Fejmozgás Alapú Gesztusok Felismerése

Bertók Kornél, Fazekas Attila

Debreceni Egyetem

Informatikai Kar

Debreceni Képfeldolgozó Csoport

H-4010 Debrecen, Pf.:12.

[bertok.kornel@inf.unideb.hu](mailto:bertok.kornel@inf.unideb.hu), [attila.fazekas@inf.unideb.hu](mailto:attila.fazekas@inf.unideb.hu)

**Absztrakt.** Jelen cikk témája egy fejmozgás alapú gesztusfelismerő rendszer, mely segítségével lehetőségünk nyílik mozdulatsorok valósidejű felismerésére és megértésére, azok rögzítésére és később adatbányászati eszközökkel történő elemzésére, valamint a már rögzített mozdulatsorok segítségével a felismerés online javítására.  
Ugyanakkor az elkészült rendszer illeszkedik egy multimodális ember-gép kommunikációt leíró modellbe is, mert használata új fejezetet nyit a metakommunikációhoz tartozó csatornák vizsgálatában.

**Kulcsszavak:** fejmozgás, gesztus felismerés, mozgás reprezentáció, dinamikus idővetemítés.

1. Bevezetés

Az ember-számítógép interakció kutatási feladatai közé tartozik, hogy olyan új, esetlegesen alternatív kommunikációs eszközöket és módszereket fejlesszen, amelyek segítik az ember-gép kapcsolatot az ember számára minél természetesebbé, magától értetődővé tenni. A különböző eszközök és programok vezérlésére sokféle megoldás létezik. Csakhogy az eszközök és programok számának növekedésével a különböző vezérlő megoldások száma is növekszik. Tehát mindenképpen szükséges lehet egy természetesebb, eszköz-független módot találni az irányítására. A kommunikáció egyszerűsítésével kapcsolatos ötleteket célszerű a mindennapi életünkben keresni.

A szóbeliség (verbális jelek halmaza) az emberi kommunikáció legtipikusabb módja, jelentős információhordozó. Ugyanakkor a gyakran lehet félreértések forrása, mivel azzal a feltételezéssel élünk, hogy egy-egy szó azonos jelentéssel bír mindenki számára. Pedig azt, hogy egy-egy szónak az adott pillanatban milyen jelentést tulajdonítunk, aktuális szükségleteink is jelentős mértékben befolyásolják. Ezért az egyes kommunikációs szituációkat kontrollálni kell.

A verbális jelek mellett a szóbeli információk kiegészítésére, ellenőrzésére vagy éppen hangsúlyozására a nem szóbeli, ún. non-verbális jelrendszert alkalmazzuk. A non-verbális jelek tipikus megnyilvánulásai a mimika, a tekintet – szemkontaktus – szemmozgás, az ún. vokális jelek, mint hangnem, hanghordozás, hangerő, hangszín; a gesztusok, a testtartás és a távolságtartás-térközszabályozás.

Jelen tanulmány a gesztusok, mint non-verbális jelek felismerésére korlátozódik. Gesztusok alatt értjük a fej, a kéz és a karok mozgását. A fejmozgások gyakoribb jelentései: az igenlés, a tagadás, a helytelenítés, a megszégyenülés, elszomorodás stb. A kéz- és karmozgások jelentése: a hívás, elutasítás, tiltakozás, kérés, könyörgés, fenyegetés, köszöntés stb. A gesztusokat a partner beszédének szabályozására (magyarázás, gyorsítás-lassítás stb.) is használjuk. E mozgásoknak jelentésük van, egy részük tudatos, másik felük öntudatlan.

Jelen cikk témája egy gesztusfelismerő rendszer ismertetése, mely segítségével lehetőségünk nyílik tudatos fejmozgások felismerésére és megértésére, azok rögzítésére és később adatbányászati eszközökkel történő elemzésére, valamint a már rögzített mozdulatsorok segítségével a felismerés online javítására.

* 1. Irodalmi áttekintés

A meglévő gesztusfelismerő rendszerekkel kapcsolatban egy áttekintő összefoglalót ismertet a [23] tanulmány. Ebben az alfejezetben csak néhány olyan munkát foglalunk össze, melyeket az előző összefoglalón kívül alaposabban tanulmányoztunk.

A fejmozgás alapú – vagy általánosságban csak a mozgás alapú – gesztusfelismerő eljárások két csoportba oszthatóak: a modell és minta alapú módszerekre. A modell alapú eljárások csoportjába a különböző rejtett Markov modellek (HMM) és azok variánsai. Marcel et al. [2] egy input-output HMM-et készített EM használatával, majd a kézfej körvonalából kinyerhető gesztusok felismerésére alkalmazta. A szakirodalomban megjelent a tradicionális HMM néhány javítása is, melyek pl. szemantikus hálókat foglalnak magukban (SNM) [3], vagy a nem-paraméteres HMM-ek [4], vagy a HCRD (Hidden Conditional Random Field) [5]. Ezen variánsok egyszerre csökkentik a tanítás költségét és az osztályozás pontosságát.

Másik népszerű modellek a véges állapotú gépek [6,7], a dinamikus Bayes hálók [8], és a topológia őrző önszervező hálózatok [9] tartoznak. Ezen eljárások feltételezik, hogy a fej mozgásának trajektóriája és ez által az artikuláció ismert. Habár ezekkel az eljárásokkal ígéretes eredményekre lehet szert tenni, a robusztusságuk nagyban függ az arc detektálásának és a mozgás követésének sikerességétől. Továbbá használatukat megelőzően sok adatra és számításigényes eljárások alkalmazására van szükség.

Ezzel szemben a minta alapú eljárások alkalmazásával elkerülhető a modell alapú módszerekben rejlő nehézségek nagy része. Mindez az egyes gesztusok vizuálisan invariáns reprezentálásával és azok közvetlen illesztésével érhető el. A meglévő módszerekben leggyakrabban tér-, és időbeli jellemzőket valamint leírókat használnak [10, 11, 12, 25, 26, 27, 15]. Laptov et al. [10] megalkotta a referencia eljárást az irányított gradiensek és optikai áramlás hisztogramjai (HOG, HOF) – mint leírók – alkalmazásával az érdekes pontok kinyerésére (STIP) és a gesztusok felismerésére.

Egyes leírók magukban foglalják a mozgás trajektóriáját [13], tér-, és időbeli gradienseket [14], és az optikai áramláshoz tartozó globális hisztogramokat [15]. Ezen eljárások legnagyobb hátránya, hogy a futás során közvetlenül illesztik az egyes gesztusokat egy már meglévő adatbázisra, mely rontja az eljárások skálázhatóságát.

1. Fejmozgás reprezentálása

Ebben a fejezetben ismertetjük részletesen az általunk kifejlesztett minta alapú gesztusfelismerő rendszert. Mindemellett megadunk egy hatékony vizuális reprezentációt a mozgást meghatározó jellemzők kinyeréséhez, amely elengedhetetlen a felismerő rendszer nagyméretű gesztus adatbázison történő használatát illetően. Ennek kapcsán bevezettünk egy új és hatékony vizuális reprezentációt a fejmozgásból kinyerhető gesztusok felismerésére vonatkozóan, mely a mozgás menetét ábrázoló képen alapul. Ezen a képen egy egyszerű FAST sarokdetektorral meghatározzuk azokat a régiókat, melyeken a mozgás a legmeghatározóbb volt. Majd egy adott gesztus sorozat minden szomszédos képkockájára kiszámoljuk az előbb kinyert régiókhoz tartozó optikai áramláshoz tartozó vektorokat és ezek alapján a globális fejmozgáshoz tartozó irányvektorokat. Ennek eredményeként egy-egy darab irányvektort kapunk a gesztus sorozat minden szomszédos képkocka párjára. Végezetül a gesztus sorozathoz tartozó irányvektorok sorozatát dinamikus idővetemítés segítségével egy előre definiált gesztus adatbázis elemeihez hasonlítjuk.

* 1. Mozgás ábrázolása

A mozgás megjelenítésére azt az irodalomban előszeretettel használt módszert alkalmaztuk, mely egy képet hoz létre a mozgás történetére vonatkozóan (Motion History Image - MHI) [REF]. Ez egy időalapú sablonozó eljárás, mely nagyon egyszerű, de ugyanakkor robusztus reprezentációt szolgáltat a mozgó objektumokra.

Rengeteg variánsa létezik és a szakirodalomban szinte megszámlálhatatlan tanulmányt találunk a felhasználására [REF]. Ebben az alfejezetben csak a módszer lényegét ismertetjük. Bobick és Davis [REF] vezette be először azt a reprezentációt a mozgás alapú gesztusok felismerésében mely külön írja le, hogy „hol” és hogy „hogyan” történik a mozgás a képen. Egy úgynevezett mozgási energiát ábrázoló bináris képet (MEI) alkottak meg, mely arra vonatkozóan tartalmaz információt, hogy hol volt mozgás egy kép szekvencián. A MEI lényegében a mozgás alakját és térbeli felosztását írja le.

A metódushoz szükség lesz még egy MHI sablon létrehozására is, amelyben minden egyes pixel a mozgásnak egy sűrűségfüggvényeként értelmezhető az adott helyen. A MEI és MHI sablonok együttesen egy kétkomponensű, az időtől is függő sablonként – vektor értékű képként – értelmezhetőek, amelyben minden egyes pixel értéke a mozgás egy függvénye az adott pixel helyén. Az eljárás az alábbi (1) képlettel számolható:

, (1)

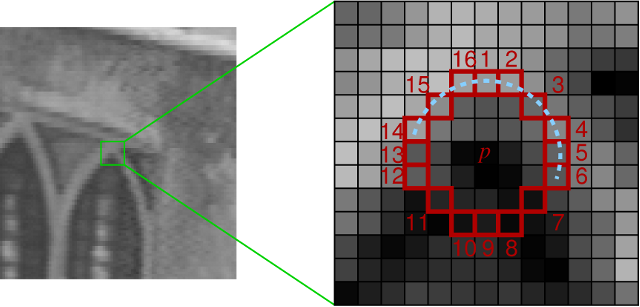
Ahol függvény egy bináris maszk, mely nem-nulla értékű pixeleket tartalmaz ott ahol mozgás volt a képen, a kép szekvencia aktuális időbélyege, pedig a maximális értéke a mozgáskövetésnek. Vagyis az -val jelölt kép összes olyan pixele, ahol mozgás volt értéket fog felvenni, még azok a részek ahol nem volt mozgás fokozatosan elhalványulnak és végül törlődnek. Az eljárás grafikus reprezentációját az 1. ábra mutatja. Az MHI eljárást a mozgás szegmentálásában is felhasználtuk. Maga a gesztus képkockák sorozata, vagyis a videó folyam szegmensei alatt fog megjelenni. A szegmensre teljesülnie kell a következő feltételnek: olyan nyitó és záró képkockák tartoznak hozzá, melyeknek a számított MHI képek átlagintenzitás értéke alacsonyabb, mint egy előre definiált küszöbérték. Vagyis olyan képkockák határolják a szegmenset, ahol a mozgás intenzitása alacsony volt.

1. ábra. A bal oldali ábrán egy mozdulatsorhoz tartozó MEI sablon látható, a jobb oldalin pedig a hozzá tartozó MHI.

* 1. A mozgást meghatározó régiók

Következő lépésben megkeressük azokat a régiókat az MHI-n, melyen meghatározóak a fejmozgásban. Erre a FAST algoritmust használtuk, mely egy egyszerű sarokdetektor – jellemző pontok kinyerésére. Hatékonysága az alacsony számításigényében rejlik, mely által valós idejű feldolgozásra alkalmas. Veszi a kép minden egyes pixelét, melyeknek egy adott sugarú környezetében vizsgálja a többi pixel értékét, lásd 2. ábra. Ha a környezetben szerepelő intenzitás értékek jelentősen nagyobbak, vagy kisebbek, mint a középpont, akkor azt sarokként osztályozza. Általában sarkok egy halmazát találja meg egy szűkebb környezetben, ezért szokás egy metrikát alkalmazni a sarkok erősségének mérésére. Általában egy kétmenetes algoritmusként implementálják, mely rendkívüli gyors számítást tesz lehetővé.



2. ábra. A FAST detektor által vizsgált tartomány egy potenciális sarokpont esetén. A FAST detektor is egy vizsgált pont körüli kör mentén – például egy 3 sugarú, 16 kerületű – vizsgálódik, ha ebből valahány – például 9 – eltér a pixelnél legalább egy küszöbbel magasabb értékkel, akkor az adott középpont egy jellemző pont.

* 1. Optikai áramlás

A FAST algoritmus jellemzőpontok egy halmazával fog visszatérni a MHI-n. A következő lépésben ezen jellemzőpontokra számítjuk ki az optikai áramláshoz tartozó vektorokat az aktuális képkockára és arra, amelynek az időbélyege megegyezik az MHI időbélyegével, vagyis az aktuális képkocka megelőzőjével.

Az optikai áramlás (Optical Flow, OF) meghatározása lényegében nem más, mint több képen azonos képrészletek megfeleltetése. Az eredmény egy vektormező, amely az elmozdulásokat, vagyis a sebességvektorokat tartalmazza. Az optikai folyamon tehát azt értjük, ahogy a képintenzitások mozgása megjelenik egymás utáni képeken. Különböző típusú képbemenetekhez az egyes optikai folyam algoritmusok más-más eredményt adhatnak, ezért célszerű a bemeneti adatok milyenségének figyelembevételével választani a lehetséges algoritmusok közül, hogy a kapott vektormező minél jobban közelítse a képeken látható objektumok valós fizikai mozgását. Inputként a videó egymáshoz közeli képkockáit szokás megadni.

Az optikai folyam algoritmusok az összetartozó képpontok megtalálásához feltételezik, hogy ezek intenzitása közel megegyezik. Szinte az összes módszer alapját ez a feltételezés adja, amit optikai folyam korlátozásként ismerünk. Jelölje egy adott pillanatban a képintenzitást, amely egy időben változó képsorozatból származik. A továbbiakhoz a következő feltételezéssel élünk: a mozgó vagy álló objektumok pontjainak intenzitása (lényegében) nem változik az idő múlásával. Legyen néhány objektum a képen, ami idő alatt (a gyakorlatban egymás utáni képvétel alatt) elmozdul távolságra. Az intenzitásértékek Taylor-sorba fejtésével és az előbb feltételezett állítások felhasználásával kapjuk:

(2)

Ezt a kifejezést rendszerint az optikai folyam feltételi egyenletének (vagy csak optikai folyamkorlátozásnak) nevezik, ahol és az optikai folyammező és koordináta irányú összetevői. Az egyenlet két ismeretlent tartalmaz. A megoldásra, a teljesség igénye nélkül az alábbi technikák a legelterjedtebbek.

A differenciális módszerek

Régebbi technológiák, de megbízhatóak. Az újabbaknak nem sikerült jelentős minőségjavulást hozniuk. Ilyen pl. a Horn–Schunck féle módszer [REF], mely ugyan az objektumok határát nem kezeli és egy úgynevezett egyenletességi korlátozással egészíti ki az optikai folyam egyenletet. Felteszi, hogy a szomszédos pontok közel azonos sebességgel mozognak, így nincs hirtelen ugrás bennük. Ezek alapján olyan vektorokat keresünk, amelyek egymáshoz képest a legkevesebbet változnak (és persze kielégítik az optikai folyamkorlátozást).

Másik népszerű módszer a Lucas–Kanade módszer, mely ellentétben az előző eljárással egy pont sebességére úgy tekint, hogy az csak a pont helyi környezetétől függ (lokális technika). Másképpen, a kiindulási feltétel az, hogy a kép kisméretű szegmenseiben a sebesség állandó értékű, így egyetlen sebességvektorhoz több egyenlet is keletkezhet. Az optikai folyam korlátozás kiegészítése a konstans lokális sebesség módszerével azt a feltételt jelenti, hogy a vizsgált pixel és a szomszédságában lévő képpontok sebessége ugyanaz. A szakirodalomban gyakran használják ennek a módszernek és ennek piramisos változatát is [REF].

Korrelációs technikák

A módszer elve [REF], hogy a videó szomszédos képkockáin olyan blokkokat keres, melyek pixelei nagyjából megegyeznek. A blokkok egymással fedésben lehetnek, és ha két egymást követő képkockán az algoritmus megtalálta ugyanazt a részt, akkor az ezek közti elmozdulás adja az optikai folyam vektorát. Az algoritmus feltételezi, hogy az adott blokkon belül az összes pixel ugyanazzal a vektorral mozdult el. Mivel a módszer az elmozdulásokat csak a blokkok közt keresi, azért így határozza meg minden egyes pixelre az optikai folyam vektort. Minél nagyobb a blokkméret, annál kevesebb optikai folyam vektort kapunk az adott képre. Az algoritmus nagyobb blokkméret esetén kevésbé érzékeny a zajra, és pontosabb, mivel egy vektor előállításához több pixelt vizsgál.

* 1. Fejmozgás iránya

Mint ahogy az előző alfejezet elején már említettük a bejövő videó folyamon a FAST sarokdetektor segítségével és a MHI képhez meghatározunk egy maszkot, melynek a fehér színű képpontjai felett kiszámoljuk az optikai áramláshoz tartozó vektorokat minden egyes képpárra. Az optikai áramláshoz tartozó vektorokat a Lukas–Kanade eljárás GPU-ra optimalizált változatával határozzuk meg.

[KÉP: FAST maszk]

A gesztus felismeréséhez szükség van egy reprezentációra mellyel egy adott szegmenshez tartozó fejmozgás leírható az optikai folyamhoz tartozó vektorok – melyet a szegmens minden képpárjához kiszámolunk – függvényében. Ha a FAST eljárás darab jellemzőpontot detektált, akkor ugyanennyi optikai folyam vektor fog keletkezni egy képpár között. A gesztusok könnyebb definiálása és felismerése érdekében meghatározzuk egy átlagvektort, melyet az darab optikai folyam vektor számtani közepeként kapunk meg. Meghatározzuk az átlagvektornak az tengely pozitív oldalával bezárt szögét a (3)-as egyenlet segítségével.

. (3)

Ahol függvény és által meghatározott érték arkusz tangensét adja vissza radiánban. Ez hasonló az arkusz tangenséhez, attól eltekintve, hogy a paraméterek előjele meghatározza, hogy az eredmény melyik körnegyedbe esik. A (4)-es egyenletek segítségével az eredményt átváltjuk szögbe és gondoskodunk arról, hogy a koordinátarendszerünk jobbsodrású legyen.

(4)

Így a vizsgált szegmens minden egyes képpárjára kapunk egy szöget, melyekhez értékétől függően egy címkét rendelünk az (5)-ös formulában definiált függvény segítségével.

. (5)

Az (5)-ös formula segítségével tulajdonképpen annyit csinálunk, hogy minden egyes szöget a teljes szög azon tizenhatodába soroljuk be, amelyikbe esik. Így a videó folyam . szegmense alatt megjelenő gesztus leírható az sorozattal, ahol az . képpárhoz számított átlagvektornak az tengely pozitív oldalával bezárt szöge.

[KÉP: mozgás szöge]

1. Gesztusfelismerés

Ebben a fejezetben egy eljárást ismertetünk, az előző fejezetben definiált címkesorozatok egymáshoz történő illesztésére vonatkozóan. Az illesztéshez szükség lesz egy előre definiált gesztus adatbázisra. Ebben az adatbázisban olyan elempárokat tárolunk, melyekben benne van a gesztus neve és az ahhoz egy előzetesen meghatározott címkesorozat pl. (6)-os formula, melyben szögletes zárójelek között soroljuk fel az adott címkesorozatot.

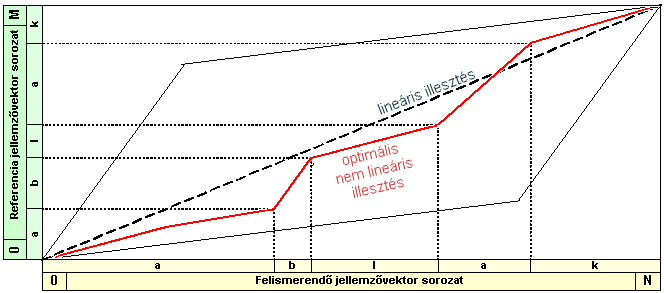
. (6)

A gesztusfelismerés során folyamatosan szegmentáljuk a videófolyamot és számítjuk a szegmensekhez tartozó címkesorozatokat. Az egyes címkesorozatokat a dinamikus idővetemítés (DTW) eljárás segítségével illesztjük az előre definiált gesztus adatbázis elemeihez. Az adatbázisban egy gesztushoz több sorozat is létezhet, így az adatbázison belül a gesztusok csoportokat alkotnak. Végezetül a szegmenshez tartozó címkesorozatot ahhoz a gesztus-csoporthoz fogjuk besorolni, melyhez a DTW átlagosan a legkisebb távolságot adta.

* 1. Dinamikus idővetemítés

A DTW feladata hogy, azonos időtengelyre vetítse az aktuálisan detektált és a tárolt fejmozgást, hogy címkesorozatot összevethessük a tárolt referenciákkal (gesztus-csoportokkal). Az összehasonlításhoz definiálni kell egy távolságot. A DTW algoritmusa lényegében egy dimenziós vektort illeszt egy dimenziós felismerendő vektorhoz. Az illesztés során a kezdőpontból a végpontba kell eljutni. Közben az útvonalkereső algoritmus lépésenként haladva a mintákat (vektorokat) egymással összehasonlítja, és a távolság minimalizálására törekszik. Az eljárás során a felismerendő címkesorozatot az összes gesztus-csoport összes referenciamintájával össze kell hasonlítani, és a legkisebb távolságú elem lesz a felismerés eredménye.

A két vektor távolságát többféleképpen számíthatjuk ki, tapasztalataink azonban azt mutatták, hogy a leggyakrabban használt módszerek közül az euklideszi távolság (ami a tagok különbségének négyzetösszegét jelenti) biztosítja a leghatékonyabb összehasonlítást, ezért a programunk is ezzel a távolsággal dolgozik. A 11. ábrán a szaggatott vonal jelzi azt az utat, amely mentén haladva egyenletesen nyújtjuk, ill. zsugorítjuk a bemenő vektorsorozatot az összehasonlításhoz. Ez a lineáris idővetemítés.



3. ábra. Az „ablak” szó illesztése

A vetemítőgörbe (az ábrán a piros lépcsős vonal) tulajdonságai közé tartozik, hogy mindig monoton növekvő, lokális korlátok jellemzik és hogy lokális optimumokon keresztül elért teljes optimum. A vetemítés útvonala tehát nem lehet tetszés szerinti. Nem haladhat visszafelé. Ezen kívül az előre haladást is sokféleképpen korlátozhatjuk, attól függően, hogy mekkora ingadozást engedünk meg az illesztés vonalán.

* 1. Mozgás adatbázis
  2. Adatbázis online bővítése

1. Kísérletek és eredmények

**Table 1.** Font sizes of headings. Table captions should always be positioned *above* the tables.

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
| Title (centered) | **Lecture Notes …** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Headings.** Text follows … | 10 point, bold |
| 4th-level heading | *Remark.* Text follows … | 10 point, italic |

References

1. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197 (1981)

2. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)

3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)

4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001)

5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

6. National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov