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 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 pl. véges állapotú gépek [REF], dinamikus Bayes hálózatok [REF], és a topológia őrző önszervező hálózatok [REF] 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.

Ugyanakkor 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 használnak [REF]. Egyes leírók magukban foglalják a mozgás trajektóriáját [REF], tér-, és időbeli gradienseket [REF], és az optical flow-hoz tartozó globális hisztogramokat [REF]. Ezen eljárások legnagyobb hátránya, hogy a futás során közvetlenül illesztik az egyes gesztusokat egy már meglévő adatbázisra, mely rontja az eljárások skálázhatóságát.

Ebben a tanulmányban egy minta alapú gesztusfelismerő rendszert ismertetünk. Mindemellett definiál 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 meghatározzuk az előbb kinyert régiókhoz tartozó optical flow vektorokat és ezek alapján a globális fejmozgáshoz tartozó irányvektorokat. A gesztus sorozathoz tartozó irányvektorok sorozatát egy előre definiált gesztus adatbázis elemeihez hasonlítjuk.

1. Fejmozgás alapú gesztusok reprezentálása
   1. Mintázatok a mozgásban
   2. Optical Flow
   3. A mozgást meghatározó régiók
2. Gesztusfelismerés
   1. Dinamikus idővetemítés
   2. Mozgás adatbázis
   3. Adatbázis online bővítése
3. 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 |



**Fig. 1.** One kernel at *xs* (*dotted kernel*) or two kernels at *xi* and *xj* (*left and right*) lead to the same summed estimate at *xs*. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption.

|  |  |
| --- | --- |
| x + y = z . | (**1**) |

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