

Automaattinen ilmeidenluokittelu liikkuvasta kuvasta

Ilari Nieminen

Helsinki 6.5 2005

Tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section Matemaattisluonnontieteellinen		Laitos – Institution – Department Tietojenkäsittelytieteen laitos	
Tekijä – Författare – Author Ilari Nieminen			
Työn nimi – Arbetets titel – Title Automaattinen ilmeidenluokittelu liikkuvasta kuvasta			
Oppiaine – Läroämne – Subject Tietojenkäsittelytiede			
Työn laji – Arbetets art – Level Tutkielma		Aika – Datum – Month and year 6.5 2005	Sivumäärä – Sidoantal – Number of pages 20 sivua + 2 liitesivua
Tiivistelmä – Referat – Abstract <p>Selvitämme, miten ihmiskasvojen ilmeitä voidaan luokitella automaattisesti tietokoneen avulla liikkuvasta kuvasta. Kognitiotieteen parissa syntynyt, formaali kasvojen toimintojen koodaustapa FACS (Facial Action Coding System) esitellään perustana automaattisen ilmeidenluokittelun menetelmille ja sovelluksille.</p> <p>ACM Computing Classification System (CSS): A.1 [Introductory and Survey], I.4 [Image Processing and Computer Vision], I.5 [Pattern Recognition], J.4 [Social and Behavioral Sciences]</p>			
Avainsanat – Nyckelord – Keywords Ilmeiden luokittelu, FACS, hahmontunnistus, kuvankäsittely			
Säilytyspaikka – Förvaringställe – Where deposited			
Muita tietoja – Övriga uppgifter – Additional information			

Sisältö

1 Johdanto	1
2 Formaali kasvontoimintojen koodaustapa	2
2.1 Koodausjärjestelmän kuvaus	2
2.2 Ilmeiden esittäminen FACS-koodien avulla	5
2.3 FACS-koodien haitat automaattisessa ilmeidenluokittelussa.	8
3 Ilmeidenluokittelu kuvasarjasta	8
3.1 Ilmeidenluokittelujärjestelmän kuvaus	9
3.2 Kuvansarjan normalisointi	10
4 FACS-koodien automaattinen muodostaminen kuvasta	13
4.1 Merkinnot	13
4.2 Optinen virtaus	13
4.3 Pääkomponenttianalyysi	17
5 Yhteenveto	20
Lähteet	21
Liite A: Esimerkki toimintayksikön määrittelystä	
Liite B: Lineaarialgebraan liittyviä käsitteitä	

1. Johdanto

Kasvojen ilmeet välittävät tietoa ihmisen tunnetilasta, toiminnan tilasta, luonteesta, temperamenttisuudesta, totuudenmukaisuudesta ja monista muista ominaisuuksista [Don99]. Käytämme jatkuvasti kasvojamme antamaan lisämerkitystä puheelle ja muulle kanssakäymiselle [EcF78].

Facial Actions Coding System (FACS) [EcF78] on käyttäytymistieteellisissä tutkimuksissa käytetyin menetelmä ihmisen kasvontoimintojen mittaamisessa ja analysoinnissa. Kyseessä on kasvontoimintojen koodaustapa, jolla kaikki tietyllä tarkkuudella havaittavissa olevat muutokset kasvojen ilmiössä voidaan havaita ja koodata systemaattisesti [EcF78].

Toistaiseksi FACS-koodausjärjestelmää käyttävät ainoastaan huolellisesti perehdytetyt ihmisasiantuntijat. Kuvankäsittelyn ja koneoppimisen menetelmät tarjoavat kuitenkin yhä parempia menetelmiä FACS-koodausjärjestelmän käytön automatisoimiseksi. Pätevän ihmisasiantuntijan kouluttaminen järjestelmään vie keskimäärin sata tuntia ja minuutin pituisen videonäytteen koodaaminen vaatii asiantuntijalta keskimäärin tunnin [Don99], joten FACS-koodausjärjestelmän automatisoinnille on ilmeistä kysyntää kognitiivisen tutkimuksen parissa.

Automaattinen ilmeidenluokittelu tietokoneen avulla on kokonaisuudessaan haastava tutkimusalue. Tehtävänä on havaita ja luokitella kaikki kasvojen ilmeet annetusta kuvasarjasta, käyttäen apuna tietokantaa erilaisten ilmeiden kuvauksista. Tehtävään liittyy useita osatehtäviä, kuten kuvan esikäsittely ja kasvojen paikantaminen kuvasta.

Luotettavaa automaattista ilmeidenluokittelua voisi hyödyntää kognitiivisen tutkimuksen lisäksi esimerkiksi interaktiivisissa käyttöliittymissä, valheenpaljastuskokeissa tai psykologien apuvälineenä potilaidensa henkisen tilan arvioinnissa. Viihdeteollisuudessa ja visuaalisissa käyttöliittymissä on kysyntää myös kasvojen ilmeiden syntetisoinnille tietokoneen avulla. Esimerkiksi animaatioelokuvissa hahmojen kasvot halutaan usein inhimillisiksi ja huulet sekä ilmeet dialogia vastaaviksi. Tähän soveltuvat hyvin samaiset menetelmät joita käytetään ilmeidenluokittelussa.

FACS-koodausjärjestelmän soveltamista automaattisessa ilmeidenluokittelussa on tutkittu runsaasti. Useimmat ilmeidenluokittelujärjestelmät määrittelevät kiinteän joukon yleisimpiä tunnetiloja, esimerkiksi *{ilo, suru, viha, inho, pelko, tuntematon}*, ja sijoittavat jokaiselle annettulle ilmeelle yhden näistä kiinnitetystä tunnetiloista [Bar99]. Usein kuitenkin pelkän ilmeen olemassaolon havaitsemisen lisäksi olemme kiinnostuneita ilmeen intensiteetistä. Esimerkiksi

vihalla on monia asteita: ärtymys, suuttumus, viha ja raivo. Näitä kaikkia vastaa samankaltainen, mutta erilainen ilme [Bar99]. FACS-koodausjärjestelmässä kasvojen toimintojen esitys erotetaan niihin liittyvien ilmeiden esityksestä, jolloin käytettävissä oleva informaatio ilmeiden intensiteetistä säilyy [EcF78].

Seuraavassa luvussa kuvailemme FACS-koodausjärjestelmän yksityiskohtaisesti, sekä selvitämme, miten sitä voi hyödyntää automaattisessa ilmeidenluokittelussa. Kolmannessa luvussa esittelemme automaattisen ilmeidenluokittelujärjestelmän toiminta- ja toteutusperiaatteet. Samalla käsittelemme ilmeidenluokitteluprosessin eri vaiheissa tarvittavia menetelmiä ja algoritmeja. Lopuksi esittelemme kaksi erityistä menetelmää FACS-koodien määrittämiseksi automaattisesti kuvasta.

2. Formaali kasvontoimintojen koodaustapa

Tässä luvussa esittelemme yksityiskohtaisesti Facial Action Coding System -koodausmenetelmän ja selvitämme miten FACS-koodausta voidaan hyödyntää automaattisessa ilmeidenluokittelussa.

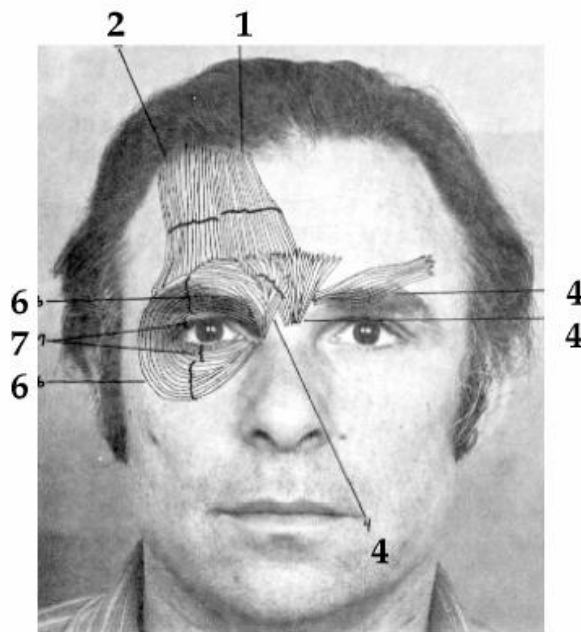
2.1 Koodausjärjestelmän kuvaus

Psykologit joutuvat työssään analysoimaan potilaiden kanssa käymiensä keskustelujen lisäksi heidän ei-verbaalista ulosantiaan. Potilaat ovat usein sulkeutuneita ja heidän tilansa arviointi tukeutuu voimakkaasti käyttäytymisen ja ilmeiden arviointiin keskustelujen lisäksi. Tavallisimpien ilmeiden ja tunnetilojen, kuten ilon, surun, pelon, vihan ja hämmästyksen luokittelu on ihmiselle triviaalia. Sen sijaan tunnetilojen tarkempi analyysi kuvasta on vaativa tehtävä, johon jotkut psykologit joutuvat ammattinsa puolesta harjaantumaan. Tarkempi analyysi käsittää erilaisten tunnetilojen yhdistelmien havaitsemista, tunnetilojen voimakkuuden arviointia, sekä tunnetilojen autenttisuuden erottamista. Tyypillisiä vaikeasti määritettäviä tunnetiloja ovat esimerkiksi teeskennelty ilo johon sekoittuu epätoivoa, itsetuhoisuuden ilmaiseminen ilman aikomusta vahingoittaa itseään tai tietoinen valehteleminen omasta voinnistaan [EcF78].

Subjektiiivisuus on keskeinen ongelma ilmeiden määrittämisessä: kun ihmisasiantuntija määrittää kuvasta tai luonnossa henkilön ilmettä hän tekee intuitiivisia tulkintoja henkilön mielentilasta, perustuen omiin synnynnäisiin ja opittuihin taitoihinsa [EcF78]. Automatisoitaessa ilmeidenluokittelua tämä subjektiiivisuus tulee poistaa ilmeidenluokitteluprosessista. Facial

Action Coding System on kasvojen toimintojen koodaustapa, joka kehitettiin alun perin juuri psykologien apuvälineeksi potilaiden ilmeiden analysoinnissa. Menetelmän taustalla on ihanne, jossa kasvojen lihasten tuottamat ilmeet ja niihin liittyvät tulkinnat erotetaan toisistaan.

FACS-koodaus jakaa kasvojen toiminnot anatomisiin perustein neljäänkymmeneenkuuteen erilaiseen *toimintayksikköön* (Action Unit, AU), jotka vastaavat käytännössä kaikkia silmin erotettavia muutoksia kasvojen ilmeissä. Osa tärkeimmistä toimintayksiköistä selityksineen on esitelty taulukossa 1. Jaottelu perustuu tarkasti ihmiskasvojen lihasten anatomiaan ja jokainen toimintayksikkö on selkeästi määritelty yhden tai useamman kasvojen lihasryhmän supistumisena [EcF78]. Kuvassa 1 [EcF78] on esimerkki muutamasta toimintayksiköstä sekä niitä vastaavista lihasryhmistä. Jaottelussa on pyritty muodostamaan minimaalinen joukko kasvontoimintoja, joka on kyllin ilmaisuvoimainen erottelemaan toisistaan kaikki ihmisen havaitsemat erilaiset ilmeet. Syy siihen, että FACS-koodit soveltuvat hyvin automaattisen ilmeidenluokittelun apuvälineeksi kuvasta on se, että jaottelu perustuu nimenomaan silmin havaittaviin, visuaalisiin muutoksiin kasvoissa.



Kuva 1: Esimerkki toimintayksiköjä vastaavista lihasryhmistä. Numerot vastaavat toimintayksikön numeroa ja tummat viivat kasvoilla osoittavat toimintayksikköön liittyviä lihaksia.

Toimintayksiköt eivät vastaa oletusarvoisesti mitään nimettyjä kasvojen ilmeitä. Sen sijaan kaikki silmin havaittavat kasvojen ilmeet voidaan esittää näiden toimintayksiköiden yhdistelmänä tarkkuudella, joka vastaa vähintään ihmisen havaintokyvyn tarkkuutta [EcF78].

Ilmeen FACS–koodi on siis 46-ulotteinen vektori, jossa alkio on 2-26 arvoisia riippuen kyseisen toiminnan luonteesta. Esimerkiksi silmänräpäytys (AU 45) on selvästi kaksiarvoinen toiminto: silmät ovat joko auki tai kiinni ja välivaiheet ovat ihmisen havaintotarkkuuden rajoissa merkityksettömiä. Leukaperien laskemiselle (AU 26) voi sen sijaan helposti määrätä useita eri asteita ja kyseinen toimintayksikkö on täten useampiarvoinen. Toimintayksiköt, jotka vastaavat kasvojen molemmilla puolilla symmetrisesti esiintyvää kasvontoimintoa, sisältävät oman koodin molemmille kasvonpuoliskoille.

AU 1	Kulmakarvojen sisemmän osan kohotus	AU 18	Huulten mutristus
AU 2	Kulmakarvojen ulomman osan kohotus	AU 19	Kielen näyttäminen
AU 4	Kulmakarvojen madallus	AU 20	Huulten pingotus leveäksi
AU 5	Ylemmän silmäluomen kohotus	AU 22	Pyöreän raon muodostaminen huulilla
AU 6	Poskipään kohotus	AU 23	Huulten pingotus kapeaksi
AU 7	Silmäluomen siristys	AU 25	Huulten raottaminen
AU 8	Huulien puristaminen yhteen	AU 26	Leukaperien laskeminen
AU 9	Nenänvarren rypistyminen	AU 27	Suun supistaminen viiruksi
AU 10	Ylähuulen kohottaminen	AU 28	Huulten imeminen
AU 11	Suupielen kuopan syventäminen	AU 30	Leukaperien siirtäminen sivuttain
AU 12	Suupielen vetäminen taakse	AU 32	Alahuulen pureminen
AU 13	Posken pullistaminen	AU 38	Sierainten laajentaminen
AU 14	Hymykuopan muodostaminen	AU 39	Sierainten supistaminen
AU 15	Suupielen painaminen alas	AU 43	Silmät rennosti suljettuna
AU 16	Alahuulen vetäminen alas	AU 44	Karsastus
AU 17	Leuan kohotus	AU 45	Silmien räpäytys

Taulukko 1: Tärkeimpiä kasvojen toimintayksiköitä.

FACS–koodit ovat yksikäsitteisesti ja täsmällisesti määriteltyjä suhteessa kasvojen anatomiaan mutta koodien määrittäminen kuvasta ei ole täysin objektiivista. Kunkin toimintayksikön määrittely sisältää neljä kohtaa:

- (1) Selostus toimintayksikön kuvaaman toiminnan aiheuttavista lihaksista sekä näiden käyttäytymisestä toiminnan aikana.
- (2) Toimintayksikön vaikutus kasvojen ilmiösuun.
- (3) Ohjeet joiden avulla ihminen voi itse aktivoida toimintayksikön kasvoillaan.
- (4) Sääntö siitä, millaisia muutoksia kasvoilla täytyy havaita ennen kuin toimintayksikön voi katsoa olevan aktiivinen.

Liitteessä A on esimerkkejä toimintayksiköiden määrittelyistä. Kuten määrittelyn asettelusta voi nähdä, kohdat (2), (3) ja (4) ovat riippuvaisia FACS-koodien arvioijasta. Puhuttaessa automaattisesta ilmeidenluokittelusta koodien arvioija on ilmeidentunnistusjärjestelmä.

FACS-koodaus on yleisin ilmeiden analysoinnin työkalu psykologien ja kognitiotieteilijöiden parissa, johtuen sen yksinkertaisuudesta ja riittävästä tarkkuudesta kasvontoimintojen erottelussa. FACS-koodauksen kehittäjät argumentoivat koulutettujen FACS-asiantuntijoiden olevan 70-95 prosenttisesti yksimielisiä kasvojen kuviin liitetystä koodeista, riippuen yksimielisyyden mitasta: kasvojen toimintayksiköiden aktiivisuudesta asiantuntijat ovat pääsääntöisesti samaa mieltä, kun taas arvioitu toimintayksikön intensiteetti vaihtelee eri arvioijien kesken suuresti [EcF78].

2.2 Ilmeiden esittäminen FACS-koodien avulla

Kasvojen toimintayksiköt eivät vastaa nimettyjä kasvojen ilmeitä, mutta FACS-koodeilla voidaan luontevasti kuvata ilmeitä. Ilmeiden esittämisen pohjaksi tarvitaan oletus että henkilön mielentila, ilmeet ja kasvontoiminnot liittyvät yksilöllistä vaihtelua lukuun ottamatta kiinteästi toisiinsa. Tässä aliluvussa esitetty matemaattinen muotoilu ilmeiden esittämiseksi ei perustu suoraan ulkoisiin lähteisiin, mutta useat ilmeidenluokittelujärjestelmät käyttävät implisiittisesti samankaltaisia menetelmiä [EsP97, CKT01]. Tässä esitetty muotoilu pyrkii yksinkertaistamaan ja havainnollistamaan ilmeiden esittämisen periaatteita, puuttumatta teknisiin yksityiskohtiin.

On ilmeistä, että useampi kasvojen asento voi vastata samaa ilmettä: esimerkiksi vihaa voi ilmaista monilla erilaisilla kasvontoiminnoilla. Tämä tulee ottaa huomioon rakennettaessa ilmeille esitystapaa FACS-koodien avulla. Sama FACS-koodi voi myös vastata useampaa kuin yhtä ilmettä, sillä kaikki ilmeet eivät ole toisiaan poissulkevia. Esimerkiksi hymy ja hämmästys ovat kaksi erilaista ilmettä, jotka esiintyvät kasvoilla usein samanaikaisesti. Tällaisen yhdistelmän voisi nimetä omaksi uudeksi ilmeekseen, mutta mahdollisia tällaisia yhdistelmiä on

useita tuhansia [EcF78] ja niiden kaikkien numeroiminen ei ole automaattisen ilmeidenluokittelun kannalta tehokasta tai mielekäästä.

Ilmeinen tapa esittää ilmeitä FACS-koodien avulla on valita esitystavaksi yksi, ilmettä hyvin karakterisoiva FACS-koodi ja sopia, että kyseistä ilmettä vastaa sellaiset FACS-koodit, jotka ovat tietyn kynnsarvon etäisyydellä tästä karikatyyristä. Etäisyyden mittana voisi käyttää vaikka FACS-koodivektorien euklidista etäisyyttä toisistaan. Tämä ei toimi käytännössä, sillä yksilöiden väliset erot kasvontoiminnoissa samaa ilmettä muodostettaessa ovat liian suuria [EcF78] ja monet toimintayksiköt ovat ilmeen kannalta merkityksettömiä, jolloin niiden ei pitäisi vaikuttaa ilmeitä esittävien koodien vertailuun.

Aiemmin mainittujen havaintojen pohjalta muodostamme seuraavanlaisen esitystavan ilmeelle I , kun tarkastellaan ilmeidenluokittelua yhden FACS-koodin perusteella: $I = \{\theta, k, \Delta\}$. Tässä θ on ilmettä vastaava keskimääräinen FACS-koodivektori eli ilmeen karikatyyri $\theta \in N^{46}$, k on painokerroinvektori, siten että k_i on arvio väliltä 0..1 sille, kuinka merkittävä kasvontoiminto θ_i on ilmeen I luokittelun kannalta. Δ on maksimipoikkeamavektori, joka antaa ylärajan sille, kuinka paljon annetun FACS-koodin kukin toimintayksikkö saa enintään poiketa karikatyyrin vastaavasta toimintayksiköstä, jotta annetun koodin voi katsoa vastaavan ilmettä I . Tällä muotoilulla annettu FACS-koodi θ' esittää ilmettä I jos ja vain jos $\forall_{i \in [1..46]} k_i |\theta_i - \theta'_i| < \Delta_i$.

Tässä tutkielmassa olemme kiinnostuneita ilmeiden luokittelusta, kun käytettävissä on kuvasarjasta määritetty järjestetty jono FACS-koodeja $(\theta'_1, \theta'_2, \dots, \theta'_n), n \in N$. Tällainen sarja kasvontoimintoja voi sisältää useita ilmeitä, samanaikaisesti rinnakkain tai sarjana peräkkäin [EcF78]. Osa kasvojen ilmeistä tai eleistä on taas luonteeltaan usean toiminnon sarjoja, kuten pään nyökytys tai välinpitämättömyyttä osoittava kulmien kohautus yhdistettynä pään kallistamiseen.

Luvussa 5 esittelemme erityisiä menetelmiä tällaisten kuvasarjojen esittämiseksi ja analysoimiseksi. Tässä yleistämme useamman kasvontoiminnon sarjana muodostetun ilmeen esityksen seuraavasti:

$$I^n = \{\theta^n, k^n, \Delta^n\}, n \in N,$$

missä θ^i on useampiosaisen ilmeen vaihetta i vastaava karikatyyri, k^i on painokerroinvektori joka kertoo, mitkä toimintayksiköt ovat oleellisia useampiosaisen ilmeen i :nnessä vaiheessa ja Δ^i on vastaavasti ilmeen i :nnen vaiheen maksimipoikkeamavektori.

Annettu jono FACS-koodeja $(\theta'_1, \theta'_2, \dots, \theta'_m)$, $m \in N$ sisältää ilmeen I^n jos ja vain jos:

$$\exists_{m' \in [1..(m-n)]} \forall_{n' \in [1..n]} \forall_{i \in [1..46]} k_i^{n'} \mid \theta'_{m'+n'} - \theta'_{m'} \mid < \Delta_i^{n'}$$

Eli annettu koodijono sisältää osajonon, joka toteuttaa useampiosaisen ilmeen asettamat rajoitteet ja vaatimukset.

Karikatyyrien muodostamiseksi tarvitaan joukko valokuvia, joissa eri henkilöt esittävät saman ilmeen kontrolloiduissa olosuhteissa. Kuvassa 2 [Coh02] on esimerkki suppeasta tällaisesta kuvajoukosta.



Kuva 2: Esimerkkejä eri henkilöiden neutraalista (1. ylhäältä), iloisesta, kuvottuneesta ja vihaisesta ilmeestä.

Tällaisista valokuvista määritettyjen FACS-koodien joukkoa kutsuttakoon *koulutusjoukoksi* $K = \{\theta''_1, \theta''_2, \dots, \theta''_n\}$, $n \in N$. Karikatyyri θ saadaan koulutusjoukon koodivektorien keskiarvosta:

$$\theta = \frac{1}{n} \sum_{i=1}^n \theta''_i$$

Painokerroinvektori k saadaan tarkastelemalla koulutusjoukon FACS-koodien sisältämien toimintayksiköiden otoskeskihajontaa. Jos toimintayksikön intensiteetin hajonta samaa ilmettä esittävässä kuvissa on suuri, on järkevää antaa toimintayksikölle pieni painokerroin, jotta tämä hajonta ei vaikuta ilmeen määrittämiseen. Vastaavasti jos toimintayksikön intensiteetin hajonta on pieni, annetaan toimintayksikölle suuri painokerroin. Näin ollen painokerroinvektorin alkio k_i voidaan esittää kääntäen verrannollisena vastaavan toimintayksikön otoskeskihajontaan, normalisoituna välille $[0..1]$:

$$k_i = 1 - \frac{1}{26} * \sqrt{\frac{1}{n-1} \sum_{j=1}^{j < n} (\theta_j - \theta_i)^2}$$

Maksimipoikkeamavektorin määrittäminen automaattisesti on kyseenalaista, sillä ihmisillä on hyvin yksilöllisiä tapoja ilmaista erilaisia ilmeitä [EcF78]. Ihmisarvioija voi määrittää pieniä maksimipoikkeama-arvoja sellaisille toimintayksiköille, jotka intuitiivisesti eivät missään nimessä saa olla aktiivisia tai joiden ehdottomasti täytyy olla aktiivisia tarkasteltavan ilmeen yhteydessä. Mikäli käytettävissä on suuri tietokanta FACS–koodeja, joita vastaavat ilmeet tiedetään etukäteen, voidaan tietokantaa hyväksi käyttäen määrittää maksimipoikkeamat siten, että kaikki poikkeamat sallitaan aluksi, ja pienennetään kutakin maksimipoikkeaman arvoa, kunnes jokin ilmeellä merkityistä FACS–koodeista ei enää tyydytä ilmeen asettamia vaatimuksia.

2.3 FACS–koodien haitat automaattisessa ilmeidenluokittelussa

FACS–koodien isoimmaksi haittapuoleksi on esitetty niiden epäformaali määrittely [EsP97]: järjestelmä on suunniteltu käytettäväksi siten, että FACS–asiantuntijaksi pyrkivä ihminen perehtyy huolellisesti kasvojen anatomiaan ja lukee FACS–koodauksen kehittäjien laatiman perehdytysoppaan, jossa on yksityiskohtainen selvitys kustakin toimintayksiköstä, sekä siitä miten koodeja määritetään valokuvista. Tämä prosessi ei ole siinä mielessä täsmällisesti määritelty, että saman koulutuksen voisi järjestää tietokonejärjestelmälle.

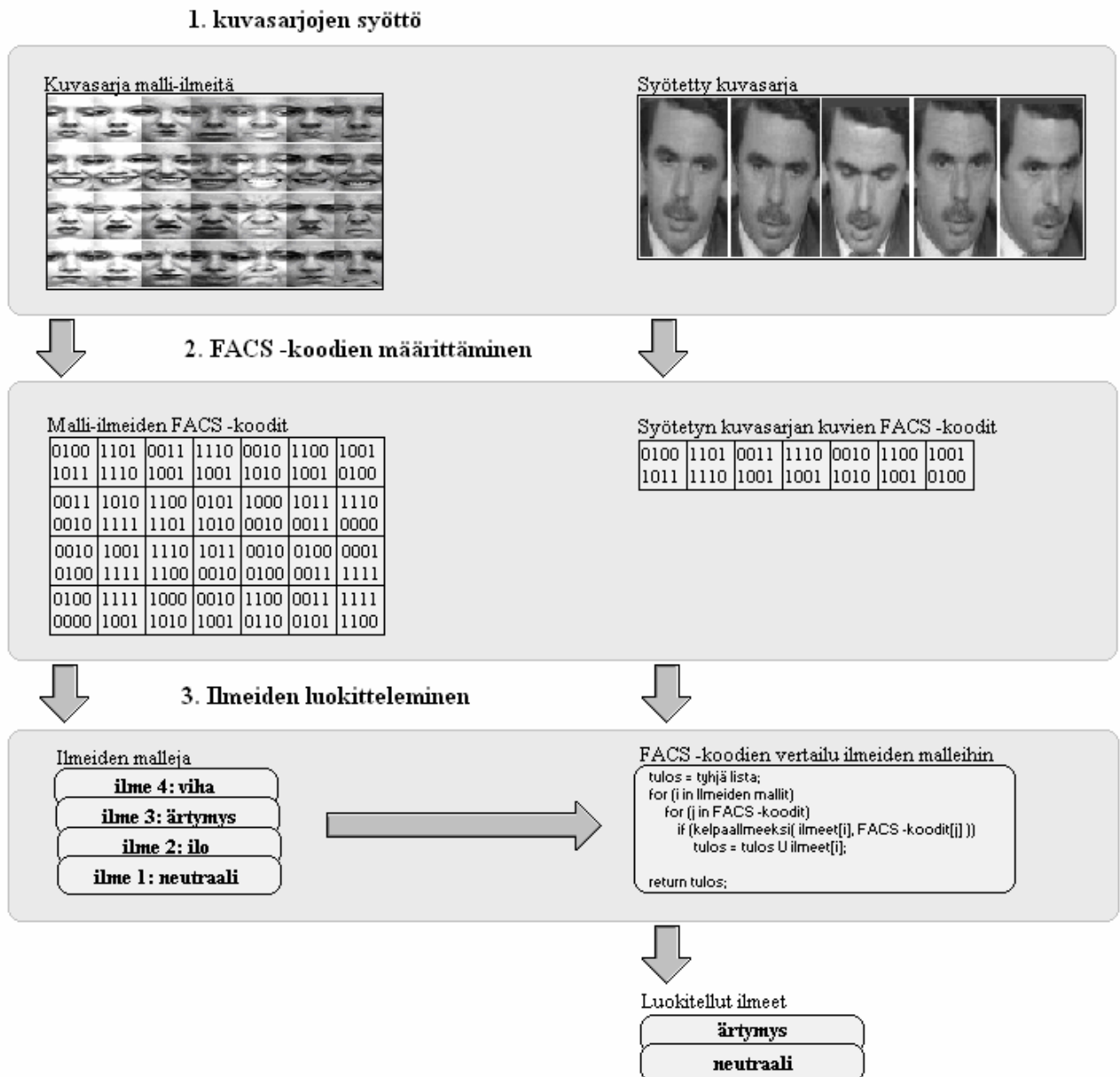
Tietokoneella ei ole käytettävissään esimerkiksi psykologien omaamaa vuosien kokemusta potilaiden ilmeiden tarkkailemisesta tai ylipäättänsä ihmisen kykyä hahmottaa ja luokitella ilmeitä kasvoilla. Mitään formaaleja takuita ei ole myöskään siitä, että FACS–koodit kykenevät erottelemaan halutulla tarkkuudella kaikki ilmeet, joista saattaisimme olla kiinnostuneita.

3. Ilmeidenluokittelu kuvasarjasta

Tässä luvussa esittelemme tyypillisen FACS–koodeihin perustuvan automaattisen ilmeidenluokittelujärjestelmän rakenteen. Lisäksi esittelemme kuvasarjan esikäsittelyyn liittyvät toimenpiteet omassa aliluvussaan, sillä ne muodostavat tärkeän osaongelman ilmeiden automaattisessa tunnistamisessa.

3.1 Ilmeidenluokittelujärjestelmän kuvaus

Automaattisen ilmeidenluokittelujärjestelmän voidaan ajatella koostuvan kolmesta komponentista: kuvasarjojen syöttö, FACS-koodien määrittäminen ja Ilmeiden luokittelu. Tyypillinen ilmeidenluokittelujärjestelmän rakenne on esitetty kuvassa 3.



Kuva 3: Automaattinen Ilmeidenluokittelujärjestelmä.

Kuvasarjojen syöttökomponentti ottaa vastaan sekä järjestelmän ylläpitäjän syöttämät kuvasarjat malli-ilmeistä että käyttäjän syöttämät kuvasarjat luokiteltavista ilmeistä. Komponentin tehtävä on käsitellä kuvasarjat siten, että tuloksena on joukko yhdenmukaisia kasvojen kuvia.

Yhdenmukaisissa kasvojen kuvissa kaikki kuvat ovat samankokoisia, keskimäärin yhtä kirkkaita sekä kierretyt siten, että kasvot ovat pystyasennossa [Don99]. Lisäksi kasvojen ulkopuolelle jäävät alueet kuvissa leikataan pois ja korvataan mustalla. Tämän komponentin suorittamia toimenpiteitä kutsutaan kuvasarjan *normalisoinniksi*.

FACS-koodien määrittäjäkomponentti saa syötteekseen normalisoidut kuvasarjat malli-ilmeistä sekä luokiteltavista ilmeistä. Komponentin tehtävä on muodostaa täsmälliset mallit kuvien kasvoista. Tässä tutkielmassa tarkastelemme ilmeidenluokittelujärjestelmää, jossa kasvojen malli on yksinkertainen FACS-koodi. Kolme erityistä menetelmää FACS-koodien määrittämiseksi esitellään seuraavassa luvussa.

On huomattava että kasvojen malli ja ilmeen malli ovat erilliset käsitteet. Kasvojen mallilla pyrimme ainoastaan mallintamaan kasvojen lihaksiston tilaa tietyllä hetkellä liittämättä siihen tietoa ilmeistä tai ihmisen tunnetiloista. Ilmeen mallilla pyrimme muodostamaan sääntöjä, jotka liittävät kasvojen lihaksiston tiloja tiettyihin nimettyihin ilmeisiin, kuten vihaan tai suruun. FACS-koodit ovat kasvojen malleja.

Ilmeiden luokittelukomponentti saa syötteekseen malli-ilmeiden ja luokiteltavien ilmeiden FACS-koodit. Komponentin tehtävä on näitä koodeja hyödyntäen selvittää, mitä ilmeitä luokiteltavassa kuvasarjassa esiintyy. Komponentin on muodostettava malli-ilmeiden FACS-koodien avulla ilmeiden malleja, joiden avulla syötetyssä kuvasarjassa esiintyviä ilmeitä voidaan luokitella.

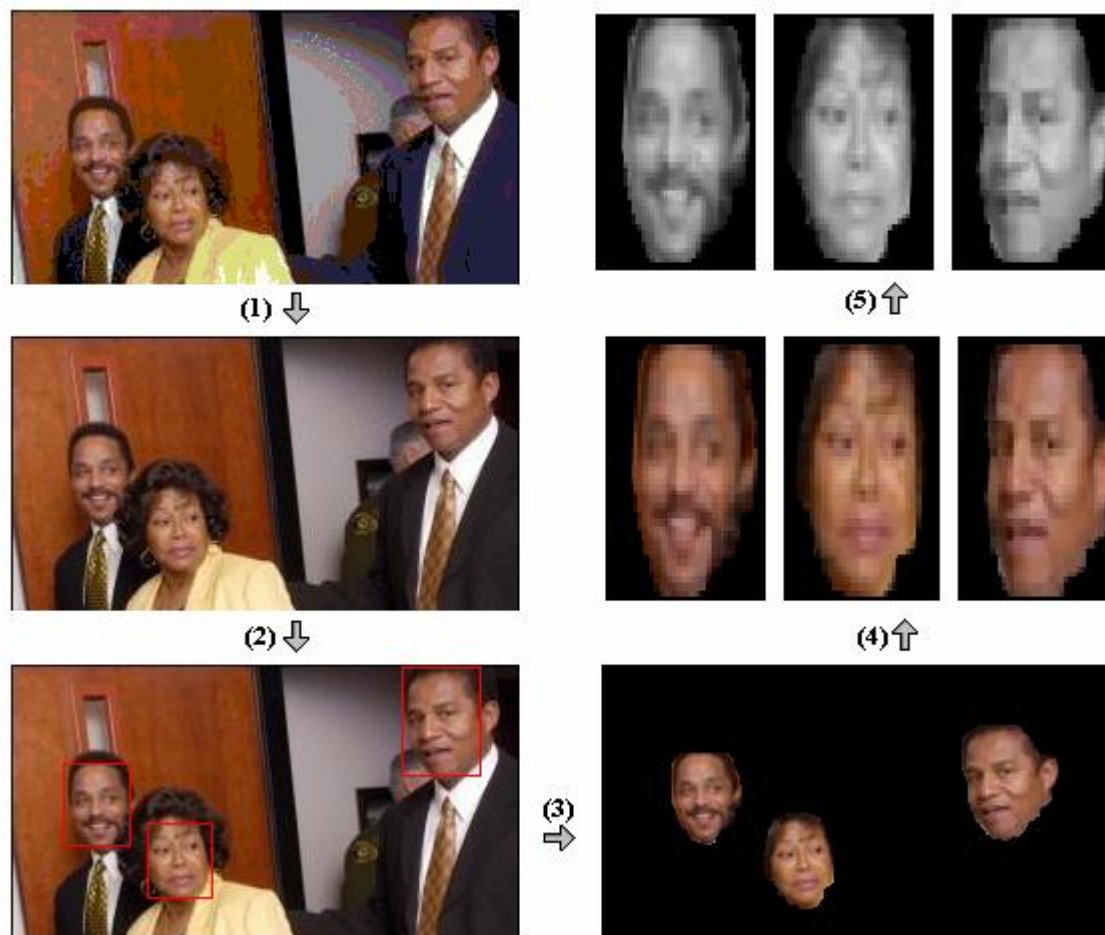
3.2 Kuvasarjan normalisointi

Kuvasarjojen normalisointi koostuu tyypillisesti viidestä osasta, joita kuva 4 havainnollistaa: kohinan poistaminen (1), kasvojen paikantaminen (2), taustan leikkaaminen pois (3), kasvojen kuvien kääntäminen pystyasentoon ja skaalaaminen samankokoisiksi (4) sekä kuvien kirkkauden tasoittaminen (5).

Tunnistusjärjestelmän syötteenä saama kuvasarja sisältää usein kameran sensoreiden mittausvirheistä johtuvaa kohinaa, joka vääristää kuvaa. Kohina vaikeuttaa muiden kuvankäsittelyalgoritmien toimintaa ja on usein poistettava kuvan normalisoinnin aluksi.

Kohinan poistaminen valokuvasta on klassinen ja hyvin tutkittu aihe digitaalisen kuvankäsittelyn parissa. Olemassa olevat kohinanpoistomenetelmät soveltuvat sellaisenaan ilmeidenluokittelujärjestelmiin. Tyypillisesti kuva suodatetaan gaussisesti, jolloin jokainen kuvan pikseli

määräytyy itsensä, sekä sitä ympäröivien pikselien keskiarvosta painotettuna normaalijakauman mukaan siten, että kauempana olevat pikselit vaikuttavat keskiarvoon vähemmän [Zha03].



Kuva 4: kuvan normalisointi.

Suodatuksen tuloksena suuret poikkeamat kahden viereisen pikselin välillä häviävät ja kuva näyttää tasaisemmalta, mutta samalla osa alkuperäisen kuvan informaatiosta katoaa. Tunnistusjärjestelmä valitsee suodatuksen määrän niin pieneksi, että tärkeät kasvojen ja ilmeiden piirteet säilyvät suodatetussa kuvassa.

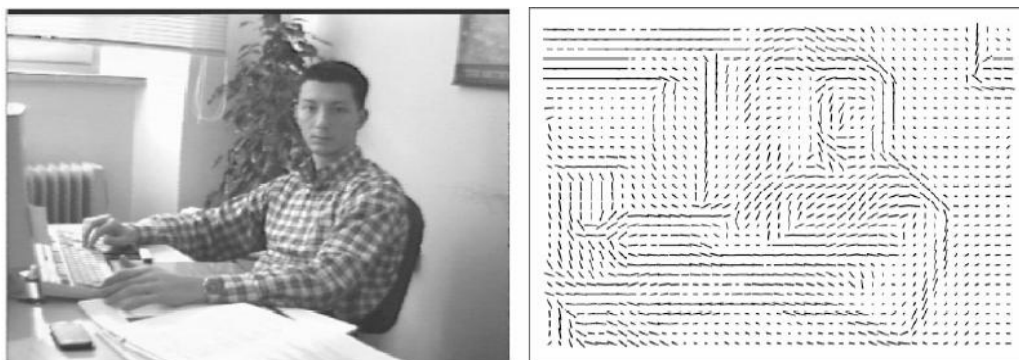
Useat kehitetyt ilmeidenluokittelumenetelmät olettavat saavansa syötteenä valmiiksi normalisoituja kuvasarjoja kasvoista [EsP97, Don99, BIY97]. Useimmissa kuviteltavissa olevista sovelluksista järjestelmän syötteenä on kuitenkin videokuvaa tosielämän tilanteista, joissa näkyy kasvojen lisäksi ilmeidenluokittelun kannalta epäoleellista ympäristöä. Tällaisissa tilanteissa ilmeidenluokittelujärjestelmä joutuu paikantamaan kasvot kuvasarjasta.

Kasvojen automaattinen paikantaminen on hankala tehtävä: kasvot voivat esiintyä missä tahansa asennossa, osittain muiden hahmojen peittäminä, oudosti valaistuin ja erikokoisina [MaM00]. Kasvojen muoto vaihtelee myös suuresti, johtuen yksilöllisten päänmuotojen lisäksi varjoista ja hiusrajasta. Jos kuva on otettu pienellä resoluutiolla, siinä esiintyviä kasvoja vastaavat alueet voivat olla hyvinkin pieniä, esimerkiksi 20x20 pikseliä.

Varhaisimmat kasvojen paikannusalgoritmit käyttivät suoraan kuvienluokitteluun käytettyjä menetelmiä, joissa käytettiin joukkoa edustavia mallikuvia ihmiskasvoista useasta eri kuvakulmasta ja erikokoisina. Mallikasvoja sovitettiin vuorollaan tarkasteltavan kuvan eri kohtiin ja kasvoiksi määriteltiin sellaiset kohdat kuvassa, jotka muistuttivat riittävästi jotakin mallikuvista [Zha03]. Tällaiset algoritmit ovat laskennallisesti vaativia ja herkkiä valaistuksen aiheuttamille vääristymille kasvojen kuvissa. Parempiin tuloksiin päästään käyttämällä *mukautuvia malleja* (deformable templates) kasvoista. Tällöin kasvot esitetään yksinkertaisena geometrisena mallina, jossa kasvot on esitetty muutaman ympyrän ja kolmion avulla, jotka kuvaavat silmien, suun ja nenän sijaintia kasvoissa. Tällaista mallia sovitetaan eri asennoissa ja erikokoisina kuvan eri kohtiin ja kasvoiksi kelpaavat sellaiset alueet kuvassa, joissa mallin suu, silmät ja muut osat vastaavat kuvassa esiintyviä yksityiskohtia.

Eräs menetelmä kasvojen paikantamisen nopeuttamiseksi ovat *suuntakuvat* (directional image) [Mam00], joiden avulla kuvasta löydetään reaaliaikaisesti kaikki potentiaaliset kasvoalueet. Suuntakuva muodostetaan muuntamalla kuva joukoksi viivoja, jotka vastaavat kuvassa esiintyvien pintojen suuntia ja kontrastia. Kapea viiva kuvaa heikommin ympäristöstään erottuvaa pinnan suuntaa ja paksu viiva kuvaa voimakkaasti kontrastoivan alueen suuntaa, kuten esimerkiksi posken ääriviivaa tummempaa taustaa vasten.

Potentiaalisia kasvoja suuntakuvassa ovat viivojen muodostamat pyörteet isomman ja vähemmän pyöristyneen alueen yläpuolella. Kuvassa 5 [MaM00] on esimerkki kuvasta sekä sitä vastaavasta suuntakuvasta. Kasvojen löytyminen on suuntakuvia käytettäessä riippumatonta sekä valaistuksesta että pään asennosta. Erityisen hyvin suuntakuvat toimivat kuvasarjoissa, sillä ihminen liikuttaa päätään yleensä hitaasti verrattuna muihin raajoihin ja kasvojen suuntakuva muuttuu hitaasti [MaM00].



Kuva 5: Valokuva ja siitä muodostettu suuntakuva

4. FACS-koodien automaattinen muodostaminen kuvasta

Tässä luvussa esittelemme yksityiskohtaisesti kaksi paljon käytettyä menetelmää, optisen virtauksen ja pääkomponenttianalyysin, joiden avulla FACS–koodeja voidaan määrittää automaattisesti kuvasarjasta.

4.1 Merkinnät

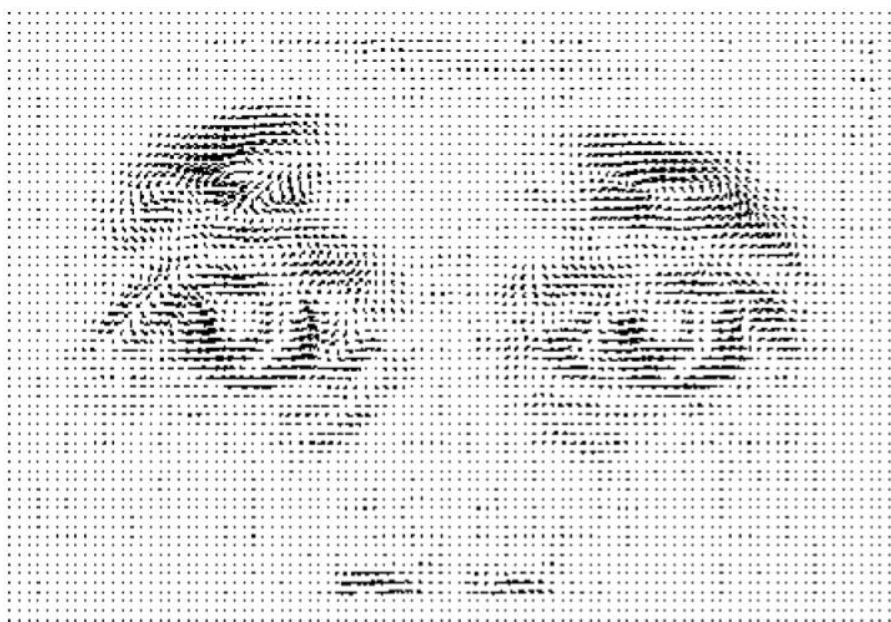
Sovimme aluksi lyhyesti merkinnöistä, joita käytämme esitellessämme FACS–koodien määritysmenetelmiä. Olkoon tarkasteltavana $x \times y$ resoluutiolla tallennettu, n kuvan mittainen ja normalisoitu kuvasarja, joka esittää kasvoja: $K = \{k_1, k_2, \dots, k_n\}, k_i \in N^{xy}$. Oletamme kuvasarjan normalisoiduksi siten että kuvat esittävät täsmälleen yhden henkilön kasvoja ja kuvasarja etenee kronologisessa järjestyksessä tasaisella nopeudella. Lisäksi oletamme, että kuvaolosuhteet pysyvät muuttumattomina kuvasarjan ajan. Kuvaolosuhteilla tarkoitamme kasvojen asentoa, etäisyyttä kamerasta sekä valaistuksen suuntaa ja voimakkuutta. Tehtävänä on määrittää FACS–koodit jokaisesta kuvasarjan K kasvojen kuvasta.

4.2 Optinen virtaus

Optinen virtaus (optic flow) tarkoittaa liikkeen analysointia kuvasarjassa. Kuvassa esiintyvien värien, pintojen, reunojen ja kontrastien sijasta olemme kiinnostuneita siitä, mikä kuvassa muuttuu siirryttäessä kuvasarjassa seuraavaan kuvaan. Erityisesti haluamme määrittää jokaiselle kuvan pikselille sen *virtauksen* (flow), eli nopeuden ja suunnan suhteessa edelliseen kuvaan.

Optinen virtaus on käytetyin menetelmä FACS-koodien automaattiseen määrittämiseen [Don99].

Kuvien K virtauskuvat (flow field) $V = \{v_1, v_2, \dots, v_n\}, v_i \in R^{4096} \times R^{4096}$ kertovat kuvien jokaisen pikselin virtauksen, eli suunnan ja nopeuden suhteessa edelliseen kuvaan. Kuvassa 6 [Don99] on esimerkkinä osa virtauskuvaa kuvasarjasta, jossa henkilö kohottaa kulmakarvojaan. Kuvassa viivat kuvaavat niitä vastaavien pikselien liikkeen suunnan ja nopeuden. Virtauskuvien vahvuus on niiden vikasietoisuus vaihteleville valaistusolosuhteille.



Kuva 6: Virtauskuva.

Määritettäessä FACS-koodeja kuvasta optisen virtauksen avulla, tarvitsemme kaikille neljällekymmenellekuudelle toimintayksikölle niitä parhaiten karakterisoivat virtauskuvat $V_{AU} = \{v_{AU1}, v_{AU2}, \dots, v_{AU46}\}, v_{AUi} \in R^{4096} \times R^{4096}$. Yksi tapa muodostaa tällaiset karikatyyrit on kuvata toimintayksiköt usean testihenkilön suorittamana, ja laskea näiden kuvasarjojen virtauskuvista keskiarvo tai mediaani [Bar99].

Kuvasarjan ensimmäisen kuvan k_0 virtaus on oletusarvoisesti nolla. Kuvan k_i pikselin $p_i(x, y)$ virtaus voidaan määrittää paikallisesti sitä edeltävän ja sitä seuraavan kuvan avulla. Tätä varten tarvitsemme tiettyjä estimaatteja.

Kahden pikselin *vastaavuuskerroin* $R(p_i(x, y), p_{i+1}(x', y'))$ on verrannollinen siihen, kuinka hyvin seuraavan kuvan pikseli p_{i+1} vastaa tarkasteltavan kuvan pikseliä p_i . Jos pikselien vastaavuuskerroin on suuri, voimme ajatella pikselin p_i siirtyneen seuraavassa kuvassa kohtaan

p_{i+1} , joten $R(p_i, p_{i+1})$ on verrannollinen siihen, kuinka todennäköisesti pikseliä p_i vastaava virtaus on $(x'-x, y'-y)$.

Kuvissa mahdollisesti esiintyvän kohinan takia pikselien vastaavuuskerroin määritellään yleensä pikselien ympäristöä hyödyntäen seuraavasti [Don99]: Muodostamme pikselien p_i ja p_{i+1} ympärille 3×3 kokoiset ikkunat ja asetamme vastaavuuskertoimen näiden pikseli-ikkunoiden korrelaatioksi:

$$R(p_i(x, y), p_{i+1}(x', y')) = \frac{1}{3 \times 3} \sum_{u=-1}^1 \sum_{v=-1}^1 \frac{(p_i(x+u, y+v) - \overline{p_i})}{s_{p_i}} \frac{(p_{i+1}(x'+u, y'+v) - \overline{p_{i+1}})}{s_{p_{i+1}}},$$

missä $\overline{p_i}$ on pikselien keskiarvo tarkasteltavassa ikkunassa ja s_{p_i} on vastaavien pikselien keskihajonta.

Kuvaa k_i vastaava virtauskuva v_i voidaan laskea pikselien vastaavuuskertoimien avulla. Jokaiselle kuvan pikselille $p_i(x, y)$:

- 1) Muodostetaan edeltävään ja seuraavaan kuvaan 5×5 pikselin ikkunat W_{i-1} ja W_{i+1} kohdan (x, y) ympärille.
- 2) Lasketaan 5×5 matriisiin R'_{i-1} edeltävän kuvan ikkunasta kaikille pikseleille vastaavuuskertoimet pikselin p_i kanssa:

$$R'_{i-1} = \begin{pmatrix} R(p_{i-1}(x-2, y-2), p_i(x, y)) & \text{]} & R(p_{i-1}(x+2, y-2), p_i(x, y)) \\ \wedge & \text{]} & \wedge \\ R(p_{i-1}(x-2, y+2), p_i(x, y)) & \text{]} & R(p_{i-1}(x+2, y+2), p_i(x, y)) \end{pmatrix}.$$

- 3) Lasketaan vastaavasti matriisiin R'_{i+1} seuraavan kuvan ikkunasta vastaavuuskertoimet:

$$R'_{i+1} = \begin{pmatrix} R(p_i(x, y), p_{i+1}(x-2, y-2)) & \text{]} & R(p_i(x, y), p_{i+1}(x+2, y-2)) \\ \wedge & \text{]} & \wedge \\ R(p_i(x, y), p_{i+1}(x-2, y+2)) & \text{]} & R(p_i(x, y), p_{i+1}(x+2, y+2)) \end{pmatrix}$$

- 4) Yhdistetään matriisit painottaen edellisestä kuvasta saatuja vastaavuuskertoimia:

$$R' = \pi R'_{i-1} + R'_{i+1}$$

- 5) Muodostetaan estimaatti pikseliä p_i vastaavasta virtauksesta (u, v) käyttäen vastaavuuskertoimilla painotettua summaa kaikista mahdollisista virtauksista ($u' \in [-2, 2], v' \in [-2, 2]$):

$$u = \frac{\sum_{i=-2}^2 \sum_{j=-2}^2 R'_{u+i+2, v+j+2} * i}{\sum_{i=-2}^2 \sum_{j=-2}^2 R'_{u+i+2, v+j+2}} \quad v = \frac{\sum_{i=-2}^2 \sum_{j=-2}^2 R'_{u+i+2, v+j+2} * j}{\sum_{i=-2}^2 \sum_{j=-2}^2 R'_{u+i+2, v+j+2}}$$

Edellä esitetyllä tavalla voimme laskea jokaiselle kuvan k_i pikselille virtauksen, ja siten muodostaa virtauskuvan v_i .

Koska kasvoissa vierekkäisten kohtien liike korreloi voimakkaasti ja kuvan pikseleisessä saattaa esiintyä kohinasta johtuvia suuria vääristymiä, suodatetaan virtauskuvia yleensä gaussisesti ennen FACS-koodien määrittämistä [Bar99]. Suodatetussa virtauskuvassa jokaisen pikselin virtaus saadaan itsensä, sekä sitä ympäröivien pikselien virtausten keskiarvona painotettuna normaalijakauman mukaan siten, että kauempana olevat pikselit vaikuttavat keskiarvoon vähemmän.

FACS-koodien määrittäminen virtauskuvien avulla on tyypillinen täsmäsongelma. Tarkasteltavana on joukko valmiiksi nimettyjä, FACS-toimintayksiköitä vastaavia virtauskuvia V_{AU} , sekä kuvasarjasta määritetty joukko virtauskuvia V . Tehtävänä on löytää kaikki sellaiset parit $(v, v_{AU}), v \in V, v_{AU} \in V_{AU}$, joille pätee $S(v, v_{AU}) < \varepsilon$, missä S on jokin virtauskuvien samankaltaisuuden mitta. Kynnysarvo ε kertoo paljonko kuvasarjasta määritetty virtauskuva saa enintään poiketa FACS-toimintayksikköä vastaavasta virtauskuvasta, jotta toimintayksikön voi katsoa olevan aktiivinen kuvassa esiintyvillä kasvoilla. Yksi tapa määrittellä $S(v, v_{AU})$ on käyttää vektorien v ja v_{AU} välisen kulman kosinia [Bar99]:

$$S(v, v_{AU}) = \frac{v \cdot v_{AU}}{|v| \cdot |v_{AU}|}$$

Tuloksena saadaan joukko riittävän samankaltaisia pareja:

$$V_{\text{täsmäivät}} = \{(v, v_{AU}) \mid v \in V, v_{AU} \in V_{AU}, S(v, v_{AU}) < \varepsilon\}$$

Tästä joukosta voidaan helposti lukea, mitkä FACS-toimintayksiköt esiintyvät täsmäivissä pareissa ja siten myös tarkasteltavassa kuvasarjassa.

4.2 Pääkomponenttianalyysi

Pääkomponenttianalyysi (principal component analysis, PCA) on lineaarialgebrallinen menetelmä, jolla löydetään annetulle joukolle vektoreita sellainen kanta, joka maksimoi hajonnan vektoreiden välillä. Samalla datan määrä vähenee, sillä analyysin tuottamia kantavektoreita on yleensä selvästi vähemmän kuin alkuperäisiä vektoreita. Liitteessä B esitellään lyhyesti joitakin lineaarialgebran käsitteitä, joita käytetään tässä luvussa. Pääkomponenttianalyysin tuottamia kantavektoreita kutsutaan datan *pääkomponenteiksi*.

Digitaalinen valokuva, jonka resoluutio olkoon $x \times y$, voidaan tulkita vektoriksi katenoimalla sen rivit peräkkäin. Tuloksena saadaan xy -ulotteinen vektori. Usein sovelluksissa kuvat ovat niin suuria (esimerkiksi 640×480), että kaikkien $x \times y$ kuvien muodostama avaruus on dimensioltaan laskennan kannalta liian suuri. Pääkomponenttianalyysillä tarkasteltavan vektoriavaruuden dimensiota voidaan pienentää huomattavasti siten, että datan sisältämästä informaatiosta katoaa vain pieni osa.

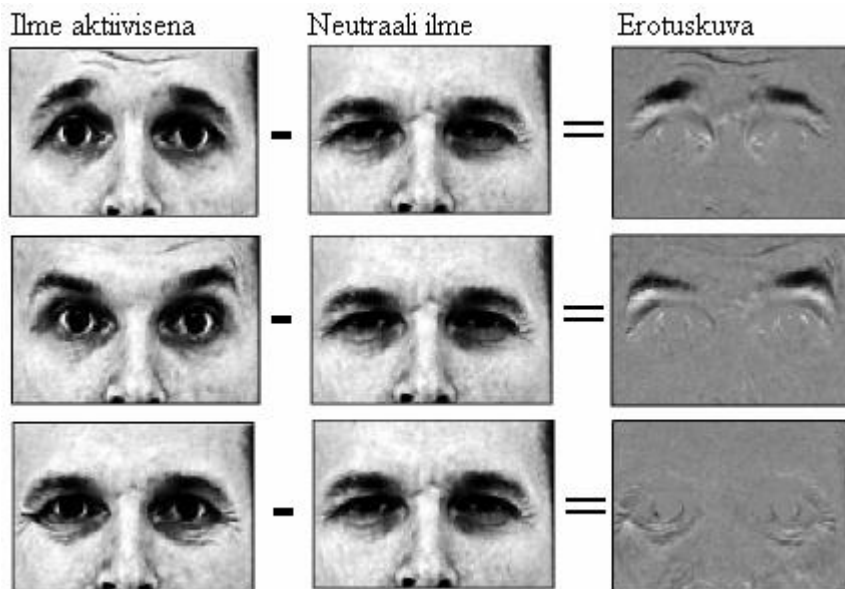
Pääkomponenttianalyysin soveltaminen ilmeidenluokitteluun perustuu oletukseen, että kuvat ihmiskasvoista erilaisine ilmeineen sijaitsevat kaikkien kuvien muodostamassa avaruudessa suppeassa ja yhtenäisessä aliavaruudessa, jonka dimensio on huomattavasti pienempi kuin xy [TuP91, Don99]. Tällöin pääkomponenttianalyysillä löydetään annetulle joukolle kasvojen ilmeiden kuvia sellaiset kantavektorit, joiden lineaarikombinaationa voidaan esittää suuri määrä erilaisia kasvojen ilmeen kuvia. Lisäksi kantavektorit ovat sellaisia, että niiden avulla esitetynä eri ilmeiden välinen hajonta on mahdollisimman suuri, jolloin ilmeiden luokittelu helpottuu.

Olkoon käytettävissä 46 kappaletta kuvasarjoja, joissa kussakin on n kuvaa:

$$K_{AU} = \begin{pmatrix} k_{AU1}^1 & \setminus & k_{AU1}^n \\ \wedge & \cdot & \wedge \\ k_{AU46}^1 &] & k_{AU46}^n \end{pmatrix}$$

Kuvasarjoissa sama testihenkilö esittää kontrolloiduissa olosuhteissa jokaisen neljästäkymmenestäkuudesta toimintayksiköstä siten, että ensimmäisessä ja viimeisessä kuvassa kasvoilla on neutraali ilme ja keskimmaisessä kuvassa toimintayksikkö on mahdollisimman aktiivinen. Merkintöjen yksinkertaistamiseksi tarkastelemme vain yhden testihenkilön esittämiä kuvasarjoja, mutta käytännössä pääkomponenttianalyysi löytää sitä paremmat kantavektorit, mitä enemmän dataa on käytettävissä [TuP91].

Kuvasarjoista voidaan muodostaa kullekin toimintayksikölle *erotuskuvat* (δ -image) vähentämällä kuvasarjan keskimmaisesta kuvasta ensimmäinen kuva, jolloin jäljelle jää pelkästään ne osat kasvojen kuvasta, jotka ovat muuttuneet toimintayksikön suorittamisen aikana [Don99]. Olkoon K'_{AU} muodostettu erotuskuvien joukko $K'_{AU} = \{k_{AU1}^{n/2} - k_{AU1}^1, \dots, k_{AU46}^{n/2} - k_{AU46}^1\}$. Kuvassa 7 [Don99] on esimerkkejä tällaisista erotuskuvista.



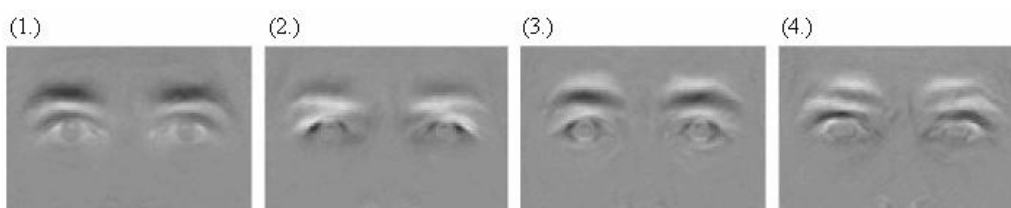
Kuva7: Ilmeitä sekä niitä vastaavia erotuskuvia.

Suoritamme erotuskuvavektorien joukolle pääkomponenttianalyysin. Joukon K'_{AU} pääkomponentit ovat joukon kovarianssimatriisin ominaisvektorit. Aloitamme muodostamalla keskitetyn datamatriisin X , jonka sarakkeita ovat normalisoidut erotuskuvavektorit. Normalisointi tapahtuu vähentämällä kustakin erotuskuvavektorista kaikkien erotuskuvien keskiarvo:

$$X = [K'_{AU0} - \overline{K'_{AU}}, K'_{AU1} - \overline{K'_{AU}}, \dots, K'_{AU46} - \overline{K'_{AU}}]$$

Keskitetyn datamatriisin X kovarianssimatriisi $\text{cov}(X)$ saadaan kertomalla matriisi käänteismatriisillaan $\text{cov}(X) = XX^T$. Etsityt ilmeiden kuvien pääkomponentit ovat tämän kovarianssimatriisin ominaisvektorit. Jokaista ominaisvektoria vastaa tietty ominaisarvo, joka on verrannollinen siihen, kuinka paljon hajontaa vastaavan ominaisvektorin suunnassa alkuperäinen data sisältää; mitä suurempi ominaisarvo ominaisvektoriin liittyy, sitä tärkeämpi kyseinen ominaisvektori on datan luokittelun kannalta.

Ominaisvektorit $E = \{e_1, e_2, \dots, e_k\}, e_i \in R^{nm}$ sekä ominaisarvot $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\}, \lambda_i \in R$ saadaan ratkaisemalla yhtälö $\text{cov}(X)E = \Lambda E$. Tuloksena on ortogonaalinen projektiomatriisi E . Tätä matriisia pienennetään vielä valitsemalla d kappaletta tärkeimpiä pääkomponentteja näitä vastaavien ominaisarvojen perusteella ja hylkäämällä muut komponentit. Sopivan d :n valinta on sovelluksen toteuttajan tehtävä. Kuvassa 8 [Don99] on erään kasvojen yläpuoliskon ilmeitä esittävän kuvasarjan neljä tärkeintä pääkomponenttia.



Kuva 8: Esimerkki kasvojen yläosan kasvontoimintojen pääkomponenteista.

FACS-toimintayksiköiden automaattinen määrittäminen kuvasta pääkomponenttianalyysin avulla tapahtuu projisoimalla kuvia pääkomponenttien virittämään avaruuteen, jota tässä nimitetään *ilmeavaruudeksi*. Kuvan y projektiio ilmeavaruuteen $\text{proj}(y)$ saadaan vähentämällä kuvasta erotuskuvien keskiarvo ja kertomalla kuva ilmeavaruuden kantavektoreilla:

$$\text{proj}(y) = E(y - \overline{K}_{AU})$$

Koska pääkomponenttianalyysin kohteena oli eri FACS-toimintayksiköitä esittäviä kuvasarjoja, tämä projektiio maksimoi hajonnan eri toimintayksiköitä esittävien kuvien välillä ja on siten hyvä apuväline toimintayksiköiden määrittämisessä. Kuvassa $K_i, i \in [1, n]$ esiintyy toimintayksikkö $AU_j, j \in [1, 46]$ jos ja vain jos $S(\text{proj}(K_{AU_j}), \text{proj}(K_i)) < \epsilon$, missä S on jokin vektoreiden etäisyyden mitta ja ϵ on sovelluksen laatijan asettama kynnyisarvo toimintayksikön hyväksymiselle. Etäisyyden S voi määrittää esimerkiksi vektorien euklidiseksi etäisyydeksi [Don99].

Pääkomponenttianalyysin käyttämistä hahmontunnistuksessa on kritisoitu siitä, että se maksimoi hajonnan kaikkien kohteena olevien dataelementtien välillä [BHK97]; siis myös hajonnan kahden samaa toimintayksikköä esittävän kuvasarjan välillä. Analyysia tehostetaan usein lisäämällä projektiioon luokkakohtainen hajautin, joka minimoi hajonnan samaan luokkaan kuuluvien dataelementtien välillä. Yksi tähän hyvin sopiva menetelmä on FLD (Fischer's Linear Discriminant) [BHK97], jota on käytetty paljon esimerkiksi automaattisessa kasvojen tunnistamisessa [BHK97, Zha03].

5. Yhteenveto

Esittelimme automaattinen ilmeidenluokittelun paradigman sekä selvitimme miten alun perin psykologien käyttöön luotua FACS-koodausta voidaan hyödyntää ilmeidenluokittelussa. FACS-koodien epäformaalien määrittelyjen vuoksi niiden automaattinen löytäminen kuvasarjasta on vaikea tehtävä.

Optinen virtaus ja pääkomponenttianalyysi ovat keskeisimmät perusmenetelmät, joilla koodeja voidaan automaattisesti määrittää. Useat ilmeidenluokittelujärjestelmät käyttävät samankaltaisia menetelmiä. Erilaisten menetelmien hyvyyden vertailu on hankalaa, sillä ilmeidenluokittelujärjestelmän testaamiseen ei ole olemassa vakiintuneita, yhtenäisiä käytäntöjä. Eri menetelmiä on testattu eri tavoin ja tulokset eivät ole keskenään vertailukelpoisia.

Ilmeidenluokittelu on nuori tutkimusala ja FACS-koodaukseen perustuvat järjestelmät ovat kehittäjiensä mukaan toistaiseksi luotettavimpia ilmeidenluokittelijoita johtuen koodauksen tarkkuudesta.

Useat olemassa olevat automaattiset ilmeidenluokittelujärjestelmät kykenevät luokittelemaan luotettavasti ilmeitä kontrolloiduissa olosuhteissa otetuista kuvasarjoista, joissa testihenkilöt suorittavat halutun ilmeen [BIY97,Don99,EsP97,KaM03]. Tilanteissa, joissa kuvattavat henkilöt eivät ole yhteistyöhaluisia tai eivät tiedä kuvauksesta, nykyiset järjestelmät eivät kykene luokittelemaan ilmeitä riittävän hyvin.

Lähteet

- Bar99 Bartlett, M. et al., *Measuring facial expressions by computer image analysis*. *Psychophysiology*, 36, 2(1999), sivut 253-278.
- BHK97 Belhumeur, P., Hespanha, J. ja Kriegman, D., *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 7(1997), sivut 711-720.
- BIY97 Black, M. ja Yacoob, Y. *Recognizing facial expressions in image sequences using local parameterized models of image motion*. *International Journal of Computer Vision*, 25, 1(1997), sivut 23–48.
- CKT01 Cohn, J., Kanade, T. ja Tian, Y., *Recognizing action units for facial expression analysis*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 2(2001), sivut 97-115.
- Coh02 Cohn, J. et al., *Facial symmetry quantification for expression invariant human identification*. *Computer Vision and Image Understanding Journal*, 91, 1 (2003), sivut 138-159.
- Don99 Donato, G. et al, *Classifying facial actions*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 10(1999), sivut 974-989.
- EcF78 Eckman, P. ja Friesen, W., toimittajat, *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press, Mountain View, California, USA, 1978.
- EsP97 Essa, I. ja Pentland, A., *Coding, analysis, interpretation and recognition of facial expressions*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 7(1997), sivut 757-763.
- MaM00 Maio, D. ja Maltoni, D., *Real-time face location on gray-scale static images*. *Pattern Recognition* 33, 9 (2000), sivut 1525-1539.
- TuP91 Turk, M. Ja Pentland, A., *Eigenfaces for recognition*. *Journal of Cognitive Neuroscience* 3, 1(1991), sivut 71-86.
- Zha03 Zhao, W. et al., *Face recognition: a literature survey*. *ACM Computing Surveys*, 35, 4(2003), sivut 399-420.

Liite A: Esimerkki toimintayksikön määrittelystä

Tässä liitteessä on esimerkki FACS–koodaukseen liittyvän toimintayksikön 15 (Suupielen painaminen alas) määrittelystä [EcF78]. Muiden toimintayksiköiden määrittelyt ovat samankaltaisia. Liitteen tarkoitus on havainnollistaa, kuinka epäformaaleja nämä määrittelyt ovat.

1) Lihasryhmä:

Lihaskunta, joka vastaa AU 15:stä toiminnasta on *Triangularis*, joka alkaa leuan sivusta ja kulkee ylöspäin kasvoilla kiinnittyen lähelle suupielä. *Triangularis* vetää supistuessaan suupielä alaspäin.

2) Vaikutukset kasvojen ilmeeseen:

- (1) Vetää suupielet alas.
- (2) Muuttaa suun muotoa siten, että huulet kaartuvat reunoilta alaspäin. Usein AU 15 leventää hieman suuta.
- (3) Aiheuttaa lievää rypistymistä tai kurtistumista ihossa suupielten alapuolella. Tämä ei välttämättä näy AU 15:stä ollessa lievää, mutta vahvassa AU 15:stä aina.
- (4) Saattaa tasoittaa tai aiheuttaa pullistumia suun alapuolelle, leuan keskelle. Saattaa tuottaa painaumuksia alahuulen alapuolelle.
- (5) Mikäli nenän ja suun välissä on pysyvä vako, se syvenee.

3) Kuinka AU 15 muodostetaan kasvoille:

Paina suupieläsi alas. Varo nostamasta alahuulta samanaikaisesti – älä käytä AU 17:stä (leuan kohotus). Jos et onnistu tässä, aseta sormesi suupielten päälle ja vedä alaspäin, huomioiden muutokset kasvojen ilmeessä. Irroita sormet suupielistä ja yritä pitää yllä samaa kasvojen ilmeä.

4) Minimivaatimukset AU 15:stä määrittämiselle:

Suun leventäminen on irrelevanttia, sillä se voi johtua myös AU:sta 20 (huulten pingotus leveälle).

- (1) Jos huuliviiva on henkilön neutraalissa ilmeessä suora tai hieman ylöspäin kaartuva, suupielten on oltava vähintään hieman alaspäin kaartuneet, jotta AU 15:stä voi katsoa aktiiviseksi.
- (2) Jos huuliviiva on neutraalissa ilmeessä hieman alaspäin kaartuva, suupielten on oltava selvästi alaspäin kaartuneet, ja tämä ei saa johtua AU 17:stä (leuan kohotus) tai AU 20:stä (huulten pingotus leveälle).

Kuva 1: Toimintayksikön 15 (suupielen painaminen alas) määrittely.

Liite B: Lineaarialgebraan liittyviä käsitteitä

Tässä liitteessä on esitelty joitakin käsitteitä ja nimityksiä, joita tutkielman luvussa 4 käytetään.

Vektori $x = [x_1, x_2, \dots, x_n]$, $x_i \in R$, $x \in R^n$ on järjestetty jono elementtejä. Tässä tutkielmassa tarkastelemme reaalisia vektoreita, joiden elementit ovat reaalilukuja.

Vektorien *linearikombinaatio* vakoilla kerrottujen vektorien summa: $ax + by + cz + \dots$, missä $a, b, c \in R$ ovat *skalaareita* ja $x, y, z, \dots \in R^n$ ovat vektoreita.

Vektoriavaruus koostuu joukosta *kantavektoreita*, sekä kaikista niistä vektoreista, jotka voidaan esittää näiden kantavektorien lineaarikombinaationa, millä tahansa skalaareilla. Kantavektorien määrää sanotaan vektoriavaruuden *dimensioksi*. Kantavektorit eivät ole yksikäsitteisiä.

Vektoreihin liittyvät laskutoimitukset ovat sitä raskaampia, mitä suurempi dimensio tarkasteltavalla vektoriavaruudella on. Kantavektorien sanotaan *virittävän* vektoriavaruuden.

Vektorialiavaruus on vektoriavaruus, joka sisältyy kokonaisuudessaan johonkin toiseen vektoriavaruuteen. Vektorialiavaruuden dimensio on aina enintään yhtä suuri kuin sitä ympäröivän vektoriavaruuden dimensio. Vektorin *projisointi* vektoriavaruudelle tarkoittaa vektorin kertomista vektoriavaruuden kantavektoreilla.