

Multi-modális gépi sakkozó – Török-2

KOVÁCS GYÖRGY, SAJÓ LEVENTE, FAZEKAS ATTILA

Debreceni Egyetem Informatikai Kar, Debreceni Képfeldolgozó Csoport
{attila.fazekas, sajolevente, gykovacs}@inf.unideb.hu

Lektorált

Kulcsszavak: multimodális ember-gép interakció, arci érzelem felismerés, arc detektálás

Az információs rendszerek használatának egy új módját jelenti a multi-modális ember-gép kommunikáción alapuló technikák használata. Ebben a cikkben egy ilyen technikán alapuló gépi sakkozó megvalósítását ismertetjük. Az ember-gép kommunikáció multi-modális mivolta abban nyilvánul meg, hogy billentyűzet nélkül, a verbális és a gesztusnyelv csatornáit is felhasználva sakkozhatunk a géppel.

1. Bevezetés

Az információs társadalom egyik alapvető igénye az információkhoz való hatékony hozzáférhetőség biztosítása. Az elmúlt évtizedben a technológia fejlődése egyre hatékonyabb eszközöket adott a kezünkbe az információ tárolására, rendszerezésére és lekérdezésére, továbbá az élet számos területén megjelentek olyan eszközök, amelyek feladata a tárolt információk lekérésének biztosítása. Ezen eszközöket gyűjtőnéven információs rendszereknek nevezzük. Ebbe a kategóriába tartoznak az általános célú számítógépek, de azok a speciális számítógépek is, amelyek valamilyen jól definiált célra készültek, mint például a menetrendekkel kapcsolatos információk tárolása, visszakeresése és az esetleges foglalások, illetve jegyek on-line vásárlásának lebonyolítására.

Annak ellenére, hogy az információs rendszerek életünk számos területén jelen vannak, a használatukkal szemben egyfajta ellenállás figyelhető meg. Ez egyrészt a technológia használatához szükséges ismeretek hiányának, másrészt az egyes rendszerek használati módja közötti különbségeknek tudható be. Ez azt jelenti, hogy életünk minden pillanatában újabb és újabb ismereteket kell elsajátítanunk az információs rendszer használatához. Ráadásul az információs rendszerek egy-egy új generációjának megjelenése egyre rövidebb idő alatt következik be és életünk egyre több területén hódítanak teret maguknak.

Az információs rendszerek használatának módja egyfajta kommunikációs nyelvnek a használatához hasonlít. Ha sok, különböző információs rendszerrel való kommunikációhoz szükséges nyelvet ismerünk, akkor az egy-egy újabb ilyen kommunikációs nyelv elsajátításához szükséges idő egyre kevesebb lesz. Továbbá, ha egy információs rendszer kommunikációs nyelvét napi szinten használjuk, akkor általában könnyebb lesz elsajátítanunk egy újabb generációjának a kommunikációs nyelvét.

A fentiek alapján teljesen nyilvánvaló, hogy jelenleg minden esetben a felhasználónak kell megtanulnia az adott információs rendszer kommunikációs nyelvét, azaz

azt, hogy milyen formában adhatunk utasításokat a rendszernek, illetve milyen formában kapjuk meg a rendszerben tárolt információkat. A multi-modális ember-gép kommunikációval kapcsolatos kutatások alap gondolata az, hogy találjuk meg annak a módját, hogy a jövő információs rendszerei képesek legyenek a felhasználóval a számukra legtermészetesebb módon kommunikálni, azaz lehetővé tenni az emberi nyelv használatát. Egy felhasználónak nem kell újabb és újabb kommunikációs nyelvet elsajátítania, hanem elegendő „szóba elegyednie” az adott rendszerrel.

Minden szempont alapján ideális multi-modális ember-gép kommunikációs rendszer még nem készült el. Ezzel is magyarázható, hogy az ipari fejlesztések területén a kivárák a jellemző. A jelenleg hazánkban elérhető technológiák képességeinek bemutatása érdekében döntöttünk úgy, hogy elkészítjük a multi-modális gépi sakkozónkat, amelyet – tisztelegve Kempelen Farkas sakkozógépe előtt – Török-2-nek neveztünk el.

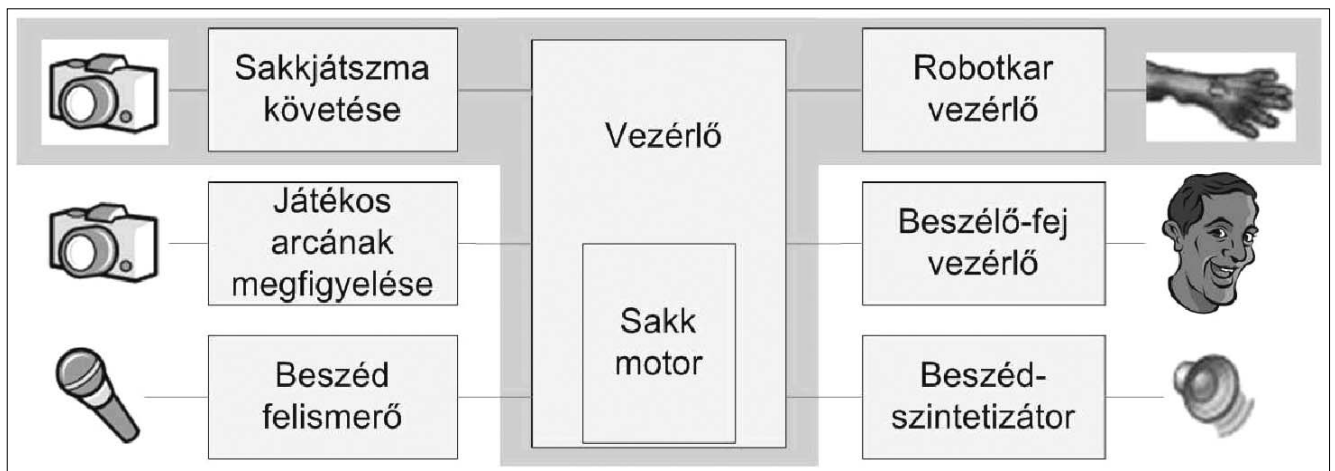
E cikk célkitűzése az, hogy az olvasó számára röviden vázolja a Török-2 felépítését és az egyes komponensek funkcionális szerepét. Áttekintjük az alapkonceptiókat, a rendszer működését, majd a rendszert alkotó egyes komponensek működését, funkcióját mutatjuk be.

A multimodális ember-gép kapcsolaton alapuló rendszerek felhasználói értékelése interdiszciplinális feladat. A rendszer használata közben készült videókat nyelvészek és pszichológusok értéklik ki a felhasználó kommunikációs tevékenységére, gesztusaira, a rendszerbeli elmélyülésére összpontosítva.

2. A Török-2 általános felépítése

A Török-2 rendszer komponenseit és a komponensek közötti kapcsolatot a következő oldali, 1. ábra szemlélteti [3].

A rendszer koncepciója, hogy egy virtuális játékost valósítsunk meg, aki a játék szempontjából a lehető legtöbb tekintetben emberként viselkedik. A virtuális sakkozó rendszert funkcionálisan két fő részre bonthatjuk:



1. ábra A Török-2 rendszer általános felépítése

a sakkjátszma lebonyolítását biztosító modulra, valamint az ember-gép kommunikációt szolgáló felület modulra.

A sakkjátszmáért felelős modul szintén több kisebb komponensből épül fel. Mivel a játszma egy valódi sakk-táblán zajlik, szükségünk volt egy eszközre, amely a virtuális játékos karját helyettesíti, azaz képes sakklépéseket végrehajtani. Ezt egy négy szabadsági fokkal rendelkező robotkar végzi. A modul bemeneti interfésze egy webkamera, amely a sakk-tábla fölött helyezkedik el. Ez a kamera felelős a játékos felismerésért, a játszma követésért.

Az ember-gép kommunikációt megvalósító komponens több emberi kommunikációs csatornát felhasználva, multi-modális kapcsolatot tesz lehetővé. A bemeneti interfészek hardver elemei: a webkamera, amely az emberi játékos arcát figyeli és egy mikrofon a játékos hangjának rögzítésére. A webkamera interfészt képfeldolgozási módszereket megvalósító szoftver egészíti ki, mely képes felismerni, hogy ül-e játékos a sakk-tábla előtt, vagy sem, felismeri a játékos nemét, életkorát és a játék alatt a játékos arcán megjelenő érzelmeket figyeli. A rögzített hangot a beszéd felismerő szoftver dolgozza fel, és a játékmenet vezérlésével kapcsolatos kulcsszavakat detektál.

Az ember-gép kommunikáció kimeneti interfészének hardver egysége a hangszóró mellett a monitor, a szoftver komponense pedig a szövegfelolvasó szoftver mellett az érzelmelek kifejezésére is képes, animált beszélő fej [4].

2.1. A sakkozógép felépítése

A sakkozógépünket a mechanikus robotkar, mint kimeneti interfész, a sakkállás-felismerő, mint bemeneti interfész, egy sakkmotor, valamint a vezérlő egység alkotják.

2.1.1. A vezérlő

A vezérlő realizálja a virtuális játékos tudatát, vagyis a játék aktuális állapotának megfelelően vezérli és szinkronizálja az egyes komponenseket: sakkozógép esetén a robotkar és a sakkállás-felismerő megfelelő időzítése kulcsfontosságú.

2.1.2. A sakkállás-felismerő

A sakkállás-felismerő modullal szemben támasztott elsődleges követelményünk volt, hogy valós időben működjön. Ehhez úgy alakítottuk ki a fizikai környezetet, hogy az ideális legyen a sakkállás optikai felismeréséhez anélkül, hogy a sakk-táblát, vagy a figurákat megváltoztatnánk. Ezzel sikerült elkerülnünk a költséges számításokat.

A webkamera a tábla fölött, középen helyezkedik el, olyan magasságban, hogy a perspektívából adódó geometriai torzítások elhanyagolhatóak legyenek. Mivel a figurák felülnézetből nem különböztethetők meg, a sakkállás-felismerő csak egy kiindulási állapothoz relatívan bekövetkező változásokat tudja detektálni.

A sakkfigurák elhelyezkedésének robusztus felismerésére két különböző módszert használunk, az első módszerrel a képet élképpé alakítjuk, azaz kiemeljük a lokális intenzitáskülönbségeket, majd ezt követően az úgynevezett Hough-transzformációval a sakkfigurák a priori ismert méretének megfelelő köröket keresünk az egyes mezőkben. A második módszerrel az egyes mezők lokális hisztogramjai alapján következtetünk arra, hogy van-e figura az adott mezőn, vagy sem. A két módszer együttesen megbízhatóan működik. [2]

Miután tudjuk, hogy mely mezőkön van figura, meghatározzuk azok színét az őket tartalmazó mezők középső régióinak világosságértékeiből.

A sakkállás-felismerő komponens a sakkállás felismerésén kívül képes érzékelni, azt, hogy a játékos mikor nyúl be a sakk-tábla fölé. Ezt az információt a virtuális játékos kezeli a következő módon: amint az emberi játékos benyúl a sakk-tábla fölé, a beszélő fej a képernyőn abbahagyja a „bámészkodást” és a sakk-táblára néz, kíváncsian várva a lépést.

2.1.3. A robotkar

A robotkar kimondottan ehhez a rendszerhez lett kifejlesztve, ennek megfelelően a következő követelményeknek kell megfelelnie:

- tudja elérni a tábla legtávolabb eső mezejét is,
- a megfelelő figurához a környező bábok érintése nélkül férjen hozzá,

- tudjon megfogni bármilyen alakú sakkfigurát,
- a bábót függőlegesen emelje föl, és tegye le,
- a bábók elhelyezését megfelelő időn belül hajtsa végre,
- lehetőség legyen a sakktáblán kívültre történő pozicionálásra is.

A robotnak négy szabadsági foka van: kettőt a váll-izület, egyet-egyét pedig a könyök és csuklóizület valószínűleg. A robot működési területe egy negyed gömb. Technikai és anyagi okok miatt a robot elektromos és mechanikai alkatrészekből áll, a meghajtást villanymotorok, az erőtovábbítást pedig bowden huzalok végzik (2. ábra).



2. ábra A robotkar

2.2. A virtuális játékos

A virtuális játékos komponens az ember-gép kommunikáció megvalósítására szolgál. Bemeneti csatornája az emberi kommunikációban is használt beszéd, valamint az arci gesztusok felismerése.

2.2.1. Érzelemfelismerés

Az emberi arc önmagában is egy információhalmaz, amelyből mi, emberek bármikor ki tudjuk nyerni az életkort, nemet és érzelmi állapotot. Ahhoz azonban, hogy mindezt fel tudjuk használni a számítógéppel történő kommunikációban, a felismerést számítógéppel kell végeznünk, ami összetett képfeldolgozási feladat. A fent említett információk (érzelem, nem, életkor) kinyerésére statisztikai tanuló algoritmusokat alkalmazunk (Support Vector Machine). A gyakorlatban azonban számos előfeldolgozási lépést kell tennünk ahhoz, hogy az arcokból információt nyerhessünk ki: meg kell találnunk az emberi játékos arcát a képen (arcdetektálás), valamint követnünk kell azt mozgás során (arckövetés), ráadásul mindezt, beleértve a tanuló algoritmusok osztályozását is, valós időben kell végeznünk.

A rendszer jelenlegi állapotában másodpercenként 2-3 érzelmdetektálást tud végrehajtani [1]. Mivel az ér-

zelmek nem váltakoznak gyorsan, a videófolyamot felhasználva a korábbi detektált érzelmelek alapján átmeneti valószínűségeket figyelembe véve a módszer még robosztusabbá tehető.

2.2.2. Beszédfelismerés

A sakkjáték tulajdonságai miatt nincs szükség teljes körű szöveg felismerésére, elegendő egyes izolált kifejezéseket azonosítani, amelyeknek az ellenfél kiválasztása illetve a játék újrakezdése esetén van szerepe.

2.2.3. Beszélő fej

A virtuális játékos komponens kimeneti interfésze a beszélő fej, amely az emberi játékoskal szemben lévő képernyőn jelenik meg. Két kimeneti csatornát használunk: a szintetizált beszédet, valamint a beszélő fejen megjelenő gesztusokat. A beszédszintetizálás a ProfiVox TTS rendszerrel történik [5], míg a megjelenő animált fej a CharToon rendszerre épült [6].

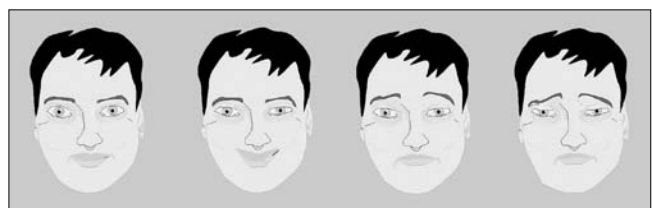
A beszédszintetizátor minden verbális megnyilvánulás esetén előállít egy hanghullám-állományt, amely a szintetizált beszédet tartalmazza, valamint egy időparaméterekkel ellátott fonéma-szekvenciát, amely azt az információt tartalmazza, hogy melyik pillanatban milyen fonéma hallható. A beszélő fej szájának animálásához definiáltuk a magyar nyelv egyes fonémáinak kiejtésükre megjelenő arckifejezéseket, ezeket a fonémához tartozó vízémának nevezzük. A beszéd animálása tehát a képernyőn megjelenő fejhez definiált vízémák lineáris interpolálása olyan módon, hogy a hangállomány párhuzamos lejátszásakor a fonéma és vízéma párok a megfelelő pillanatban jelenjenek meg.

A másik csatorna a beszélő fej esetén az érzelmelek kifejezése. Jelenleg 4 érzelmi állapotot tudunk megjeleníteni: természetes, szomorú, vidám, unott állapotokat (3. ábra). Mindemellett a fej véletlenszerűen kisebb arckifejezéseket tesz (pislogás, szájhúzogató), ezzel is élethűbbé téve a fej viselkedését. A véletlenszerű folyamatok vezérlésére a fej esetén a Perlin-zaj függvényt használjuk, mivel az más zajfüggvényeknél jobban közelíti a valós világban előforduló természetes viselkedésmintákat.

3. A rendszer működése

Ha nincs játékos a táblánál, a rendszer felismeri azt, és a beszélő fej embereket szólít meg, hogy üljenek le, és játsszanak vele. Ha valaki úgy dönt, hogy leül, azt a korának és nemének megfelelő köszöntéssel üdvözlöli.

3. ábra A beszélő fej érzelmi állapotai



Ezt követően játékost választhatunk. A virtuális játékos kiválasztása után kezdődik a játék. Az emberi játékos kezd. Lépését a sakkállás-felismerő modul felismeri és a sakkmotor meghatározza a következő lépést, melyet a robotkar elvégez. Mindeközben a játékosal szemben lévő webkamera képét feldolgozó szoftver érzelmeket detektál, ezen érzelmeknek és a sakkjátszma állapotának függvényében a beszélő fej az arcán érzelmeket kifejezve, az érzelmi állapotának megfelelő megnyilvánulásokkal kommentálja a játékot. A virtuális játékos igyekszik kapcsolatban maradni és kommunikációt folytatni az emberi játékosal annak lépése során is, sőt ha kell egy kicsit sürgetni lépése megtételére. A játék természetesen valamelyik játékos győzelméig tart, ami után egy újabb játékot ajánl fel a rendszer.

4. Összefoglalás

A rendszer tesztelése folyamatban van, eredményekről még nem tudunk számot adni. Összefoglalva tehát van egy rendszerünk, amely a multimodális ember-gép kapcsolatok lehetőségeit demonstrálja. A rendszer komponensei (érzelem-, nem-, kor-felismerés, beszélő fej, robotkar, beszédfeldolgozás) függetlenek egymástól, a vezérlő komponens módosításával újabb alkalmazásokat alakíthatunk ki.

A sakkozó tesztelésével párhuzamosan lassan elkészül a multimodális dámajátékos. A meglévő komponensek felhasználásával és újabb modulok (például kézgesztusok felismerése) fejlesztésével szeretnénk megvalósítani nem táblás játékokat is, például barkochba, kő-papír-olló.

A szerzőkről

Fazekas Attila egyetemi docens a Debreceni Egyetem Informatikai Karán, ahol digitális képfeldolgozással és alakfelismeréssel kapcsolatos kutatásokat folytat 1992 óta. PhD fokozatát Matematika és Számítástudomány területén 1999-ben kapta meg. Végzés után tudományos ösztöndíjasként (1992-1995), majd egyetemi tanársegédként (1995-2000), illetve egyetemi adjunktusként (2000-2005) dolgozott a Kossuth Lajos Tudományegyetemen (2000-tól jogutódján a Debreceni Egyetemen). Dr. Fazekas elnöke az International Association for Pattern Recognition magyarországi tagszervezetének, azaz a Képfeldolgozók és Alakfelismerők Társaságának. Közel 60 publikációja jelent meg a szakma folyóirataiban és konferenciák kiadványaiban. Közel 40 alkalommal tartott előadást különböző nemzetközi és hazai fórumokon kutatásainak eredményéről. Munkáját számos alkalommal elismerték, többek között 1997-ben a Kalmár László Alapítvány díját, 2002-ben a Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíját, 2004-ben a Debreceni Egyetem Informatikai Karának díját, 2006-ban a Matsumae International Foundation kutatási ösztöndíját kapta meg. Számos konferencia szervezője volt, szakmai folyóiratok bírálója.

Sajó Levente jelenleg a Debreceni Egyetem Informatikai Karának második éves PhD hallgatója multimodális ember gép kapcsolat és arci érzelmefelismerés témákban. Egyetemi tanulmányait 2000-ben kezdte Programtervező Matematikus szakon. Már hallgató korában bekapcsolódott a Képfeldolgozó Csoport munkájába, az elért eredményeket TDK dolgozat formájában publikálta. Sikeresen pályázott meg egy Finnországi tanulmányi ösztöndíjat, melynek következtében 2005-ös évben a tanulmányait külföldön folytathatta. Hazatérése után szerezte meg MSc diplomáját (2006), sikeresen felvételített PhD-ra és tovább folytatta az egyetemi hallgató korában megkezdett munkát. Az eddigi eredményeit az Universitas Alapítvány ösztöndíjával ismerték el.

Kovács György a Debreceni Egyetem Informatikai Karának első éves PhD hallgatója. Egyetemi tanulmányait 2002-ben kezdte programtervező-matematikai szakon. Már hallgató korában bekapcsolódott a kar munkájába, demonstrátorként tevékenykedett, nyári szakmai ösztöndíjat több ízben nyert el, köztársasági ösztöndíjas volt. Több kutatási projektben is részt vett, majd miután megszerezte MSc. diplomáját (2007), PhD-hallgatóként folytatta tanulmányait képfeldolgozás területén.

Irodalom

- [1] A. Fazekas, Gy. Hingyi, L. Sajó, Multi-modális gépi sakkozó – Török 2, Proc. of KÉPAF'07, 25-27 January 2007, Debrecen, Hungary, pp.173–181.
- [2] A. Fazekas, A. Nagy, L. Sajó, Török 2 gépi sakkozó, Proc. of KÉPAF'07, 25-27 January 2007, Debrecen, Hungary, pp.182–189.
- [3] A. Fazekas, L. Sajó, Multi-modal human-computer chess player: The Turk 2, Proc. of ITI'07, 25-28 June 2007, Dubrovnik, Croatia, pp.29–30.
- [4] Gy. Kovács, Zs. Ruttkay, A. Fazekas, Virtual Chess Player with Emotions, Proc. of 4th Hungarian Conference on Computer Graphics and Geometry, 13-14 November 2007, Budapest, pp.182–188.
- [5] Olasz, G., Németh G., Olasz, P., Kiss, G., Gordos, G., "PROFIVOX – A Hungarian Professional TTS System for Telecommunications Applications", International Journal of Speech Technology, Vol. 3, No.3/4, December 2000, pp.201–216.
- [6] Zs. Ruttkay, H. Noot, Animated CharToon Faces. Proc. of NPAR 2000 – 1st International Symposium on Non-Photorealistic Animation and Rendering, June 2000, pp.91–100.