön

tunnistamista.

6.1 Uutislähetyksen rakenteelliset mallit

Vaikka käytännöt vaihtelevat instituutiosta toiseen, TV-uutislähetys koostuu aina pienehköstä joukosta ajallispaikallisia elementtejä, joita järjestelemällä uutislähetys rakennetaan. Hietalan [1996, 63] mukaan uutislähetys muodostuu seuraavista osista: (1) näkyvistä kertojista eli uutisankkureista, (2) kertovasta tilasta eli uutisstudiosta ja (3) kerrotusta tilasta eli studion ulkopuolisesta maailmasta, josta uutiset kertovat. Kertovaan tilaan kuuluu uutisankkureiden ja kommentaattoreiden lisäksi olennaisena elementtinä myös uutisikkuna, jossa voidaan esittää graafinen uutistunnus, maailmankartta tai käsiteltävän uutisjutun otsikko; uutisikkunan sijaan tai sen lisäksi voidaan näyttää kuvaa myös uutiskeskuksesta [ks. Hietala 1996, 63]. Muita elementtejä ovat aloitus- ja lopetustunnukset, jotka eivät ole niinkään kontekseja tai tiloja, joissa asiat tapahtuvat, vaan ajallisia kehyksiä, joilla osoitetaan siirtymistä genrejen välillä, uutisten kertovaan tilaan ja sieltä pois. Uutistunnuksen ja uutisstudion kaltaiset elementit ovat itsenäisiä merkityksellisiä segmenttejä, jotka tietyssä järjestyksessä esitettynä muodostavat uutislähetyksille ja uutisaiheille kehykset, joiden sisään on koottu kutakin aihetta koskevia merkityksellisiä alijaksoja: Pietilän [1995, 197] mukaan kehystäminen tarkoittaa “jutun tai sen jonkin osan aloittamista ja päättämistä toisiaan vastaavin kuvin”. Uutislähetyksessä uutisstudio muodostaa linkittävän kehysmallin, joka kehystää uutisjutut ja -sähkeet ja siirtää katsojan kerrottuun tilaan.

Seuraavaksi kuvaillaan YLE:n uutislähetyksen yksittäisiä kehyksiä koskevia paikkasidonnaisia malleja ja ajallisia yksiköitä koskevia malleja yleisellä tasolla. Kuten luvussa 4 esitetään, mallit koostuvat (1) aluemalleista, joilla tarkoitetaan uutisankkuria, uutisikonia, uutisohjelman otsikkopalkkia, uutisankkurin nimipalkkia ja taustaa, (2) kehysmalleista paikkasidonnaisina asetelmina edellä mainittuja aluemalleja ja (3) otosmalleista sekvensseinä edellä mainittuja kehysmalleja.

6.1.1 Aluemallit ja kehysmallit

Uutisstudio on indeksointijärjestelmän kannalta kehysmalli, joka muodostetaan joukosta aluemalleja. Mahdollisia aluemalleja ovat:

1. *Uutisankkuri*: Uutisankkuri on uutisstudion silmiinpistävin objekti uutisikkunan lisäksi.

Uutisankkuri on sijoitettu joko keskelle ruutua tai keskustasta hieman oikealle ja on läsnä uutisstudiossa koko lähetyksen ajan.

2. *Uutisikoni* (uutislähetyksen tunnus tai logo).
3. *Uutisikkuna* ja uutisaiheen otsikkopalkki: Uutisikkuna (tai "luukku", ks. Pietilä 1995, 23) sijaitsee ruudun vasemmassa yläreunassa; siinä esitetään kunkin uutisjutun aihe tai muuta informaatiota. Uutisikkuna kuuluu uutisstudion eri malleista kaikkiin muihin versioihin paitsi uutissähkeisiin.
4. *Nimipalkki*: Uutisankkurin nimi näytetään valkoisella kuvatekstillä läpikuultavan harmaassa taustassa ruudun alaosassa ensimmäisen uutisjutun yhteydessä. Uutisjuttujen ja -sähkeiden sisällä nimipalkkia käytetään toimittajien tai haastateltavien nimien esittämiseen.
5. *Tausta*: Taustalla saatetaan näyttää uutiskeskusta monitoreineen.

Uutisstudiosta on kuusi eri variaatiota: Kaikissa niistä kuvakulma on suoraan edestä, mutta kuvausetäisyys vaihtelee. Hallitsevimpia värisävyjä uutisstudiossa ovat taustan vihreä, musta, turkoosi ja harmaa sekä etualan kellertävän ruskea (pöydän väri) ja ankkurin ihon ja vaatetuksen värit. Variaatiot uutisstudion kehysmallista ovat seuraavat:

1. *Uutistunnukseen liittyvä malli*: Tätä mallia käytetään, kun uutisstudioon siirrytään ensimmäistä kertaa. Aluemalleihin kuuluvat (a) uutisankkuri, (b) uutisikkuna (tyhjä) ja (c) tausta. Kuvausetäisyys vaihtuu kaukaa lähelle.
2. *Uutisjutun juontoon liittyvä malli*: Aluemalleja ovat: (a) uutisankkuri, (b) uutisikkuna, (c) nimipalkki, (d) tausta. Nimipalkki näytetään vain ensimmäisen uutisjutun yhteydessä. Kuvausetäisyys on normaali.
3. *Uutissähkeeseen liittyvä malli*: Aluemalleja ovat: (a) uutisankkuri ja (b) tausta. Kuvausetäisyys on läheltä.
4. *Toiseen sisällysluetteloon liittyvä malli*: Aluemalleja ovat (a) uutisankkuri, (b) uutisikkuna ja (c) tausta. Kuvausetäisyys vaihtuu kaukaa lähelle.
5. *Säätiedotuksen juontoon liittyvä malli*: Aluemalleja ovat (a) uutisankkuri, (b) uutisikkuna ja (c) tausta. Kuvausetäisyys on normaali.
6. *Lopetustunnukseen liittyvä malli*: Aluemalleja ovat (a) uutisankkuri, (b) uutisikkuna (seuraavat lähetykset) ja (c) tausta. Kuvausetäisyys on normaali.

Myös uutistunnuksesta voitaisiin tehdä vastaava jäsennys, mutta sitä ei voi pitää kovinkaan tarkoituksenmukaisena videon semanttisen sisällön indeksoimisen kannalta. Seuraavaksi tar-

kastellaan, kuinka kehysmalleja järjestellään otostyyppien yhteydessä.

6.1.2 Otostyyppejä koskevat mallit

Uutislähetyksen ajallisten elementtien järjestys on YLE:llä nykyään [ks. Pietilä 1995, 173] seuraava: uutistunnus, tervehdys, uutisvinkit, uutisjutut ja -sähkeet, toinen sisällysluettelo (eli lisää vinkkejä), loput uutisjutut ja -sähkeet, sää, lopetuspuhe ja lopetustunnus. Näistä suurin osa käydään läpi vain kerran lähetyksen aikana. Lähetys on saavuttanut suurin piirtein puolivälinsä toisessa sisällysluettelossa (“tässä lähetyksessä...”) ja se alkaa olla lopussa päästäessä säätiedotukseen. Näitä elementtejä käsitellään seuraavaksi tarkemmin ja seuraavassa luvussa käsitellään niiden tunnistamista ja mallien avulla. Elementit on jaettu kolmeen ryhmään: tunnuksiin, sisällysluetteloihin ja juontoihin (joihin myös sähkeet lasketaan).

6.1.2.1 Tunnukset

Uutistunnus on otosmalli, joka toimii linkkinä kertovaan tilaan ja siitä pois [liite 9]. Tunnuksen tarkka mallintaminen alue- ja kehysmalleilla ei ole kuitenkaan tarkoituksenmukaista, kuten edellä todettiin: riittää, että se havaitaan. Uutistunnus muodostuu kahdesta otoksesta [liite 8: O1 ja O2], jotka muodostavat ensimmäisen ryhmän [liite8: R1]. Uutistunnus alkaa kuvan valoisuuden voimistamisella tummasta vaaleaan. Ensimmäisessä otoksessa on kaksi aluetta: alaosassa on musta ja turkoosi palkki, jossa otoksen loppuvaiheessa vilistää numerosarja, ja yläosassa on värisävyiltään pääosin turkoosi ja valkoinen graafinen animoitu tekstuuri. Toisessa otoksessa alueiden koko ja sijainti säilyy ennallaan, mutta alaosan palkkiin ilmestyy teksti “20:30” vaalealla tekstillä, jonka tilalle vaihtuu otoksen loppupuolella teksti “uutiset”. Ensimmäisen ja toisen otoksen välisen asteittaisen siirtymätehosteen loputtua kuvan suuremmassa alueessa näkyy valkoisista viivoista muodostuva objekti, joka on ilmeisesti abstrahoitu kello, joka liikkuu myötäpäivään. Sen taustalla näytetään kuvaa uutisstudiosta ja -ankkurista, kuvattuna kaukaa yläoikealta. Kameran liikettä ei ole kummassakaan otoksessa. Ensimmäisen ja toisen otoksen välinen siirtymä käyttää asteittaista pyyhkäisyä tehosteenaan. Uutistunnuksesta siirrytään uutisstudioon toisen otoksen jälkeen liuotukselta vaikuttavan asteittaisen siirtymätehosteen kautta [liite 10].

Lopetustunnuksen vasemmassa reunassa on vaaleansininen läpikuultava alue, joka peittää lähes puolet ruudusta; sen yläosasta, jossa esitetään juokseva kellonaika, ei kuitenkaan näy läpi [liite 11]. Kyseiseen alueeseen on kirjoitettu valkoisella tekstillä

“uutispäällikkö”, “kuvaussihteeri”, “ohjaaja” ja lyhyesti yhteystiedot. Oikeanpuoleisessa alueessa kuvakulma uutisstudioon on yläoikealta kuten uutistunnuksen toisessa otoksessa. Uutiset loppuvat kuvan häivyttämiseen mustan ruudun suuntaan. Tiedonhaun kannalta tunnukset eivät sisällä mitään erityisen mielenkiintoista.

6.1.2.2 Sisällysluettelot

Uutisvinkkien tarkoitus on esitellä uutislähetyksen aiheita [liite 12]. Pietilän [1995, 175] mukaan “niin YLE:llä kuin MTV:llakin ensimmäinen vinkki viittaa lähetyksen pääjuttuun, jonka paikka on heti vinkkijakson jälkeen”. Lisäksi vain varsinaisiin uutisjuttuihin vinkataan; vinkatut jutut ovat yleensä lähetyksen pisimpiä ja ne on sijoitettu lähetyksen alkupuolelle [Pietilä 1995, 175]. Uutisvinkit ovat siis eräänlainen sisällysluettelo, jossa näytetään illan tärkeimmät uutiset. Uutisvinkit ovat eräänlaisia uutislähetyksen keskeisten uutisjuttujen kohokohtia ('skims'). Uutisvinkeissä kuva on jaettu kahteen alueeseen: varsinaiseen kuvaan ja alaosan mustaturkoosiin palkkiin, jossa lukee “uutiset”. Uutisvinkeissä näytettävä kuvamateriaali on suoraan uutisjutuista leikkauksia myöten. Vinkeissä esiteltävät uutisjutut on erotettu selvästi havaittavalla pyyhkäisyllä. Uutisvinkkien osuus aloitetaan ja lopetetaan liuotuksella.

Toinen sisällysluettelo eli “tässä lähetyksessä kerromme vielä” on neljäs versio uutisstudiota koskevista kehysmalleista. Lähetyksen loppuosan uutisaiheet esitetään uutisikkunassa samaan tapaan kuin uutisvinkeissä; käytetty vaihtumistehostekin on sama. Erot alun uutisvinkkeihin ovat lähinnä siinä, että toisessa sisällysluettelossa vinkatut uutiset esitetään uutisikkunassa eikä koko ruudun kokoisina; uutisankkuri kommentoi sekä uutisvinkkejä että lähetyksen puolivälin sisällysluetteloa. Lähetyksen lopun uutisia vinkattaessa taustalla soitetaan uutistunnuksen musiikkia hiljaa. Kuvausetäisyys on aluksi suurempi kuin uutisstudiossa yleensä, mutta se tarkennetaan normaalietäisyydelle ennen siirtymistä lähetyksen loppupuolen ensimmäisen uutisaiheen käsittelyyn [vrt. liite 10]. Ensimmäiseen uutisjutun juontoon siirrytään hitaalla pyyhkäisyllä uutisikkunassa. Kyseisen siirtymän havaitseminen voi tuottaa ongelmia; jos sitä ei havaita, uutisvinkit päätyvät osaksi niiden jälkeen tulevaa uutisjuttua.

6.1.2.3 Juonnot, uutisjutut ja uutissähkeet

Uutistunnuksen jälkeen siirrytään uutisstudioon [liite 10]. Uutisankkuri tervehtii nopeasti katsojia, minkä jälkeen siirrytään ensimmäiseen sisällysluetteloon eli uutisvinkkeihin. Uutisstudiossa ei ole liikkuvia objekteja, mutta kamera tarkentaa kuvausetäisyyden ensimmäisellä ker-

ralla uutisstudioon siirryttäessä kaukaa normaalietäisyydelle. Juonnosta poistutaan uutisvinkkeihin ja säätiedotukseen liuotuksen välityksellä, mutta siirtymä varsinaisiin uutisjuttuihin on välitön.

Uutisjuttu aloitetaan uutisankkurin juonnolla uutisstudiossa [liite 13]. Uutisikkunassa näytetään uutisaiheen otsikko ja kuva uutisjutusta. Alhaalla näytetään toimittajan tai jonkin haastateltavan henkilön nimi harmaalla läpikuultavalla taustalla. Uutisjuttua juonnettaessa kuvakulma ja -etäisyys ovat kuten uutisstudiossa normaalisti. Uutisjuttuun ja siitä pois siirrytään välittömästi ilman siirtymätehostetta. Myös sähkeeseen siirrytään uutisjutun jälkeen välittömästi. Vain uutisstudion sisällä siirryttäessä normaalilta kuvausetäisyydeltä lähelle käytetään siirtymätehostetta: näin tehdään esimerkiksi, jos ankkuri kommentoi uutisjuttua toimittajan osuuden jälkeen, ja seuraavaksi vuorossa on sähke.

Uutisankkuri lukee uutisähkeen uutisstudiosta, mutta sähkeeseen saattaa kuulua kuvamateriaalia muualtakin. Jos sähkeeseen siirrytään uutisstudion kautta, siirtymätehosteena on liuotus, mutta uutisjutusta siirrytään sähkeeseen välittömästi. Sähkeen aikana kuvausetäisyys on lähempää ankkuria ja uutisikkunaa ei näy [liite 13]. Uutissähkeen jälkeen siirrytään välittömästi seuraavan uutisjutun juontoon, jossa kuvausetäisyys on taas normaali.

Säätiedotus muodostaa oman erillisen osansa uutislähetyksessä, mutta se juonnetaan uutisstudiossa. Uutisstudiosta siirryttään säätiedotukseen liuotuksen välityksellä. Siirtymä säätiedotuksesta takaisin uutisstudioon on kuitenkin välitön. Lopetuspuheenvuoron ankkuri esittää uutisstudiossa. Uutisikkunassa näkyy seuraavien lähetysten ajankohdat “21:50” ja “23:00”. Lopetustunnukseen siirrytään liuotuksen avulla.

6.2 Uutislähetyksen mallipohjainen indeksointi

Seuraavaksi tarkastellaan YLE:n uutislähetyksen automaattista indeksointia näkökulmasta, jossa otetaan huomioon nykymenetelmien tila: pääasiallisena indeksointitavoitteena on uutislähetyksen jäsentäminen uutisjuttuihin. Arviot visuaalisesta sisällöstä jätetään pääosin hakijalle, mutta hakujen rajaamiseksi ääniraidan transkriptioiden tuottaminen on välttämätöntä. Kaikki huomiot ovat tietysti teoreettisia, sillä ilman laboratoriotestejä on vaikea arvioida, kuinka indeksointijärjestelmä suoriutuisi käytännössä.

6.2.1 Indeksoitavat nimekkeet

Eräs keskeisimmistä päätöksistä, jotka hakujärjestelmää suunniteltaessa pitää tehdä, on määri-

tellä tiedonhaun perusyksiköt eli ne nimekkeet, joihin haut kohdistetaan. TV-uutisten yhteydessä on ilmiselvää pitää uutisjuttuja ja -sähkeitä haettavina tiedonhaun perusyksikköinä: ne eivät ole liian pitkiä ja ne käsittelevät yhtä tiettyä aihetta: semanttisesta sisällöstä kiinnostuneille hakijoille ne ovat ilmeisiä hakukohteita. [ks. Mills et al. 2000, 4]. On otettava huomioon, että jos haku kohdistetaan liian pitkiin segmentteihin, ne sisältävät hakijan kannalta liikaa irrelevanttia informaatiota; jos haku kohdistetaan liian lyhyisiin segmentteihin, informaatio on liian fragmentoitunutta [Mills et al. 2000, 4]. Uutiset voivat kiinnostaa tiedonhakijoita ainakin seuraavista syistä:

1. *Aihe*: Käyttäjät ovat kiinnostuneita aiheesta, jota uutisissa käsitellään. He haluavat etsiä kokonaisia uutisjuttuja tai -sähkeitä ja saada sen informaation, jonka uutislähetyksiä kokonaisuudessaan katsomalla voi saada.
2. *Yleiset objektit*: Käyttäjät hakevat tietyntyyppisiä objekteja, joista he ovat kiinnostuneita esteettisessä mielessä tai sitten käyttäjät ovat realistisia ja hakevat objekteja niiden havaittavien visuaalisten piirteiden avulla, joiden he olettavat olevan yhteydessä semanttisiin käsitteisiin.
3. *Tunnistetut objektit*: Käyttäjät ovat kiinnostuneita tunnistetuista ja nimetyistä objekteista, kuten henkilöistä, joita he haluavat hakea semanttisella tasolla.

Tämän luettelon perusteella voidaan erottaa dokumenttikeskeinen ja objektikeskeinen videotiedonhaku. Jos hakija on kiinnostunut jostakin aiheesta, hän etsii uutisjuttuja tai -sähkeitä; toisaalta käyttäjä voi olla kiinnostunut esimerkiksi uutisjutuissa esiintyvistä henkilöistä tai hän saattaa etsiä uutisaiheita henkilöiden (siis esimerkkikuvien) perusteella. Kohtien 1 ja 3 toteuttaminen vaatii tekstipohjaisen käyttöliittymän, jonka avulla käyttäjä kirjoittaa esimerkiksi hakemansa henkilön nimen ja sitten etsii tulosjoukosta selaimella ne kuvat, jotka todella vastaavat etsittyä henkilöä. Yleisiä objekteja (kohta 2) haetaan käytännössä visuaalisten kyselyjen avulla: käyttäjä suorittaa kyselyn esimerkkikuvalla, hahmotelmalla tai vaikka määrittelemänsä hahmon liikemallilla.

Jos indeksoinnissa tyydytään pelkkään uutislähetysten jäsentämiseen uutisjuttuihin ja -sähkeisiin niin, että varsinainen sisällön tunnistaminen jätetään pääosin käyttäjälle, on automaattinen indeksointi huomattavasti realistisemmin toteutettavissa kuin pyrittäessä tunnistamaan sisältöä semanttisella tasolla. Uutislähetyksissä pelkkä jäsentäminen on siinä mielessä riittävä vaihtoehto, että avainkehyksiä ei tule yhdestä uutisaiheesta selattavaksi kohtuutonta määrää: liitteessä 8 analysoitu uutisjuttu koostui 17:stä otoksesta. Varsinaiseksi ongelmaksi dokumenttikeskeisessä videotiedonhaussa muodostuu kuitenkin uutisaiheen tunnistami-

nen: tarvitaan joku keino, jolla hakija voi rajata selattavien nimekkeiden määrää sitä aihetta koskeviin, joista hän on kiinnostunut. Ei siis riitä, että uutisjutut ja -sähkeet jäsennetään, on pääteltävä aihe, josta ne kertovat. Tässä luvussa käsitellään seuraavaksi uutislähetyksen jäsentämistä, jäsentämisen toteuttamismahdollisuuksia sekä uutisaiheen tunnistamista sisällönanalyysilla.

6.2.2 Uutislähetyksen otostyyppien ajallinen malli

Uutislähetystä voidaan pitää eräänlaisena kaaviona, joka indeksointijärjestelmän näkökulmasta muodostuu elementeistä, jotka luvussa 6.1 esitettiin kehys- ja otostyyppejä koskevina malleina. Käsitellyt mallit koskevat kertovaa tilaa eli uutisstudiota ja kertojia; uutisstudion ulkopuolinen tila, josta varsinaiset uutisjutut- ja sähkeet kertovat, ei ole niin säännönmukainen, että se voitaisiin mallintaa. Seuraavaksi käsitellään otostyyppejä koskevien mallien järjestelyä YLE:n uutislähetyksessä [ks. liite 8].

Kuten luvussa 4 esitettiin, otokset voidaan merkitä symboleilla niiden sisällön visuaalisen samankaltaisuuden perusteella, jolloin symbolisia koodisekvenssejä tutkimalla voidaan päätellä, mitkä otokset esimerkiksi kuuluvat samaan ryhmään. Visuaalista samankaltaisuutta voidaan arvioida monilla eri perusteilla (ks. luvut 3 ja 4), joiden pitäisi vastata ihmisten arvioita; tilanteesta ja tarkoituksesta riippuen voidaan vertailla esimerkiksi koko kehyksen värijakaumia tai sitten kehysten alialueita, kuten objekteja – tästä lisää myöhemmin. Niin kauan, kun arvioidaan vain havaittavia piirteitä ja niiden vastaavuutta kahden eri otoksen välillä, ryhmittely perustuu visuaaliseen samankaltaisuuteen. Kohtaukset, uutisjutut ja -sähkeet ovat kuitenkin sisällöltään semanttisesti yhteneviä. Otosryhmän visuaalinen samankaltaisuus, jos se esiintyy tietyn mallin sisällä, voi merkitä myös semanttista yhteneväisyyttä, jolloin otosryhmä muodostaa kohtauksen. Mallilla voidaan tässä yhteydessä tarkoittaa esimerkiksi aikarajaa tai muuta ryhmien muodostamista koskevaa tietoa. Semantiikkaa ei kuitenkaan voi johtaa ilman apriorista tietoa siitä, missä yhteydessä mikäkin ero ja samankaltaisuus merkitsee mitäkin. Koodisekvensseistä voidaan päätellä mistä uutisjuttu tai -sähke alkaa ja mihin ne loppuvat, jos tunnetaan uutislähetyksen visuaalinen ja semanttinen rakenne, joita käsiteltiin edellä. Uutislähetyksen rakenne vain pitää ilmaista indeksointijärjestelmälle.

Seuraavassa esimerkissä esitetään symbolisin koodein YLE:n uutislähetyksen alkuosan rakenne. Sekvenssissä jokainen samankaltainen otos on merkitty samalla koodilla. Otostyypin perässä oleva luku osoittaa sitä, kuinka mones otos samantyyppisistä on kyseessä, ja lyhyt viiva tarkoittaa välitöntä ja pitkä asteittaista siirtymää:

A1---A2---B1---C1-C2-C3-C4-C5-C6---B2-D1---E1---F1-G1---H1---I1-J1-J2-K1---H2---D2-D3-D4-L1-M1-M2-B3---

Sekvenssissä A = uutistunnus, B = uutisstudio ja C = uutisvinkit. Varsinaista uutisjuttua edustavat merkit D–M. Uutislähetyksen rakenteen tuntien pelkän koodisekvenssin avulla on mahdollista luokitella ryhmiä etsimällä uutisstudiota koskevia kehysmalleja (B). Tästä sekvenssistä on helppo erotella omaksi ryhmäkseen esimerkiksi uutisvinkit – nehän on kehystetty uutisstudion merkeillä ja asteittaisilla siirtymillä. Myös ensimmäisen uutisjutun alku on helppo tunnistaa. Sen sijaan uutisjutun loppupäässä tulee ongelmia: kun B3 normaalisti tarkoittaisi seuraavan uutisjutun alkua, niin esimerkkitapauksessa B3 kuuluu vielä ensimmäiseen uutisjuttuun. Lisäksi siitä siirrytään uutisjutun sijaan uutissähkeeseen, joka luetaan uutisstudiosta – siksi asteittainen siirtymä, joka on merkitty sekvenssin loppuun. Tämä uutisjuttujen erottamista koskeva ongelma voidaan kuitenkin ratkaista kehysmallin avulla tarkastelemalla uutisikkunaa, jonka sisällön perusteella voidaan päätellä, että B3 kuuluu vielä ensimmäiseen uutisjuttuun sen sijaan, että se aloittaisi uuden. Toisen sisällysluettelon (“tässä lähetyksessä kerromme...”) uutisikkunassa käytettävää tehostetta käytetään myös siirryttäessä uutisstudion sisällä uutisjutusta toiseen. Uutisikkuna onkin viimeinen keino havaita uutisjutun vaihtuminen. Seuraavaksi arvioidaan, kuinka edellä symbolisesti esitetyn uutisjutun tunnistaminen mahdollisesti onnistuisi käytännössä ja millaisia menetelmiä kannattaisi käyttää indeksoinnissa.

6.2.3 Ajallisen rakenteen automaattinen jäsentäminen

Uutislähetyksen ajallisen rakenteen jäsentäminen aloitetaan segmentoimalla lähetys otoksiin. Tämän jälkeen otokset luokitellaan visuaalisen samankaltaisuuden perusteella uutistunnuksiin, uutisstudiota ja -ankkuria koskeviin, uutisvinkkeihin ja varsinaisia uutisia koskeviin otoksiin vertailemalla otoksista poimittuja avainkehyksiä otosmalleihin. Uutislähetyksen rakennetta koskevan apriorisen tiedon avulla otostyyppien järjestyksestä voidaan päätellä, mitkä otostyypit kuuluvat mihinkin uutislähetyksen jaksoon. Kuten luvussa 6.2.2 esitellystä koodisekvenssistä käy ilmi, uutisten jäsentäminen sisällön perusteella on periaatteessa yksinkertaista. Jäsentäminen perustuu uutisstudion tunnistamiseen. Uutisjuttu alkaa aina uutisstudiosta ja päättyy uutisstudioon: uutisjutun jälkeen uutisankkuri joko kommentoi edeltänyttä uutisjuttua tai siirtyy heti seuraavaan aiheeseen. Uutisjutun vaihtuminen voidaan havaita viimeistään vertaamalla uutisikkunaa uutisjutun aloittaneen uutisstudiota koskevan otoksen uutisikkunaan.

Uutisjuttua voi seurata myös sähke, joka tunnistetaan uutisikkunan puuttumisesta ja lähemmästä kuvausetäisyydestä.

Luvussa 4 määriteltiin ensimmäiseksi indeksointitehtäväksi segmentointi ja ajallisen rakenteen jäsentäminen, ja tätä tarkoitusta varten suunniteltuja menetelmiä käsiteltiin vastaavasti. Samassa luvussa tuotiin myös ilmi, että leikkausten tunnistaminen ei yleensä aiheuta suuria ongelmia, vaikka asteittaisista tehosteista liuotuksen on raportoitu olleen joissakin tapauksissa vaikeasti tunnistettavissa. Liitteessä 8 analysoidun uutisjutun 17:stä otoksesta yksitoista päättyy asteittaiseen siirtymään; yleisin asteittainen siirtymä on nimenomaan liuotus. Segmentoinnissa on otettava huomioon, että aina ei ole mahdollista tai mielekästä muodostaa segmenttejä otosrajojen perusteella: siirtymätehosteet saattavat tehdä otosten välisistä eroista vaikeasti havaittavia, ja joissain tapauksissa objektit aiheuttavat kuvassa otosrajoihin verrattavissa olevia silmiinpistäviä muutoksia. Esimerkiksi uutistunnuksen otosrajat eivät pelkästään ole epäselvät kehysten välisten pienten erojen vuoksi, koska niiden välinen siirtymätehoste kestää kauan, vaan lisäksi otoksien sisällä ilmestyy ja katoaa objekteja. Pitäisikö segmentit muodostaa otosrajojen perusteella, vai pitäisikö segmentoinnissa ottaa huomioon keskellä otosten välisen siirtymän kuvaan ilmestyvät silmiinpistävät objektit?

Koska kertovaa tilaa koskevissa otostyypeissä ei ole nopeaa kameran tai objektien liikettä, histogrammipohjaisten segmentointimenetelmien pitäisi sopia hyvin niiden erottelemiseen kerrottua tilaa koskevista otoksista. Valitettavasti siirtymiä uutisstudion sisällä ei voida havaita histogrammipohjaisten menetelmien avulla pienten kehysten välisten muutoksien vuoksi. Esimerkiksi uutistunnuksen yhteydessä ja ensimmäistä kertaa uutisstudioon siirryttäessä ongelmia tulisi erittäin hitaiden siirtymätehosteiden takia. Uutisstudion sisäisissä siirtymissä segmentointia varten tarvitaan hahmontunnistusta, jolla etsitään kehysmallien perusteella uutisikkunaa, ja lohkopohjaisia indeksointimenetelmiä, joita sovelletaan uutisikkunassa tapahtuviin muutoksiin. Segmentointia ei voi kuitenkaan pitää varsinaisena ongelmana, sillä algoritmit voidaan sovittaa uutislähetyksiä varten. Kynnysten asettaminen segmentointimenetelmiä varten on helppoa, koska uutislähetysten visuaalinen ilme pysyy samana pitkään. Kun tiedetään alue ja otostyyppi, jossa muutokset tapahtuvat, ei uutisstudion sisäistenkään siirtymien havaitsemisen pitäisi muodostaa ylitsepääsemättömiä ongelmia. Asteittaisten siirtymätehosteiden tunnistamiseen soveltuvat parhaiten mallintamiseen perustuvat menetelmät, sillä siirtymätehosteet ovat samanlaisia lähetyksestä toiseen. Myös kaksoisvertailuun perustuvia menetelmiä voidaan käyttää asteittaisten siirtymien tunnistamisessa. Mitä tahansa menetelmää käytetäänkin, tärkeää on estää uutisjuttuihin kuuluvien otosten sekoittuminen uutissähkeisiin ja lähetyksen puolivälin vinkkien sekoittuminen niitä seuraavaan uutisjuttuun.

Kun otokset on segmentoitu, niistä valitaan avainkehyksiä, joita vertailemalla

otokset ryhmitellään ja merkitään symbolisilla koodeilla. Otosten välistä samankaltaisuutta voidaan arvioida monilla eri perusteilla. Histogrammipohjaiset menetelmät soveltuvat hyvin uutisstudiota koskevien avainkehysten erottelemiseen uutisjuttuja koskevista avainkehyksistä. Uutisstudiolle ovat ominaisia grafiikasta johtuvat vihreät, vaalean siniset ja turkoosit värisävyt. Paljon vaikeampaa on tunnistaa uutisstudiota koskevat eri otostyypit. Kuvausetäisyyden tunnistamiseksi tarvitaan objektien muotoihin perustuvia menetelmiä. Objektien muotojen havaitsemisen ei pitäisi tuottaa ylipääsemättömiä ongelmia, sillä riittää, että löydetään kehysmallin aluemallien kanssa riittävän samankaltaisia alueita ja niiden välille kehysmallissa määritellyt suhteet, esimerkiksi uutisstudiota vastaava paikkasidonnaisten elementtien sijoittelu. Uutisankkuri on silmiinpistävä objekti jokaisessa uutisstudiota koskevassa kehysmallissa. Jos pystytään tunnistamaan uutisankkuri sen viemän tilan perusteella, voidaan erotella uutissähkeet muista uutisstudiota koskevista kehysmalleista: sähkeissähän ankkuri on keskellä ruutua ja kuvattu huomattavasti lähempää kuin normaalisti. Uutistunnuksen ja lopetuspuheenvuoroa koskevien otosten erottaminen uutisjuttujen juonnoista vaatii uutisikkunan tarkastelua esimerkiksi tekstuurien perusteella; uutisikkunan tunnistaminen ei tuota ongelmia, sillä se on silmiinpistävästi kehystetty neliskulmainen alue hieman ruudun vasemmassa yläreunassa, uutisankkurin vasemmalla puolella. Uutistunnuksen ja lopetuspuheenvuoron yhteydessä uutisikkunassa on vihertävää grafiikkaa, mutta uutisjuttujen juontojen yhteydessä uutisikkunassa esitetään uutisjutun otsikko ja pysäytyskuva itse jutusta. Näiden väliset erot ovat yleensä selkeät kuten liitettä 10 ja 13 vertaamalla voi todeta. Koska ensimmäistä kertaa uutisstudioon siirryttäessä ja toisen sisällysluettelon yhteydessä kamera tarkentaa kaukaa lähelle, kameran liikkeen tunnistamisella voidaan vielä varmistua, että mikä uutisstudiota koskeva otostyyppi on kyseessä. Tarkennukseen liittyen voidaan vertailua varten ottaa useampia avainkehyksiä, esimerkiksi otoksen ensimmäinen, keskimmäinen ja viimeinen. Uutisjuttujen juonnoissa uutisankkurin koko pysyy samana koko otoksen läpi, joten niitä varten riittää yksi ainoa kehysmalli.

Jos otosten segmentointi on realistisesti mahdollista, ja sitä on tutkittu paljon, otoksia korkeampia yksikköjä muodostettaessa on oltava huomattavasti varautuneempi joskaan ei pessimistinen, sillä myös uutislähetysten mallintamista on tutkittu paljon. Menetelmiä käsiteltiin tätä tarkoitusta varten luvussa 4, ja edellä pohdittiin otosten välisen samankaltaisuuden arvioimista ja otosten luokittelemista kehysmallien avulla. Kun uutislähetys on segmentoitu otoksiin ja nämä otokset ryhmitelty visuaalisin perustein, apriorisen tietämyksen avulla niille annetaan semanttinen merkitys eli jäsennetään otosryhmät niiden sisällön perusteella. Visuaalisten piirteiden ja semanttisten merkitysten samankaltaisuuksien ja erojen käyttäminen mielekkäällä tavalla on yksi uutislähetyksen jäsentämisen suurimpia haasteita: joskus

visuaalisesti samankaltaiset otokset kuuluvat semanttisesti eri ryhmiin, joskus semanttisesti yhtenevät otokset näyttävät erilaisilta.

Esimerkiksi uutislähetyksen esittäminen kohtauksensiirtymiskaavion avulla ilman älykästä otosten ryhmittelyä on ongelmallista: Uutislähetyksessä erityispiirteenä on se, että siinä esiintyy hyvin pitkällä ajanjaksolla (n. 25 minuuttia) hyvin samanlaisia otoksia, jotka kuuluvat semanttisesti eri ryhmiin. Uutisjutun juonto lasketaan uutisjutun osaksi, koska se sisältää uutisjuttuun liittyvää informaatiota. Tätä huomiota tukee se, että uutisjutun juonnosta siirrytään varsinaiseen uutisjuttuun välittömän siirtymän kautta. Valitettavasti juontoja koskevat otokset muistuttavat enemmän toisiaan kuin uutisjuttuja joihin ne kuuluvat. Jos kaikki uutisstudioita koskevat otokset ryhmitellään pelkästään visuaalisten piirteiden perusteella yhdeksi solmuksi, leikkausyhteyttä solmujen välille ei löydy, koska kohtauksensiirtymiskaavio on koko uutislähetyksen ajan aktiivinen uutisstudiota koskevan solmun ympärillä, johon valtaosa muista solmuista on yhteydessä. Normaalisti kohtauksen raja tulisi vastaan kaavion ollessa ohut eli kun otoksesta on yhteys vain yhteen sellaiseen otokseen, jota ei ole aiemmin käytetty. Koska uutislähetyksessä uutisstudiota koskevia otoksia käytetään koko lähetyksen ajan, leikkausyhteyttä ei löydetä ja järjestelmä luokittelee koko uutislähetyksen yhdeksi isoksi kohtaukseksi. Uutisstudiota koskevien otoksien pitäisi siis kuulua eri ryhmiin. Uutislähetysten yhteydessä ryhmittelyssä ei ole niinkään kysymys otosten ajallisesta esiintymisestä vaan visuaalisesta ulkonäöstä ja otosten esiintymisjärjestyksestä. Ryhmittelyn ei siis pidä olla aikarajattua tai aikamukautuvaa vaan sellaiseen mallintamiseen pohjautuvaa, jossa järjestelmälle ilmaistaan eksplisiittisesti, että varsinaista uutisjuttua edeltävä juonto kuuluu uutisjutun yhteyteen.

Edellä pohdittiin keinoja erottaa toisistaan otoksia, jotka ovat visuaalisesti hyvin samankaltaisia, mutta joiden tiedetään kuuluvan semanttisesti eri ryhmiin. Joissain tilanteissa ongelma on päinvastainen: uutistunnuksen kahden otoksen ja uutisvinkkien kohdalla on huolehdittava siitä, että tietyt otokset sijoitetaan samaan ryhmään visuaalisista eroista huolimatta. Ongelman ratkaisu on samankaltainen kuin uutisstudion eri otostyyppejä eroteltaessa: sen sijaan, että eroja etsittäisiin kuvan yksittäisten piirteiden avulla, etsitään samankaltaisuutta kehyksen yhtenevistä piirteistä; tällaisesta piirteestä käy uutisvinkkien alaosan mustaturkoosi palkki.

6.2.4 Sisällön ja aiheen tunnistaminen

Kuten aiemmin esitettiin, pelkkä uutislähetysten ajallinen jäsentäminen ei riitä, jos käyttäjät ovat kiinnostuneita uutisjuttujen ja -sähkeiden aiheista. Tätä varten tarvitaan jonkin tasoista

sisällöntunnistamista tai ainakin havaittaviin piirteisiin pohjautuvia visuaalisia hakuja, joilla hakija voi kysellä uutisdokumentteja niiden piirteiden perusteella, joiden hän olettaa liittyvän niihin semanttisiin merkityksiin, joista hän on kiinnostunut. Käytännössä kuitenkin puhe on keskeisessä asemassa uutisaiheiden tunnistamisessa ja hakemisessa.

Vaikka kertovan tilan objektien havaitseminen on tärkeää uutislähetyksen jäsentämisen kannalta, tiedonhaun kannalta ne eivät ole kovinkaan tärkeässä asemassa – lukuun ottamatta uutisjuttujen juonto-osuuksien uutisikkunoita. Objektien tunnistaminen semanttisella tasolla kerrotussa tilassa eli yleisellä alueella on tällä hetkellä tavoite, jonka toteuttamista ei voida pitää kovinkaan realistisena. Kasvontunnistusta voidaan yrittää uutisjutuista esimerkiksi Name-It-järjestelmällä, joka tunnistaa kasvoja kuvatekstien ja ääniraidan avulla. Teoriassa myös yleisten objektien tunnistamista – kasvojen lisäksi – rajoittamattomissa konteksteissa voitaisiin yrittää ääniraidan transkription avulla. Ongelmaksi muodostuu kuitenkin kuvan ja äänen satunnainen vastaavuus. Otoksessa 10, josta uutisjuttu alkaa, näkyy uutisikkunassa teksti “Doping” ja transkriptiossa mainitaan, että “maastohiihto on jälleen joutunut doping-huhujen kohteeksi – – “. Seuraavassa otoksessa mainitaan, että “suomalaisen maastohiihdon kärki kärysi dopingista” ja samaan aikaan ruudussa näkyy hiihtäjä ladulla. Otoksessa 15 mainitaan kuvatekstissä “lääkintäeversti Heikki Laapio” samaan aikaan kun ääniraidalla puhuu Laapio – puhe tosin alkaa jo edellisestä otoksesta. Edellä mainituista yhteensattumista huolimatta yleisten objektien tunnistaminen ääniraidan perusteella on ongelmallista: vaikka joskus ääniraitaa voitaisiinkin käyttää objektin tunnistamiseen, suuren osan ajasta kuva ja ääni eivät vastaa toisiaan suoraan. On hyvin vaikea muodostaa luotettavaa mallia, jonka perusteella järjestelmä tietäisi, että nyt kuvasta tunnistettu alue tunnistetaan objektiksi ääniraidalta kuullun avainsanan avulla, kun suuren osan ajasta kuvan ja äänen yhteys on symbolinen tai indeksinen.

Kuvatekstit ja ääniraita ovat huomattavasti läheisemmässä suhteessa semanttiseen sisältöön kuin visuaalisen sisällön havaittavat piirteet yleensä. Kuvasta realistisesti tunnistettavissa olevia objekteja ovatkin uutisikkunan, uutisjuttujen ja -sähkeiden kuvatekstit, joita esimerkiksi Name-It käyttää kasvojen tunnistamiseen ja nimeämiseen ääniraidan lisäksi. Kuten luvussa 4 mainittiin, kuvatekstit havaitaan yli 90 %:n varmuudella ja tunnistetaan yli 70 %:n tarkkuudella. Kuvatekstit eivät tarjoa uutisikkunan lisäksi juurikaan tietoa uutisjutun aiheesta, mutta niistä on apua henkilöiden tunnistamisessa. Jos halutaan tukea semanttisia hakuja sekvensseihin niissä esiintyvien henkilöiden perusteella, on otettava huomioon, että (1) joskus ääniraita alkaa aiemmin kuin henkilö ilmestyy kuvaan (vrt. liite 8: O14–O15), (2) joskus sekä kuvateksti että ääniraita alkavat ennen kuin henkilö ilmestyy kuvaan (vrt. liite 8: O23–O24), ja joskus (3) kuvateksti ei liity kuvaan vaan ääniraidalla puhuvaan toimittajaan

(vrt. liite 8: O21). Henkilöiden tunnistaminen edellyttää, että kuvaraidalla puheen kanssa samaan aikaan esiintyviä objekteja vertaillaan toisiinsa ja johonkin malliin; periaatteessa kuvatekstin yhteydessä tai ääniraidalla esiintyvät nimet jätetään huomiotta, mikäli ne eivät liity kuvasta tunnistettuun ihmisen muotoiseen objektiin.

Liikkeen ja tapahtumien havaitseminen ja tunnistaminen määriteltiin erääksi indeksointitehtäväksi. Edellä mainittiin, että kameran liikkeen tunnistuksesta voi olla apua uutisstudioita koskevien otostyyppien eri versioiden erottelemisessa keskenään. Kameran liikkeen tunnistamista haittaavia liikkuvia objekteja ei uutisstudiossa ole. Varsinaisissa uutisjutuissa sen sijaan on liikkuvia objekteja, jotka saattavat kiinnostaa hakijoita. Käytännössä jonkin liikkeen tyypin tunnistaminen semanttisella tasolla edellyttäisi liikkuvan objektin ja kontekstin tunnistamista, mikä ei ole rajoittamattomissa konteksteissa realistista. Liikkeen suhteen on paljon todennäköisempää, että hakija käyttäisi määrittelemiään hahmoja ja niiden liikemalleja kyselyinä ja toivoisi, että ne vastaavat niitä objekteja ja objektien liikettä, joista hän on (semanttisella tasolla) kiinnostunut. On hyvin epätodennäköistä, että nykytietämyksen mukaan voitaisiin rakentaa järjestelmä, joka tunnistaisi esimerkiksi “junan” objektityyppinä. Käyttäjä voisi kuitenkin määritellä junanmuotoisen objektin esimerkiksi olettaen, että junat kuvataan yleensä sivulta, ja tälle hahmolle liikemallin; tämän jälkeen hän toivoisi, että tämä kysely tuottaisi todella kuvia junista eikä esimerkiksi maantiellä kuvatuista rekka-autoista. Koska on muitakin lähteitä, joista voi etsiä videokuvaa junista, kuin TV-uutiset, voidaan olettaa, että käyttäjä pelkkien junavideoiden sijaan on kiinnostunut vaikkapa VR:ää koskevista uutisaiheista, joissa yleensä esiintyy junia. Koska videotiedonhaku on usein sumeaa, ja tiedontarpeiden määrittely on vaikeaa ja niiden ilmaiseminen mielekkäällä tavalla hakujärjestelmälle vielä vaikeampaa, käyttäjä päätyy usein selailemaan tulosjoukkoa. Tästä syystä erääksi indeksointitehtäväksi asetettiin videodatan uudelleen esittäminen tiivistetyssä muodossa.

Videoinformaation visualisointia ja videotiedonhaunkäyttöliittymiä sekä -hakuvälineitä käsiteltiin luvussa 5. Aihehakua varten tarvitaan tekstuaalinen käyttöliittymä, jolla haut kohdistetaan uutisjuttujen transkriptioihin tai uutisikkunoista poimittuihin kuvaruututeksteihin. Järjestelmä voi toimia käyttökelpoisella tavalla, vaikka aihehaut eivät olisikaan tarkkoja: Otosten tarkkuudella analysoitu uutisjuttu oli pituudeltaan vain runsaat kaksi minuuttia. Mikäli sen jokaisesta otoksesta poimittaisiin yksi avainkehys, selattavaksi jäisi vain 17 avainkehystä, joiden avulla saisi jo riittävän kuvan uutisjutun visuaalisesta sisällöstä. Otoksista poimitut avainkuvat voitaisiin esittää ruudulla vaikkapa kronologisesti järjestettyinä ja näyttää niiden alapuolella kunkin otoksen kohdalla siihen liittyvä ääniraidan transkriptio. Ylimpänä tasona käyttöliittymässä voisivat olla aihehakuun vastaavat uutisjutut, joita edustettaisiin niiden toisen otoksen avainkehyksillä (ensimmäinen otostahan olisi uutisstudiosta). Näitä avain-

kehyksiä napauttamalla saisi tietyn uutisjutun kaikkien otosten avainkehykset kronologisesti järjestettyinä; napauttamalla otosten avainkehyksiä hiirellä järjestelmä toistaisi kyseisen otoksen. Uutisjuttujen ja -sähkeiden suhteellinen lyhyys tekee niiden selaamisesta täysin realistisen hakumenetelmän, kunhan vain ensin selailtavien nimekkeiden eli uutislähetyksen tapauksessa uutisjuttujen määrä saadaan tarpeeksi pieneksi.

Nykyiset järjestelmät sopivat huonosti videoiden visuaalisten sisältöjen semanttisiin hakuihin: automaattinen semanttinen päättely on liian kehittymätöntä esimerkiksi objektien tyyppien (esim. hiihtäjä) perusteella hakemiseen. Aihehakujen toteuttaminen on realistista, mikäli vain transkriptioiden tuottaminen saadaan riittävän luotettavaksi. Jos oletetaan, että uutisankkurin juonto on eräänlainen tiivistelmä uutisjutun sisällöstä ja että ankkureiden (ja toimittajien) määrä on rajallinen, voidaan puheentunnistusjärjestelmä optimoida suhteellisen pienelle joukolle puhujia. Nykyisillä menetelmillä on saavutettu parhaimmillaan 70 %:n tarkkuus uutislähetyksessä.

Uutislähetyksen automaattisen indeksoinnin toteuttamismahdollisuuksista voidaan esittää yhteenvetona kirjallisuuden ja YLE:n uutislähetyksen perusteella, että

1. uutisten ajallisen rakenteen jäsentäminen uutisjuttuihin ja -sähkeisiin on realistista toteuttaa nykytietämyksellä
2. uutisjuttujen aiheiden ja kasvojen nimeäminen kuvaruututekstejä tunnistamalla ja tekstuaalisten tiivistelmien laatiminen ääniraidan transkription avulla on mahdollista, vaikkakaan ei täysin ongelmatonta
3. epärealistista on sen sijaan yleisten objektien tunnistaminen ja nimeäminen uutisjuttujen rajoittamattomissa konteksteissa.

Tämän luvun tarkoitus oli soveltaa aiempaa tietämystä konkreettiseen uutislähetykseen. Esitettiin toteutettavissa olevat asiat ja ne asiat, joiden toteuttaminen on varsin epärealistista. Lisäksi pohdittiin indeksoinnissa kohdattavia ongelmia. Tarkkoja arvioita indeksointimenetelmien suorituskyvystä on kuitenkin vaikea tehdä ilman laboratorio- ja käyttäjätestejä. Loppujen lopuksi järjestelmän käyttökelpoisuus riippuu käyttäjästä ja siitä mitä järjestelmältä odotetaan. Seuraavassa luvussa esitetään koko tutkielmaa koskevat johtopäätökset.

7 KESKUSTELU JA JOHTOPÄÄTÖKSET

7.1 Johtopäätökset

Tutkielman tarkoitus on ollut selvittää, (1) mitkä ovat videoiden ja erityisesti TV-uutisten keskeisimmät ominaispiirteet indeksoinnin kannalta tarkasteltuna, (2) kuinka merkitys muodostuu havaittavien piirteiden pohjalta ja kuinka hahmopohjaiset indeksointimenetelmät pyrkivät mallintamaan tätä prosessia; lisäksi on pyritty selvittämään (3) videoiden visualisoinnin ja videotiedonhaun käyttöliittymien periaatteita eli sitä, miten alkuperäinen sekventiaalinen videodata esitetään uudelleen tiedonhakuun paremmin sopivassa muodossa. Näitä kohtia pohdittiin lisää YLE:n uutislähetyksen yhteydessä.

Johtopäätöksinä esitetään, että automaattiset indeksointijärjestelmät toimivat tällä hetkellä parhaiten videoiden ajallisen rakenteen jäsentämisessä. Otokset voidaan tunnistaa kehysten välisiä eroja etsimällä, ja otoksista voidaan muodostaa havaittavien piirteiden avulla ryhmiä vertailemalla niiden avainkehyksiä. Otoksia korkeammat tarinan yksiköt voidaan päätellä visuaalisten piirteiden ja otosten ajallista järjestämistä koskevan aihetietämyksen perusteella ilman, että videosisällöstä pitäisi yrittää tuottaa tulkintoja semanttisella tasolla. Aihetietämykseen kuuluu se, että tiedetään minkätyyppinen ja merkitykseltään minkäkinlainen otos seuraa missäkin vaiheessa. Luvussa 6 indeksointiprosessia käsiteltiin uutislähetyksen yhteydessä, jossa osoitettiin, kuinka uutislähetyksen rakenne voidaan jäsentää uutisjutut kehystävien uutisstudiota koskevien otostyyppien esiintymisen avulla. Vaihtoehtoisesti ajallinen rakenne voidaan jäsentää myös havaittujen objektien ja liikkeen perusteella, jolloin segmentti katsotaan alkaneeksi, kun kuvassa havaitaan objekti, ja se loppuu, kun objekti häviää. Sen sijaan objektien tunnistaminen semanttisella tasolla on realistisesti mahdollista lähinnä kasvojen ja kuvatekstien osalta. Automaattisten menetelmien rajoituksista kertoo se, että joissakin lähteissä annettiin ymmärtää, että videoiden indeksointi voi vaatia myös ihmisen väliintuloa.

Koska automaattisten indeksointijärjestelmien kyky tunnistaa sisältöjä on rajoittunut, käytännön hakutilanteissa samankaltaisuuden arvioiminen ja objektien tunnistaminen jäävät usein hakijan tehtäväksi, mikä pitäisi ottaa huomioon järjestelmiä suunniteltaessa: visualisointia pitäisi erityisesti kehittää ja selailua tukea hakustrategiana. Se, että hakija arvioi videosisältöä ei ole välttämättä huono asia, sillä hän on joka tapauksessa oman hakuongelmansa asiantuntija ja pystyy tekemään sisällöstä luotettavampia arvioita kuin mikään automaattinen järjestelmä. Automaattiset indeksointimenetelmät voivat tehdä sen, mikä inhimilli-

selle hakijalle tai indeksoijalle on vaikeaa: suurten videodokumenttimäärien jäsentämisen hallittavan kokoisiin yksiköihin ja niiden tuomisen saataville tiedonhakua tukevassa muodossa erilaisten korvikkeiden avulla. Objektien tunnistamisen ongelmat eivät siis välttämättä tarkoita, että nykyiset järjestelmät olisivat käyttökelvottomia videotiedonhaussa. Vaikka oletettaisiinkin, että kasvoista ei voitaisi tehdä semanttisia hakukohteita, joiden nimiä kyseltäisiin tekstuaalisesti, semanttiset haut voitaisiin kuitenkin kohdistaa uutisjuttujen transkriptioihin, ja halutut kohdat voitaisiin etsiä selailemalla. Vaihtoehtoisesti tutkielmassa on ehdotettu, että objekteja voitaisiin hakea niihin liittyvien havaittavien piirteiden avulla.

Kirjallisuudesta on havaittavissa, että tutkimuksessa on pyritty useampien erilaisten menetelmien integrointiin yksittäisten menetelmien suorituskyvyn parantamiseksi. Tämä käy ilmi puhetiedonhaun yhteydessä, jossa on yhdistetty foneemi- ja sanapohjaisia menetelmiä, ja erityisesti multimodaalisen kasvojen tunnistamisen kohdalla (esim. Name-It). Jos video on multimodaalinen viestintäväline, sitä kannattaa käyttää hyödyksi.

Kirjallisuuden keskeisimpiä ongelmia on epäjohdonmukainen terminologia: Esimerkiksi termillä olio ('entity') tarkoitetaan joskus vain havaittavaa aluetta ja joskus semanttisella tasolla tunnistettua objektia. Vastaavasti VIMSYS-mallin yhteydessä Del Bimbo [1999, 48] ja Lu [1999, 182–183] käyttävät kirjoittaessaan samasta asiasta eri termejä ('feature' ja 'entity'). Idris ja Panchanathan [1997, 159] käyttävät termiä ”representative frame”, kun muut käyttävät termiä ”keyframe”. Vastaavia esimerkkejä löytyisi lisääkin, varsinkin liittyen videon ajallisen hierarkian elementteihin (ks. luku 2.1): kehysten ('frame') ja otosten ('shot') kohdalla terminologia on yhdenmukaista eri lähteissä, mutta näiden käsitteiden yläpuolella variaatiota riittää, ja erityisesti kohtaus eli 'scene' vaikuttaa epämääräisesti käytetyltä termiltä.

Tutkielmassa esitettyihin viittauksiin ja arvioihin järjestelmien suorituskyvystä on suhtauduttava varauksin, sillä kirjallisuudessa raportoidut saanti- ja tarkkuusarvot eivät ole keskenään vertailukelpoisia, koska automaattisten indeksointimenetelmien suorituskyvyn arvioimiseen ei ole olemassa yhtenäistettyä testitietokantaa. Pelkän kirjallisuuden valossa ei ole mahdollista tehdä päteviä johtopäätöksiä menetelmien suorituskyvystä käytännössä. Tämän lisäksi on vältettävä tekemästä liiallisia johtopäätöksiä TV-toimituksen tarpeista luvun 5 pohjalta: vaikka Markkulan [2002] esiraportista käy ilmi, millaisia käytäntöjä toimituksella on tällä hetkellä, siitä ei voida vielä päätellä, miten toimitus haluaisi ja miten sen pitäisi toimia käytännössä. Tutkielma ei suoranaisesti käsittele TV-toimitusta, vaan se otetaan huomioon esimerkkinä aineiston rajaamisen ja jäsentämisen apuna.

7.2 Jatkotutkimus

Brunellin ja muiden [1999, 105–106] mukaan tutkimusala on pystynyt tuottamaan hyödyllisiä välineitä multimedia- ja videodatan käyttämiseksi, mutta avoimia ongelmia on vielä useita. Automaattinen segmentointi on joissakin sovelluksissa hyväksyttävän tarkkaa, mutta sitä voitaisiin parantaa edelleen huomattavasti yhdistelemällä useita eri algoritmeja, jotta varsinkin väärin tunnistettujen otosrajojen määrää saataisiin vähennettyä. Semanttisten merkitysten automaattinen päättely vaatii lisätutkimusta ja mahdollisesti parantaa järjestelmien käyttökelpoisuutta. Lisäksi käyttöliittymiä ja selaimia on kehitettävä lisää. Kaikki tämä vaatii poikkitieteellistä lähestymistapaa. [Brunelli et al. 1999, 105–106.] Käytetyt lähteet keskittyivät indeksointimenetelmien kuvailemiseen, mutta algoritmien tai järjestelmien suorituskykyä ei käsitelty yhtä kattavasti [vrt. Brunelli et al. 1999, 89–90; Antani et al 2002, 958–959]. Tähän on vaikuttanut *TREC*:n kaltaisen standardisoidun testikokoelman ja järjestelmien arviointia koskevan ja yleisesti hyväksytyn videoaineistolle sopivan metodologian puuttuminen. Geisler ja muut [2001] käsittelevät vapaata videokokoelmaa [18], jota voidaan käyttää tulevaisuudessa yhtenäisissä laboratoriotesteissä.

Järjestelmän suunnittelussa pitäisi ottaa huomioon virhetilanteet indeksoinnissa. Jos algoritmien tarkkuus ei ole täydellinen, on todennäköistä, että videoindeksiin tulee virheitä: esimerkiksi otosrajoja tunnistetaan sinne, missä niitä ei ole. Suunnittelussa olisi otettava huomioon virheiden mahdollinen kumuloituminen videohierarkiassa. Olisi tutkittava, kuinka esimerkiksi otosrajojen tunnistamisvirheet vaikuttavat koko hakujärjestelmän käytettävyyteen loppukäyttäjien näkökulmasta, ja kuinka otosrajojen tunnistamisvirheet vaikuttavat korkeamman tason tarinan yksikköjen muodostamiseen. Sisällön suhteen käyttökelpoinen järjestelmä ei vaadi täydellistä tunnistamistarkkuutta, sillä hyvällä käyttöliittymällä ja selailuominaisuuksilla voidaan korvata aiheen tunnistuksessa tapahtuneita virheitä, mutta jäsentämättä jäänyt otos voisi pahimmillaan jäädä pois tulosjoukosta, jos se on päätynyt väärään ryhmään. Näihin seikkoihin liittyen pitäisi tutkia, että millaiseen uutisjuttujen aiheiden ja objektien tunnistamistarkkuuteen käyttäjät olisivat tyytyväisiä ja millaisia käytännön ongelmia järjestelmien tunnistusvirheet aiheuttavat.

Dokumenttikeskeisessä videotiedonhaussa aiheen tunnistaminen on kriittinen kohta, mitä varten puheentunnistuksen tarkkuutta pitäisi kehittää. Uutislähetystä tarkemmin katsoessa (vrt. liite 8) ei voi olla kiinnittämättä huomiota siihen, kuinka pienen osan uutisjutun aiheen kannalta relevantista informaatiosta kuva välittää. Äänen keskeinen osuus erottaa uutisvideotiedonhaun monista muista visuaalisen tiedonhaun alueista: elokuvissa saattavat

18 The Open Video Project: <URL: http://www.open-video.org>

kiinnostaa visuaaliset piirteet, mutta uutisissa uutisankkurin, toimittajan ja haastateltavien puhe välittää suurimman osan siitä, mistä katsojat ovat kiinnostuneita. Kuvan ja tekstin yhdistäminen, silloin kun ne vastaavat suoraan toisiaan, on keskeinen ongelma tutkimuksessa: jos puhe vastaa suoraan kuvaa, puhetta voidaan käyttää kuvassa esiintyvien objektien tunnistamiseen; jos puhe ei vastaa suoraan kuvaa, vaan niiden välinen yhteys on vain uutisjutun sisäinen, puheen liittämistä kuvan objekteihin pitäisi välttää.

Kirjallisuuteen itseensä liittyy edellä mainittu terminologian kirjavuuden muodostama ongelma. Eräs lisätutkimuksen aihe voisi olla käsiteanalyysi: terminologian lainaaminen elokuvateoriasta ja semiotiikan teorioista auttaisi ainakin videoiden ajallisen hierarkian termistön yhdenmukaistamista.

Konenäön ja hahmontunnistuksen alojen kirjallisuus lähestyy aihepiiriä matemaattisesta näkökulmasta; piirteiden poimimista käsitellään kattavasti, mutta niiden kartoittamista semanttisiksi käsitteiksi ei ole käsitelty analysoidussa kirjallisuudessa kuin pinnallisesti. Uutisia koskevassa jatkotutkimuksessa voitaisiin kiinnittää enemmän huomiota mallien rakentamiseen ja niiden käsitteelliseen ja hierarkkiseen esittämiseen syvällisemmällä ja teoreettisemmalla tasolla. Uutislähetyksiä analysoimalla voitaisiin muodostaa ajallisia malleja, niiden välisiä siirtymiä koskevia malleja ja kertovan tilan objekteja koskevia malleja. Hierarkkia alkaisi yksittäisten kehysten sommittelusta, etenisi otosten tasolle ja käsittäisi fyysiset kontekstit (esimerkiksi uutisstudion) ja semanttiset kontekstit (kuuluko uutisstudiota koskeva otos uutisjutun juontoon, sähkeeseen vaiko lähetyksen puolivälin vinkkeihin). Lisäksi mallit käsittäisivät, kuinka otoksia linkitetään toisiinsa. Näiden mallien avulla esitettäisiin kaikki ne mahdolliset variaatiot, joilla uutislähetys voidaan koota. Mallien rakentamisen jälkeen voitaisiin pohtia konkreettisesti, mitkä algoritmit ja tekniikat sopisivat parhaiten indeksointijärjestelmän toteuttamiseksi. Tämä viimeksi mainittu ehdotus esitetään tutkielman päätteeksi jatkotutkimuksen kannalta suositeltavimpana ja realistisemmin toteutettavissa olevimpana vaihtoehtona.

LÄHTEET

Abberley, D., Kirby, D., Renals, S. & Robinson, T. 1999. The THISL Broadcast News Retrieval System. Saatavilla www-muodossa: <URL: http://svr-www.eng.cam.ac.uk/~ajr/esca99/Abberley.pdf>. (Luettu 3.5.2002.)

Ahenger, G. & Little, T. D. C. 1995. A Survey of Technologies for Parsing and Indexing Digital Video. Saatavilla www-muodossa: <URL:http://hulk.bu.edu/pubs/papers/1995/ahanger-jvcir95/TR-11-01-95.html>. (Luettu 9.4.2001.)

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.

* Arnheim, R. 1972. Art and Visual Perception. Berkeley and Los Angeles: University of California Press.

* Arnheim, R. 1982. The Power of the Center. Berkeley and Los Angeles: University of California Press.

Bolle, R. M., Yeo, B-L. & Yeung, M. M. 1998. Video query: Research directions. IBM Journal of Research and Development 42 (2). Saatavilla www-muodossa: <URL:http://www.research.ibm.com/journal/rd/422/bolle.html>. (Luettu 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.

Chang, Y., Zeng W., Kamel, I., Alonso, R. 1996. Integrated image and speech analysis for content-based video indexing. Saatavilla www-muodossa: <URL:

http://www.ee.princeton.edu/~wzeng/indexing.html>. (Luettu 10.4.2001.)

Corner, J. 1995. Television Form and Public Address. London: Edward Arnold.

Del Bimbo, A. 1999. Visual Information Retrieval. San Francisco, California: Morgan Khaufmann Publishers Inc.

Ellis, J. 1992. Visible fictions. Cinema, television, video. 2nd ed. London and New York: Routledge.

Feuer, J. 1992. Genre study and television. In R. C. Allen (Ed.) Channels of Discourse, Reassembled. 2nd ed. London: Routledge, 138–160.

Fiske, J. 1992. Merkkien kieli. Johdatus viestinnän tutkimiseen. 2. painos. Tampere: Vastapaino.

Gauch, S. 1998. The VISION Digital Video Library System. Saatavilla www-muodossa: <URL: http://www.ittc.ukans.edu/~sgauch/DVLS.html>. (Luettu 14.11.2002.)

Geisler, G. Marchionini, G., Nelson, M., Spinks, R. & Yang, M. 2001. Interface Concepts for the Open Video Project. Proceedings of the Annual Conference of the American Society for Information Science. November, 58–75.

Glavitsch, U. & Schäuble, P. 1992. A System for Retrieving Speech Documents. In Proceedings of SIGIR'92. New York: ACM, 168–176.

Gripsrud, J. 1995. The Dynasty Years. Hollywood Television and Critical Media Studies. London and New York: Routledge.

Grosky, W. I. 1997. Managing Multimedia Information in Database Systems. Communications of the ACM 40 (12), 73–80.

Gross, R., Shi, J. & Cohn, J. 2001. Quo vadis Face Recognition? Paper presented at Third Workshop on Empirical Evaluation Methods in Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition 2001. 9.–14.12.2001. Hawaii, USA.

Grossberg, L., Wartella, E. & Whitney, D.C. 1998. MediaMaking: Mass media in a popular culture. Thousand Oaks CA, London, New Delhi: Sage Publications.

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.

* Gupta, A., Weimouth, T. & Jain, R. 1991. Semantic queries with pictures: The VIMSYS model. Paper presented at the Proceeding of the 17th International Conference on Very Large Databases. 3.–6.9.1991. Barcelona.

Hall, S. 1999. Identiteetti. Tampere: Vastapaino.

Hietala, V. 1996. Ruudun hurma: johdatus TV-kulttuuriin. Jyväskylä: Gummerus.

Hjelsvold, R,. Lagorgen, S., Midtstraum, R. & Sandsta, O. 1995. Integrated video archive tools. In Proceedings of ACM International Conference on Multimedia '95. San Francisco, CA: ACM, 283–293.

Idris, F. & Panchanathan, S. 1997. Review of Image and Video Indexing Techniques. Journal of Visual Communication and Image Representation 8 (2), 146–166.

* Jain, R. & Hampapur, A. 1994. Metadata in video databases. SIGMOD Record 23 (4). Saatavilla www-muodossa: <URL: http://www.acm.org/sigmod/record/issues/9412/jain.ps>. (Luettu 14.12.2002.)

Kender, J. R. & Yeo, B-L. 1998. Video Scene Segmentation Via Continuous Video Coherence. Oral presentation. Saatavilla www-muodossa: <URL: http://www.cs.columbia.edu/~jrk/research/video-coherence.ps>. (Luettu 28.2.2003.)

Kress, G. & van Leeuwen, T. 1996. Reading Images. The Grammar of visual design. London

and New York: Routledge.

Leavers, V. F. & Burley, C. E. 2001. The use of cognitive processing strategies and linguistic cues for efficient automatic language identification. Language Sciences 23, 639–650.

Lee, H. & Smeaton, A. F. 1999. User-interface Issues for Browsing Digital Video. Paper presented at the 21st Annual Colloquium on IR Research (IRSG 99). 19.–20.4.1999. Glasgow, UK. Saatavilla www-muodossa: <URL: http://www.cdvp.dcu.ie/Papers/IRSG99.pdf>. (Luettu 12.11.2002.)

Lee, H. & Smeaton, A. F. 2002. Designing the User Interface for the Físchlár Digital Video Library. Journal of Digital Information 2 (4). Saatavilla www-muodossa: <URL: http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Lee/>. (Luettu 4.12.2002.)

Lienhart, R., Pfeiffer, S. & Effelsberg, W. 1997. Video Abstracting. Communications of the ACM 40 (12), 55–62.

Lu, G. 1999. Design issues of multimedia information indexing and retrieval systems. Journal of Network and Computer Applications 1999 (22), 175–198.

Markkula, M. 2002. Methods and techniques for next generation information retrieval and management of digital resources: Video IR. [Tutkimuksen esiraportti.]

Markkula, M. & Sormunen, E. 2000. End-User Searching Challenges Indexing Practises in the Digital Newspaper Photo Archive. Information Retrieval 1, 259–285.

Markkula, M., Tico, M., Sepponen, B., Nirkkonen, K. & Sormunen, E. 2001. A Test Collection for the Evaluation of Content-Based Image Retrieval Algorithms – A User and Task-Based Approach. Information Retrieval 4 (3/4), 275–294.

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.

Petković, 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.

Ponceleon, D. B. & Srinivasan, S. 2002. Tutorial on Practical Guidelines for Multimedia Information Retrieval. Tutorial held at the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 11.8.2002. Tampere, Finland.

Prabhakaran, B. 1997. Multimedia Database Management Systems. Boston, London and Dordrecht: Kluwer Academic Publishers.

Rasmussen, E. M. 1997. Indexing Images. In M.E. Williams (Ed.) Annual Review of Information Science and Technology 32. New Jersey: Information Today, 169–196.

Roth, V. 1998. Content-Based Retrieval from Digital Video. Saatavilla www-muodossa: <URL:http://www.igd.fhg.de/igd-a8/publications/OtherSubjects/98_Content-BasedRetrievalFromDigitalVideo.html>. (Luettu 8.4.2001.)

Rui, Y., Huang, T. S. & Mehrotra, S. 1999. Constructing table-of-content for videos. Multimedia Systems 7, 359–368.

Satoh, S., Nakamura, Y. & Kanade, T. 1999. Name-It: Naming and Detecting Faces in News Videos. IEEE Multimedia 6 (1), 22–35.

Seiter, E. 1992. Semiotics, structuralism and television. In R. C. Allen (Ed.) Channels of Discourse, Reassembled. 2nd ed. London: Routledge, 31–66 .

Sheridan, P., Wechsler, M. & Schäuble, P. 1997. Cross-Language Speech Retrieval: Establishing a Baseline Performance. In Proceedings of SIGIR-1997. New York: ACM, 99–-108.

Shneiderman, B. 1998. Designing the user interface: Strategies for effective human computer

interaction. 3rd ed. Reading, MA: Addison-Wesley.

Smith, J. R. 2001. Quantitative Assessment of Image Retrieval Effectiveness. Journal of the American Society for Information Science and Technology 52 (11), 969–979.

Sormunen, E., Markkula, M & Järvelin, K. 1999. The Perceived Similarity of Photos – A Test-Collection Based Evaluation Framework for the Content-Based Image Retrieval Algorithms. In Draper S. et al. (Eds.) Mira 99: Evaluating interactive information retrieval. Electronic Workshops in Computing (eWic).

Srinivasan, S. & Petkovic, D. 2000. Phonetic Confusion Matrix Based Spoken Document Retrieval. In Proceedings of SIGIR-2000. New York: ACM, 81–87.

* Swanberg, D., Shu, C. F. & Jain, R. 1993. Knowledge guided parsing in video databases. In Proceedings of SPIE Electronic Imaging: Science and Technology. San Jose, 13–24.

* Tonomura, Y., Akutsu, A., Taniguchi, Y. & Suzuki, G. 1994. Structured video computing. IEEE Multimedia 1 (3), 34–43.

Whittaker, S., Hirschberg, J., Choi, J., Hindle, D., Pereira, F. & Singhal, A. 1999. SCAN: Designing and evaluating user interfaces to support retrieval from speech archives. In Proceedings of SIGIR'99. New York: ACM, 26–33.

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.

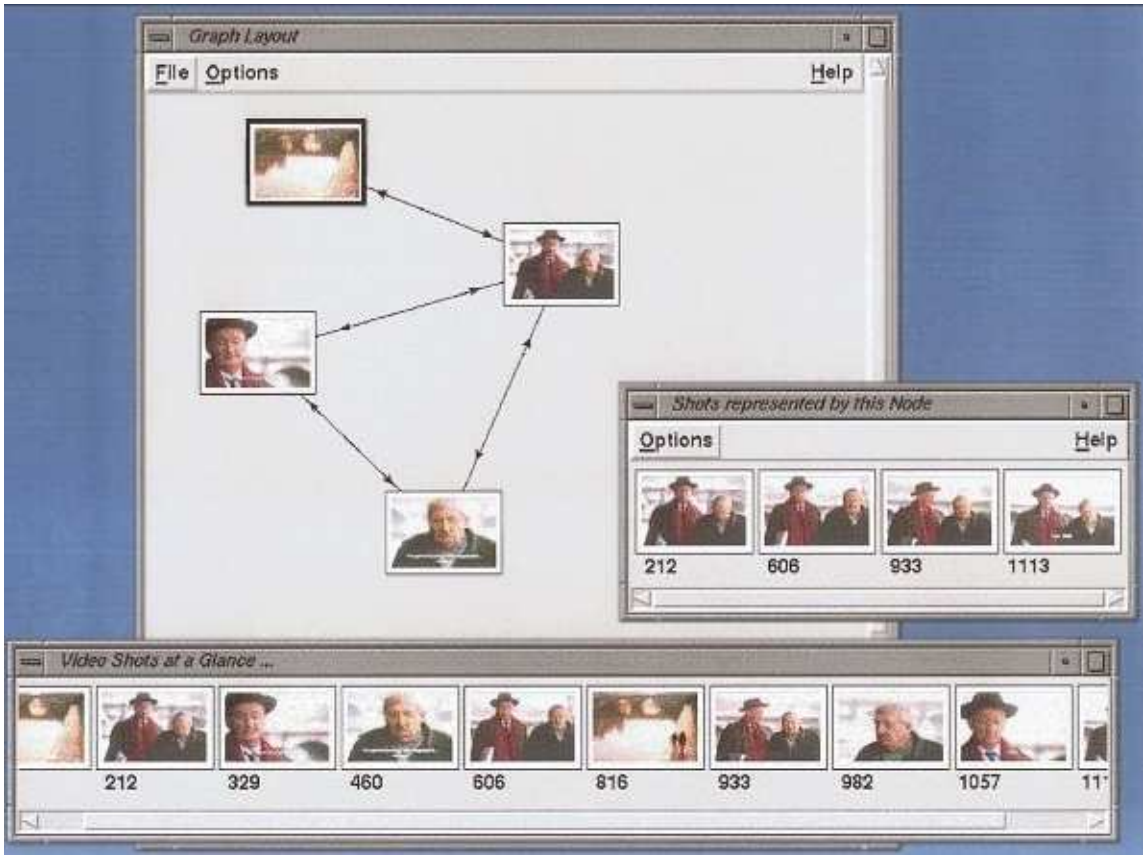
* Zettl, H. 1984. Television Production Handbook. Belmont, California: Wadsworth Publishing Company.

Zissman, M. A. & Berkling, K. M. 2001. Automatic language identification. Speech Communication 35 (2001), 115–124.

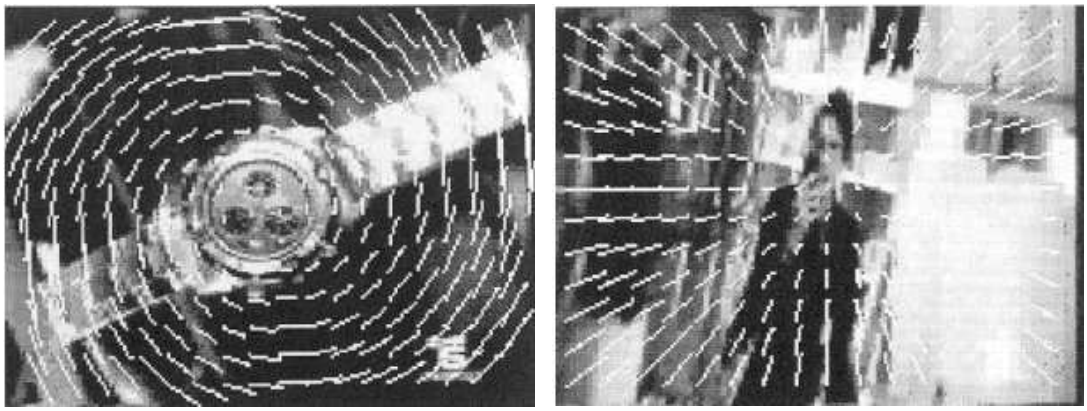
(Tähdellä (*) merkittyihin on viitattu toisen lähteen kautta.)

LIITTEET

Liite 1. Kohtauksensiirtymiskaavio, jossa on neljä solmua [Bolle et al. 1998]. Solmussa näkyvät samaan ryhmään kuuluvat otokset. Alimmassa ikkunassa näkyvät videon kaikkien otosten avainkehykset; 212, 606, 933 ja 1113 ovat avainkehyksien numeroita. Janat solmujen välillä ovat linkkejä: ylimmäisellä solmulla on yhteys siitä alaoikeaan sijaitsevaan solmuun, mutta ei alapuolellaan olevaan solmuun. Tämä voidaan varmistaa alimmaisesta ikkunasta: kaavion ylimmäinen otos ei ole missään vaiheessa vierekkäin yksittäisistä miehistä otettujen otosten kanssa, joten niillä ei ole yhteyttä.

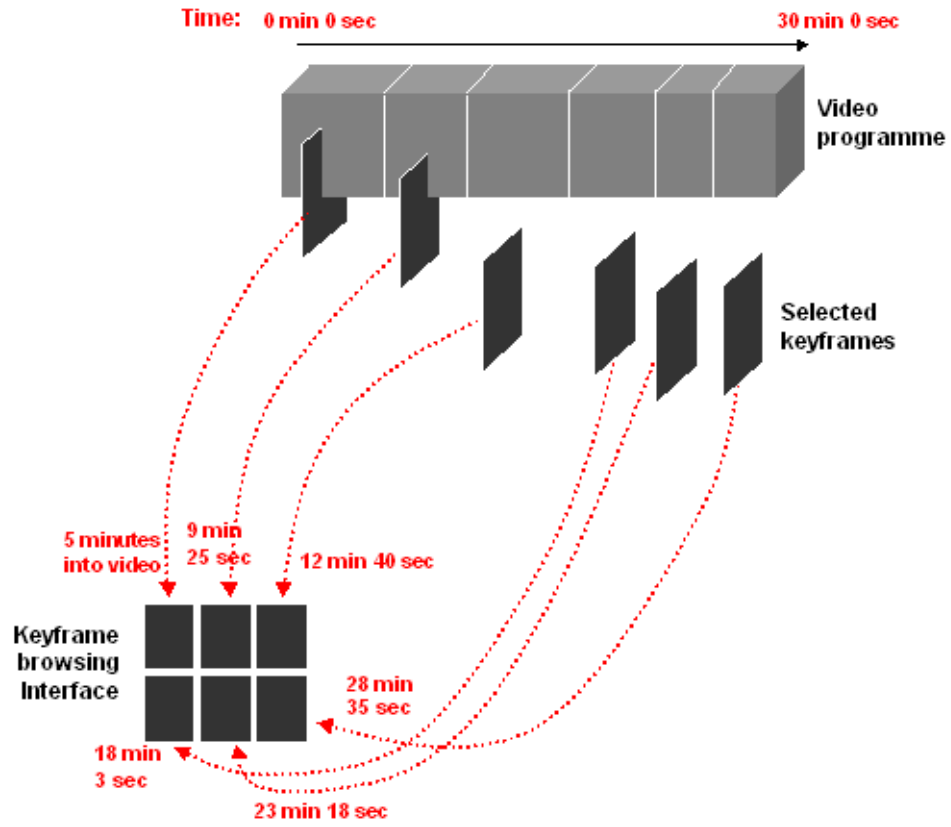


Liite 2. Pyörivän objektin ja zoomaavan kameran liikekentät [Del Bimbo 1999, 235].



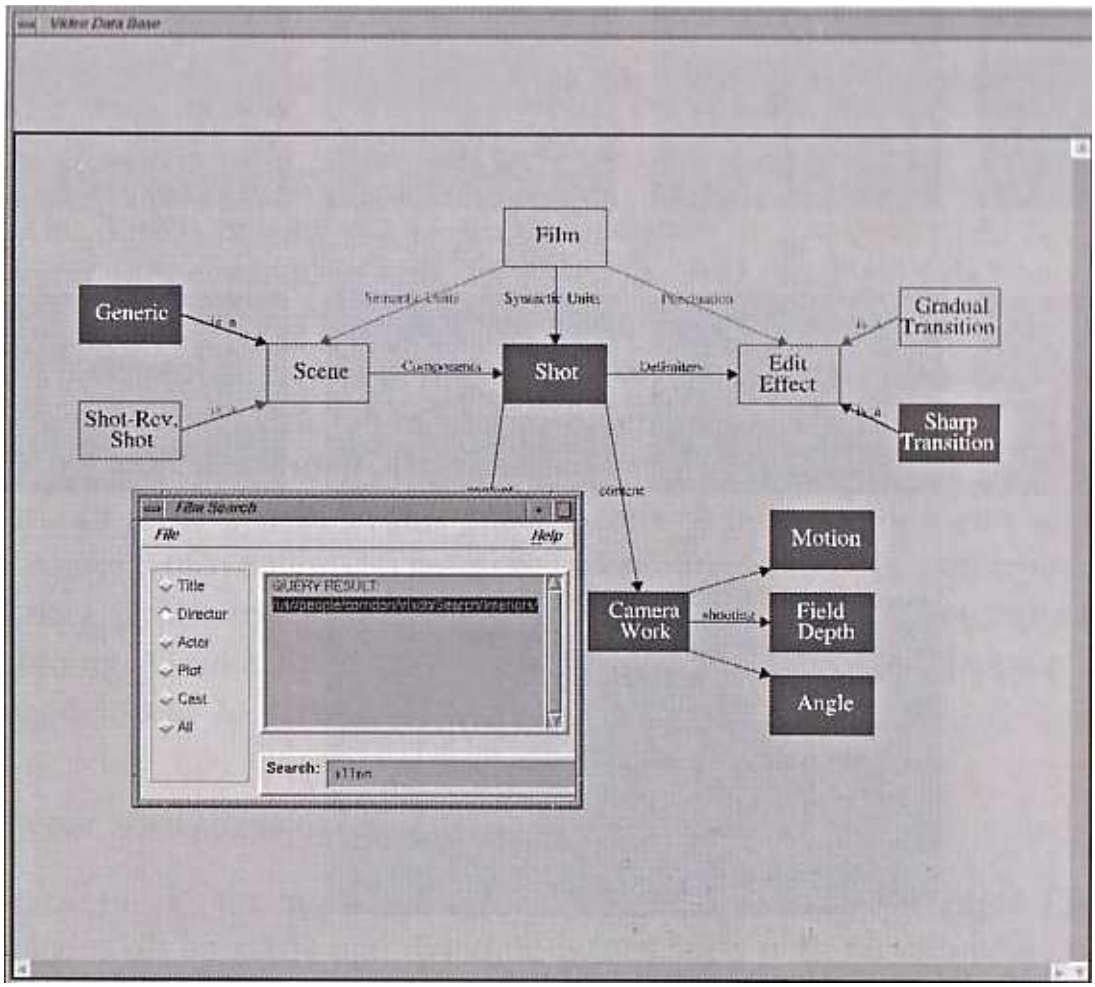
Liite 3. Mosaiikki [Del Bimbo 1999, 17].

Liite 4. Avainkehyksien valitseminen videosta [Lee & Smeaton 2002].



Liite 5. Microsoft Mediaplayer-ohjelman sekventiaalinen selain [Lee & Smeaton 2002].

Liite 6. Videon rakenteen esittäminen graafisen kaavion avulla [Del Bimbo 1999, 250].



Liite 7. Hakuvälineiden ominaisuuksien arviointitaulukko.

Ominaisuus	***VISION***	***VIDEOS-TAR***	***OLIVE***	***VideoQ***	***Netra-V***	***Screening Room***	***Videolog-ger***
1. Luonnollisen kielen kyselyt	2	2	2	2	0	2	2
2. Ääniraidan indeksointi ja haut	2	-2	2	-2	-2	2	2
3. Ääniraidan transkriptio	-2	-2	2	-2	-2	-2	-2
4. Histogrammin käsittely	0	0	0	0	0	0	0
5. Avainkehyspohjainen hahmon piirtäminen	0	0	0	0	0	0	0
6. Avainkehyspohjainen kysely esimerkkikuvalla	0	0	0	0	0	0	0

Ominaisuus	*VISION*	*VIDEOS-TAR*	*OLIVE*	*VideoQ*	*Netra-V*	*Screening Room*	*Videolog-ger*
7. Liikepohjainen hahmon piirtäminen	0	0	0	2	0	0	0
8. Liikepohjainen kysely esimerkillä	0	0	0	2	2	0	0
9. Avainsana ja kategorialuettelot	2	0	2	2	0	0	2
10. Tekstuaalinen kuvailu	2	2	0	2	0	2	2
11. Transkriptio	0	0	2	0	0	2	2
12. Yksittäinen avainkehys	2	-2	2	2	2	2	2
13. Kronologinen avainkehysluettelo	0	0	2	0	0	2	2
14. Eri tasoilla tiivistetty avainkehysluettelo	0	0	0	0	0	2	0
15. Vuorovaikutteinen hierarkkinen avainkehysselain	0	0	0	0	0	0	0
16. Ajastettu avainkehyksen toisto	0	0	0	0	0	0	0
17. Kohokohtien toisto	0	0	0	0	0	0	0
18. Nimekkeiden (yleinen) toisto	2	2	2	2	2	2	2
19. Transkriptio + toisto synkronisoituna	0	0	2	0	0	0	0
20. Avainkehys + toisto synkronisoituna	0	0	0	0	0	0	2
21. Tekstuaalinen haku + toisto synkronisoituna	0	2	0	0	0	0	2
22. Avainkehyksien älykäs valitseminen	0	0	0	0	0	0	0
YHTEENSÄ	10	4	18	10	2	14	18

Liite 8. Sisällönanalyysi YLE:n TV1:n klo 20:30 uutislähetyksestä 29.11.2002. Otokset on jäsennetty videonauhuria käyttäen sekunnin tarkkuudella, joten lopputulos siis voi olla täysin tarkka. Toisessa sarakkeessa “R” tarkoittaa otosryhmää ja “O” otosta (eli segmenttiä). Kuvasarakkeessa videoraitaa kommentoidaan seuraavassa järjestyksessä (niiltä osin, kun se on tarkoituksenmukaista ja sisältö ei ole toisteista):

1. Mainitaan otoksessa esiintyvät (a) *keskeiset objektit* (ne, joilla on aiheen kannalta semanttista merkitystä) eli lähinnä kasvot ja kuvatekstit, (b) *silmiinpistävien objektien si-*

jainnit eli sommittelu (otoksen alussa), (c) *kuvakulma ja etäisyys* (läheltä, normaali tai kaukaa), (d) *hallitsevat värisävyt* ja (e) *kuvauspaikka eli konteksti*, johon otos kuuluu.

2. Liikkeestä mainitaan, että onko otos (a) *staattinen* vaiko *dynaaminen* sekä (b) *liikkeen tyyppi* (kamera vai objektit), (c) *suunta* ja mahdollisesti kuvaillaan (d) *keskeisten objektien liiketapahtumia* (esimerkiksi “auto ajaa ohi” tai “uutisankkuri puhuu”).
3. Mainitaan segmenttien (otosten) välillä käytetyn *siirtymätehosteen tyyppi*, välitön vai asteittainen (häivytys, liuotus, pyyhkäisy tai himmennys).

Äänisarakkeessa on mainittu puhuja (myös kieli olisi mainittu, jos siihen olisi ollut tarvetta) ja puhe on litteroitu. Analyysissa käsitellään otosten tasolla vain alkurituaalia ja ensimmäistä varsinaista uutisaihetta eli tarinaa. Lähetyksen muita uutisaihetta (A2–) varten on kuitenkin oma rivinsä (“A”), ja niistä mainitaan vain juttujen pituus, niiden aihe ja tyyppi (varsinainen uutisjuttu tai -sähke).

Aika	***R/O***	***Kuva***	***Ääni***
0:00 – 0:05	1/1	Uutistunnus:	- -
0:05 – 0:09	1/2	Uutistunnus (2):	- -
0:09 – 0:12	2/3	Uutisstudio: - Asteittainen (liuotus)	[Ankkuri]: Hyvää iltaa.
0:12 – 0:15	3/4	Uutisvinkit: - Lääkeruisku, käsi, “uutiset” - Edestä, läheltä, ihon väri - Kamera: staattinen; objektit: pyörivä laite taustalla - Välitön	Uudet doping-väitteet herättävät ihmetystä. Kansainvälisen hiihtoliiton veritestejä
0:15 – 0:19	3/5	- Veripulloteline, “uutiset” - Etuviisto, läheltä; valkoinen, metalli, puu - Kamera: staattinen; pyörivä teline - Asteittainen (pyyhkäisy)	epäillään peukaloidun.
0:19 – 0:25	3/6	- Öljytankkeri (oikealla puolella), kumilautta, laituri, “uutiset”, “Byzantio” - Edestä, normaali; musta, punaruskea, vesi, harmaa. - Kamera: zoom; objektit: kumilautta kelluu aalloilla - Asteittainen (pyyhkäisy)	Greenpeace meni Tallinnaan estämään kiistellyn öljytankkerin lähtemistä satamasta.

Aika	*R/O*	*Kuva*	*Ääni*
0:25 – 0:32	3/7	- Lentokone, “uutiset” - Sivulta, kaukaa,; valkoinen, harmaa - Kamera: jäljitys; objektit: lentokone nousee oikealta vasemmalle - Asteittainen (pyyhkäisy)	Suomen lentomarkkinoille on tulijoita. Vireillä on viisi lentoyhtiöhanketta.
0:32 – 0:38	3/8	- Lapsi (keskellä kuvaa), kirja, “uutiset” - Edestä, normaali; vaalea, punainen, vakoinen, keltainen - Kamera ja objektit staattisia - Asteittainen (pyyhkäisy)	Suomalaiset lapsiperheet ovat köyhtyneet kymmenen viime vuoden aikana.
0:38 – 0:45	3/9	- Ei selviä objekteja, ihmisiä, “uutiset” - Normaali kuvausetäisyys; ei selviä värejä, punainen paita - Kamera: siirto oikealle; objektit staattisia - Asteittainen (liuotus)	Yleisradion radiokanavat hiovat ilmettään. Nimet muuttuvat ja ohjelmisto uudistuu tammikuussa.
0:45 – 1:03	4/10	Uutisstudio: - Uutisikkunassa “Doping” ja koeputkia, alareunassa “Arvi Lind”. - Välitön	[Ankkuri/Suomi]: Maastohiihto on jälleen joutunut doping-huhujen kohteeksi. Tällä kertaa väärinkäytöksistä epäillään kansainvälistä hiihtoliittoa. Tämän päivän Helsingin sanomat väittää, että FIS olisi laimentanut hiihtäjien verinäytteitä omissa testeissään Lahden MM-kisoissa viime vuonna.
1:03 – 1:14	5/11	Uutisjuttu: - Hiihtäjä, kuvateksteissä mm. “Harri Kirvesniemi”, valkoinen teksti sinistä taustaa vasten; (kuvattu) edestä; valkoista, sinistä, tummaa ja vihreää; kuvattu hiihtoladulla metsässä - Kamera: jäljitys; objekti (hiihtäjä) liikkuu suoraan edestä ohi oikealle - Asteittainen (liuotus)	[Toimittaja/suomi]: Lahden MM-kisoissa suomalaisen maastohiihdon kärki kärysi dopingista. Siitä asti ainakin tiedotusvälineissä on herätetty epäilyjä, että kaikkea ei ole vielä kerrottu.
1:14 – 1:25	6/12	- Kuvassa teksti “Helsingin Sanomat” ja “FIS:n epäillään laimentaneen verinäytteitä Lahden MM-hiihdoissa” (kuvattu hieman vinossa suoraan lehdistä), läheltä; mustaa, valkoista ja turkoosia - Objektit ja kamera staattisia - Asteittainen (liuotus)	Tämän päivän Helsingin Sanomat väittää, että FIS olisi laimentanut kisoissa ottamiaan ottamiaan verinäytteitä ja näin saanut hiihtäjien hemoglobiiniarvot näyttämään alhaisemmilta. Sillä olisi peitelty EPO-hormonin käyttöä.
1:25 – 1:35	7/13	- Kuvassa kaksi kävelevää ihmistä, hiihtäjä ja valmentaja; kuvattu takaa ylhäältä; normaali etäisyys, hiihtäjällä sini-valkoinen asu, valkoinen lumi. - Kamera (jäljittää ja zoomaa); objekteja, (jotka kävelevät) - Välitön	EPO lisää veren punasolujen tuotantoa ja sitä kautta lisää suorituskykyä. Epäilyt eivät koske varsinaisia doping-testejä. Ne teki Lahdessa Suomen anti-doping toimikunta

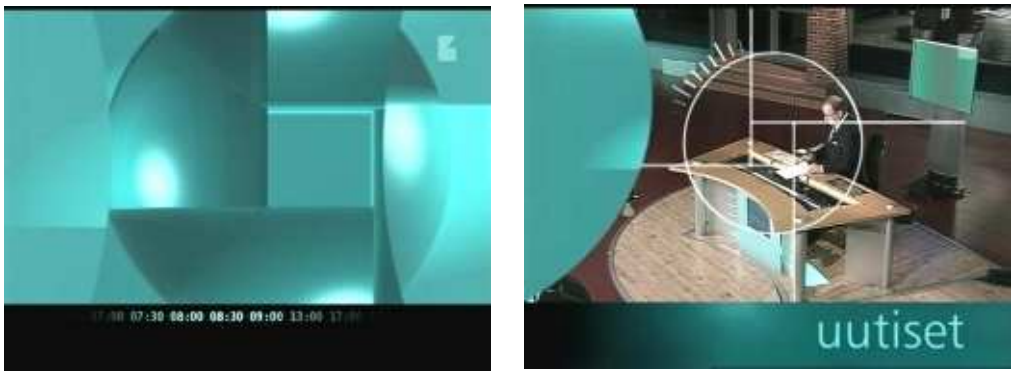
Aika	*R/O*	*Kuva*	*Ääni*
1:35 – 1:39	7/14	- Hiihtäjä ja valmentaja kumartunut maahan; kuvattu takaviistosta, läheltä; sinistä, mustaa ja valkoista - Kamera ja objektit staattisia. - Asteittainen (liuotus)	lääkäri Heikki Laakion johdolla. [Heikki Laapio]: Doping-testejä
1:39 – 1:47	8/15	- Kuvateksti "lääkintäeversti Heikki Laapio" mustalla tekstillä vaaleaa ja turkoosia taustaa vasten; Laapio kuvattu lähes suoraan edestä, läheltä; sisätiloissa, pysäytyskuvassa - Kamera ja objektit staattisia - Asteittainen (liuotus)	siellä teki virallisesti tää kisaorganisaatio eli me ja sitten WADA, joka tuli sinne tän suomalaisten hässäkän takia.
1:47 – 1:50	9/16	- Kuvassa käsiä ja lääkeruisku; kuvattu sivulta, läheltä; ihon väri, tumma, vaalea - Kamera staattinen, kuvassa pyörivää liikettä - Välitön	[Toimittaja]: Lisäksi kansainvälinen hiihtoliitto FIS teki
1:50 – 1:53	10/17	- Kuvassa pyörivä koeputkiteline; kuvattu etuviistosta - Kamera staattinen, objekti pyörii - Välitön	kisoissa omia veritestejään. Niillä
1:53 – 1:57	10/18	- Kuvassa laite; kuvattu etuviistosta, läheltä; harmaan sävyjä - Kamera staattinen, vasemmassa yläreunassa liikettä otoksen loppupuolella (joku käyttää konetta) - Välitön	mitattiin hemoglobiiniarvoja. Verta kerättiin myös
1:57 – 2:04	11/19	- Kuvassa laitteen näyttöpaneeli, jossa lukee "RINSING" ja "NO. 26"; kuvattu läheltä, teksti vinossa. - Kamera (panoroi ylhäältä alas), objekti staattinen - Asteittainen (liuotus)	FIS:n oman EPO-testin kehittämiseen. Ja vain FIS tietää tarkalleen mitä näytteitä on otettu ja miten.
2:04 – 2:09	12/20	- Kuvassa taas Laapio - Asteittainen (liuotus)	[Laapio:] Ne teki viereisessä huoneessa ja me ei yksityiskohtaisesti nähty miten ne tehtiin.
2:09 – 2:20	13/21	- Kuvassa hiihtäjä hiihtoladulla ja kuvatekstejä mm. "Finland" ja myöhemmin toimittaja "Maria Stenroos"; kuvattu etuviistosta (kohde liikkuu), läheltä. - Kamera (jäljittää), hiihtäjä (liikkuu etuvasemmalta oikealle) - Välitön	[Toimittaja]: Niin kauan kuin FIS ei kerro koko totuutta testeistä, voidaan vain arvailla mitä on tapahtunut. Suomen antidoping-toimikunnan Timo Seppälä
2:20 – 2:23	13/22	- Ks. Edellinen. Ei toimittajan nimeä. - Hiihtäjä (liikkuu vasemmalta oikealle) - Välitön	pitää kuitenkin lantrausväitteitä epäuskottavina.

Aika	*R/O*	*Kuva*	*Ääni*
2:23 – 2:25	13/23	- Ks. Edellinen. Ruudussa teksti “Timo Seppälä” - Välitön	[Seppälä]: Minä olettaisin, että
2:25 – 2:37	14/24	- Ruudussa tekstit “lääketieteellinen johtaja”, “Timo Seppälä”, “Suomen antidoping-toimikunta”; kuvassa hieman etuviistosta kuvattu Seppälä vihreässä takissa, kuvausetäisyys normaali; paljon huonosti erottuvia värisävyjä; kasvot tunnistettavissa? - Kamera ja kuva staattisia - Välitön	FIS ei ole ainakaan tällä tavalla manipuloinut, kun on ollut esillä, näitä näytteitä. Jos FIS olisi halunnut manipuloida näytteitä, niin olishan ollut paljon helpompi tapa esimerkiksi kalibroida
2:37 – 2:46	15/25	- Kuvassa paljon ihmisiä hiihtostadionilla, vaikea erottaa toisistaan; etäisyys läheltä/normaali; tummaa, vaaleaa, sinistä. - Kamera (panoroi vasemmalta oikealle) - Välitön	nämä näytelaitteet näyttämään väärin. [Toimittaja]: Jos näytteet ovat laimentuneet, se on Seppälän mukaan voinut johtua FIS:n huonosta näytteidenotosta tai
2:46 – 2:55	15/26	- Ihmismassa läheltä, vaikeasti erotettavissa; etuvasemmalta; mustaa sinistä, valkoista, tummanvihreää ja keltaista - Kamera (panoroi oikealta vasemmalle), objektit (taputtavat) - Välitön	laboratoriotekniikasta. Selvyys asiaan saatanee vain, jos FIS julkaisee testi menetelmänsä ja hiihtäjien veriarvot.
2:55 – 3:00	16/27	Uutisstudio: - Doping-uutisikkuna edelleen. - Asteittainen (liuotus)	[Ankkuri]: Lisää asiasta illan urheiluruudussa.
		(Seuraava aihe eli uutisähke.)	

Aika	*A*	*Uutisaihe*	*Kommentit*
3:00 – 3:32	2	Sonera n toimitusjohtajan vangitseminen	Uutissähke
3:32 – 5:48	3	Greenpeace ja Byzantio-öljytankkeri	
5:48 – 6:16	4	Kenia ja Mombasan terrori-isku	Uutissähke
6:17 – 8:37	5	Kilpailu kotimaan lentomarkkinoilla	
8:37 – 9:13	6	Työsopimukset	Uutissähke
9:13 – 9:40	7	Välikysymys ja harmaan talouden torjuminen.	Uutissähke
9:40 – 12:13	8	Valtion tilintarkastajat moittivat puolustusvoimia.	
12:13 – 12:39	9	Liikenneturma johtaa syytteisiin	Uutissähke
12:39 – 12:56	10	Tässä lähetyksessä kerromme vielä: Lapsiperheet köyhtyvät Radiouudistus Pakkanen ja Rukan maailmancup	
12:56 – 14:43	11	Taloussuhdanteet	
14:43 – 16:57	12	Lapsiperheet ja köyhyys.	

Aika	*A*	*Uutisaihe*	*Kommentit*
16:57 – 19:20	13	Ruotsi ja EMU	
19:20 – 19:48	14	Tupakoinnin kieltäminen Norjassa julkisilla paikoilla	Uutissähke
19:48 – 22:18	15	Radiouudistus	
22:18 – 24:23	16	Säätiedotus	
24:23 – 24:33	17	Lopetuspuhe	
24:33 – 24:43	18	Lopetustunnus	

Liite 9. Uutistunnus [Yleisradio, TV1. 1.2.2003].



Liite 10. Uutisstudio uutistunnuksen jälkeen, kaukaa ja tarkennuksen jälkeen normaalietäisyydeltä [Yleisradio, TV1. 1.2.2003].

Liite 11. Loppukommentit ja lopetustunnus [Yleisradio, TV1. 1.2.2003].

Liite 12. Uutisvinkki ja siirtymä [Yleisradio, TV1. 1.2.2003].

Liite 13. Uutisjutun alku ja uutisikkuna sekä -sähke [Yleisradio, TV1. 1.2.2003].