Joseph D. Anderson és Jessica K. Hodgins

Az emberi mozgás észlelése számítógéppel előállított képekben

A speciális hatásokért felelős animátorok célja az, hogy komputeren olyan eseményrészleteket szerkesszenek meg, amelyek az esemény valóságos környezetben felvett részeivel összevághatók vagy egybeilleszthetők úgy, hogy a számítógépes technikával készített eseményrészletek nem különböztethetők meg a valóságos részletektől. E feladat megvalósítása nem kis dolog, mégis a legtöbb filmben előfordul számítógépes animáció, amelynek jelentős részét a néző nem veszi észre. Habár a speciális hatásokért felelős művészek és technikai szakemberek kiválón végzik feladatukat, általában próba-szerencse alapon dolgoznak, s az elméleti megértés jelentős mértékben le van maradva hozzájuk képest.

Ha a cél az, hogy a számítógépes technikával készült képek illeszkedjenek a fotografált képekhez, a legkézenfekvőbb kérdés: mi a különbség a kétfajta kép között. Az egyik nyilvánvaló különbség, hogy a film képfelületén több ezüst haloid szemcse található, mint amennyi pixelt a számítógépes kép mátrixa tartalmaz, s így a fénykép felbontása és szürke skálájának terjedelme is nagyobb. De talán ennél jóval fontosabbak azok a különbségek, amelyek a kétféle képtípus készítésének módjából erednek. A számítógépes technikával készült képeket szerkesztéssel hozzák létre, a filmképet rögzítik (felveszik). Így a kérdés a következőképpen módosul: mi a különbség a szerkesztett és a rögzített képek között?

A digitalizált kép létrehozásához a tárgyakat és az eseményeket matematikailag modellálják. A létrejött kép azért lehet valószerű, mert az általunk ismert világ, úgy tűnik, fizikai törvényeket követ, és így matematikailag leírható. Elméletileg a világban létező tárgy és a vásznon megjelenő tárgy viszonya matematikailag ugyanolyan pontosan meghatározható, mintha a fénysugarak mechanikai mozgása által jött volna létre – ha eltekintünk az egyediség (specificitás) kérdésétől. Ha a világban egy esemény bekövetkezik, maga az esemény egyedi, és a változás, amit az optikai mintázatban (*array*) előidéz szintén egyedi. Mindazonáltal a mintázatban előidézett változás nem kép, nem az esemény „utánzata". Az esemény és az optikai mintázatban neki megfelelő változás között sajátosan egyedi (specifikus) viszony áll fenn. Az optikai mintázat által hordozott információ *specifikálja* az eseményt. Ahogyan James Gibson megjegyezte, „különböző fajta eseményeknek különböző fajta változások felelnek meg (Gibson 1979: 102). Az animátornak ezzel az egyediséggel kell megküzdeni.

Amikor például a felvevőgép egy felrobbanó autót, a moziban gyakori látványt filmez, a robbanás a fizika törvényei szerint következik be, a gázok szabálytalan alakban tágulnak és égnek, a kocsi millió kiszámíthatatlan alakú és nagyságú darabra hullik, s e darabok bonyolult kiszámíthatatlan pályákon repülnek szerteszét. A felvevőgép az eseménynek az adott nézőpontból minden egyes látható részletét megörökíti, s amikor a filmet később a vászonra kivetítik, a néző az egybefüggő képet és a részletek gazdagságát hihetőnek találja. Habár a számítógépes grafikus a robbanás törvényeit, például a gázok tágulását és a kocsidarabok szétrepülő mozgását általában helyesen adja meg, minden egyes apró láng- és füstmozgást, vagy a felrobbant autó minden egyes darabjának nagyságát és alakját nem tudja modellálni és kiszámítani, be kell érnie a matematikai általánosítással. S amikor a filmet később kivetítik a vászonra, a néző a szintetikusan létrehozott látványt gyakran a valóságtól eltérőnek észleli.

Ha adottak a különbségek a fotografált és a számítógépes technikával készült képek között, hogyan lehet képes az animációs művész egy adott eseményről olyan képeket alkotni, amelyek összevághatók ugyanazon esemény fotografált részleteivel, anélkül, hogy a néző észrevenné a különbséget? Ha eltekintünk a szakma olyasféle trükkjeitől, mint amikor a számítógépes technikával készült képeket árnyékolják, vagy a néző figyelmét a hangsávra vagy a kép más részeire terelik, a válasz csak az lehet, hogy a tárgyak tulajdonságait és

mozgásait az animátorok pontatlanul modellálják. A mozgóképet a mozgás különbözteti meg a fényképtől, ez teszi képessé az élő valóság visszaadására. Mert ahogyan Gibson észrevette, „a szem arra a célra fejlődött ki, hogy a változást és az átalakulást észlelje. A retinális kép ritkán egyenlő az élet kimerevített képével. Ennek megfelelően a mozgóképet az ábrázolás (*depiction*) alapvető formájának, a festményt és a fényképet pedig a speciális esetének kellene tekintenünk… A filmkészítők közelebb állnak az élethez, mint a fényképészek" (Gibson 1979: 182). A mozi készítőinek tehát csak meg kell ragadni a mozgást az életben, az animátoroknak viszont olyan algoritmusokat kell gyártani, amelyek az élő mozgást ábrázolják.

Johansson, Kozlowski, Cutting és mások korán felismerték, hogy levezethetők olyan matematikai egyenletek, amelyek leírják azoknak a fehér pontoknak fekete képernyőn történő mozgását, amelyek az [emberi test főbb mozgó pontjaira illesztett] pontszerű fényforrásokkal megjelölt mozgó vagy futó emberek lefényképezéséből vagy videoszalagra való rögzítéséből származtak (Johansson 19873, Cutting & Kozlowski 1977, Kozlowski & Cutting 1977, 1978). E rögzített adatok eredményeképpen nem volt többé akadálya annak, hogy Kozlowski, Cutting és más kutatók az emberi mozgást szimulálják. De mennyire tűnnek valószerűnek ezek a szintetikus mozgások a nézők számára? Egy érdekes problémát vettek észre akkor, amikor az animátorok az emberi mozgást vagy futást megpróbálták a képernyőn néhány másodpercen keresztül folyamatosan szimulálni. A szintetikus figurák mozgása láthatóan túl szabályos volt. A nézők azt mondták, hogy a futó figurák merevnek és természetellenesnek hatnak. Jessica Hodgins és munkatársai megvizsgálták, hogy szimulálhatók-e azok a kis szabálytalanságok, amelyek az emberi mozgással együtt járnak, és amelyeket a felvevőgép természetszerűleg megragad. A figurák mozgását leíró egyenletekbe „véletlenszerű zajt" (vagyis kis véletlenszerű változásokat) vagy „hibákat" építettek be. Az eredmény az volt, hogy a nézők a bizonyos szabálytalansággal futó figurák mozgását kicsit természetesebbnek találták, mint a szabályosakat (Bodenheimer, Shleifman & Hodgins 1999).

Egy 1998-as tanulmányban Hodgins és munkatársai egy másik gyakorlati probléma megoldását tűzték ki célul, amelynek, mint kiderült, mélyebb elméleti következményei vannak (Hodgins, O'Brien & Tumblin 1998). Amikor az animátorok egy nagyobb munkába fognak, először gyakran vázlatokat készítenek a kívánt mozgás némelyikéről úgy, hogy egyszerű formákkal dolgoznak, mint például a hagyományos pálcika figurák. Ekkor az a gyakorlati kérdés vetődik fel, hogy a kezdetben egyszerűen megrajzolt figurák mozgása ugyanolyannak tűnik-e a nézőknek, mint a teljesen kidolgozott figurák mozgása. Ahogyan Hodgins és munkatársai a kérdést megfogalmazták: „Vajon a geometriai ábra befolyásolja-e a néző ítéletét a mozgást illetően, vagy képes a néző a geometriai ábrától függetlenül pontos ítéleteket alkotni?" A kérdést 20-20 párba állított animáció segítségével kívánták ellenőrizni. A minta fele (20) pálcika figura, a másik fele (20) felöltöztetett poligonális emberi alak volt.

A nézőktől egyszerűen azt kérték, mondják meg, hogy a párokban lévő figurák mozgása egyforma vagy különböző. Valójában az egyes párokban lévő figurák mozgása különböző mértékben eltért egymástól. Általános eredményként azt kapták, hogy nézők a mozgásban meglévő kis különbségeket jobban észrevették a poligonális figurák esetében, mint a pálcika figuráknál.

A kísérlet elején a kutatók három lehetséges magyarázatot állítottak fel három lehetséges kimenetet illetően: 1) Ha a nézők finomabb különbségeket észlelnek a pálcika figurák mozgásában, annak az lehet a magyarázata, hogy „az egyszerűbb ábrákat könnyebb feldolgozni, mint az összetettebbeket, mivel a nézők jobban odafigyelhetnek a mozgás részleteire, mint a figurákéira. 2) Ha a nézők pontosabb különbségeket tesznek a poligonális figurák között, az ezért lehet, mert „az emberek jóval nagyobb tapasztalattal rendelkeznek a valódi emberi alakok helyzetét és mozgását illetően, mint amikor a pálcika figurákhoz hasonló absztrakt ábrákat kell megítélniük". 3) Ha a nézők a kétfajta figura mozgása között

egyforma mértékben tesznek különbséget, annak az lehet az oka, hogy „az emberi látórendszer az észlelt képet csak arra használja, hogy egy háromdimenziós reprezentációt működtessen. A mozgást a mentális reprezentáció, s nem a közvetlenül észlelt kép alapján ítéli meg."

Az adatok összegyűjtése és kielemzése után a kutatók arra jutottak, hogy a kísérlet sem az első, sem a második lehetőséget nem igazolja. Viszont úgy ítélték meg, hogy „empirikusan kizárja a harmadik lehetőséget, mivel megmutatta, hogy a nézők érzékenysége a mozgásváltozatokat illetően lényegesen különbözik a pálcika figurák és a poligonális ábrák esetében…". A nézők sokkal finomabb különbségeket észleltek a teljesen megrajzolt poligonális figurák mozgásait illetően, mint a pálcika figurák mozgásaiban.

Emlékezzük arra, hogy ha a nézők a kétfajta figura mozgása között egyforma mértékben tettek volna különbséget, az elméleti magyarázat kimondta volna, hogy „az emberi látórendszer az észlelt képet csak arra használja, hogy egy háromdimenziós reprezentációt működtessen. A mozgást a mentális reprezentáció, s nem a közvetlenül észlelt kép alapján ítéli meg." Mármost, az a tény, hogy a nézők a kétfajta figura mozgásváltozatai között különböző mértékben tettek különbséget, nem feltétlenül cáfolja az elméletnek azt a részét, amely feltételezi a mentális reprezentációk meglétét, jóllehet a kísérlet eredményei kétségessé tesznek egy efféle elméletet.

Vannak egyéb lehetőségek is. Mi van akkor, ha az észlelő rendszer *nem* úgy működik, hogy mentális reprezentációkat hoz létre, és ezek alapján ítéli meg a mozgást? Mi van akkor, ha az észlelő rendszer képes arra, hogy közvetlenül felismerje az emberi mozgást, akkor is, ha az minimális mértékben jelölt, ahogyan ezt Johansson, Kozlowski és Cutting javasolja? Nem lehetséges, hogy egy összetett figura esetében (mint a Hodgins, O'Brien és mások által használt poligonális figurák), amely több információt hordoz, a többletinformáció bizonyos része összeegyeztethetetlen az emberi mozgással? Nem lehet, hogy ez a többletinformáció felnagyítja a figura mozgásának változásait, s ezeket a nézők a természetes emberi mozgással összeegyeztethetetlenként észlelik? Ezeknek a kérdéseknek az ellenőrzésére szerkesztettük meg a jelen kísérletet, melyben a nézőknek háromféle futó alakot mutattunk: egy poligonális, egy pálcikákból és egy pontokból álló figurát.

Közöltük a nézőkkel, hogy mindegyik figura esetében matematikailag modelláltuk a futó mozgás, és megkértük őket, hogy rangsorolják a figurákat aszerint, hogy mennyire hasonlít a mozgásuk a természetes emberi mozgáshoz. (A korábbi – Hodgins, O'Brien és mtsi által végzett – kísérletben a nézőknek csak azt kellett megmondani, hogy a párba állított figurák mozgása hasonló vagy eltérő.) Azt feltételeztük, hogy ha a nézői percepció maga után vonja a reprezentációk megalkotását, akkor a nézők a teljesen kidolgozott poligonális figurák mozgását ítélik az emberi mozgáshoz leginkább hasonlónak, a pálcika figurákét sorolják a második, és a legkevesebb információt hordozó pontfigurák mozgását a harmadik helyre. Ha viszont a nézők közvetlenül felismerik az emberi mozgást, akkor is, ha minimálisan jelölt, és bármilyen nem emberre vonatkozó információt is ekként ismernek fel, akkor a nézők a pontokból álló figurát ítélik az emberi mozgáshoz leginkább hasonlónak, mivel az emberi futás felismeréséhez minimálisan szükséges információt, de ennél nem sokkal többet tartalmaz. A pálcika figura lenne ekkor a második, és a teljes, poligonális figura a harmadik, mivel ez utóbbi tartalmazza átlagban a legtöbb információt, de ennek legnagyobb része valószínűleg összeegyeztethetetlen az emberi mozgással.

A figurákat tartalmazó videószalagot 105 személynek mutattuk meg, akiknek nagy része műszaki és kereskedelmi szakos egyetemi hallgató volt és egy bevezető jellegű kommunikációs kurzusra jártak. A képeket kivetítőn keresztül teljes sötétben egy 15 láb [kb. 5 méter] széles vászonra vetítettük. Mind a három figura esetében ugyanazt az algoritmust használtuk azért, hogy a mozgásuk teljesen egyforma legyen. (A mozgás generálásának

részleteit l. Hodgins és mktsi 1995) Minden figurát 10 másodpercig mutattunk, 5 másodpercnyi szünetet (sötétet) tartva két figura között. A figurákat különböző sorrendben háromszor mutattuk meg.

A nézők egy válaszlapot kaptak, melyen három ikon, a pont, a pálcika és egy teljes emberi alak ikonja szerepelt, alattuk egy-egy kockával. A válaszlapon jelezni kellett a kort, a nemi és faji hovatartozást, valamint a következő útmutató volt olvasható: „Három számítógépes technikával készült futó emberi alakot fognak látni. Minden kép háromszor jelenik meg. Figyeljék a képeket, majd válaszoljanak a következő két kérdésre: Melyik számítógépes technikával készült kép közelíti meg leginkább a valóságos emberi futómozgást? Írjanak egy 1-est a megfelelő ikon alatti kockába. Melyik számítógépes technikával készült képet tennék a második helyre? Írjanak egy 2-est a megfelelő ikon alatti kockába."

Eredményként azt kaptuk, hogy habár a mozgás mindegyik figura minden egyes megjelenése során ugyanaz volt, a nézők következesen a pontokból álló figurának a mozgását ítélték úgy, hogy a leginkább megközelíti a valóságos emberi futómozgást, a pálcika figurát tették a második helyre, míg a teljes, poligonális alakot tartották a legkevésbé hasonlatosnak a valóságos emberi mozgáshoz.

Miért is tűnik úgy, hogy a pontfigura mozgása sokkal hasonlóbb a valóságos emberi mozgáshoz, mint a pálcika és a poligonális figura mozgása? Szerintünk legalább kétféle válasz lehetséges. Először is, a mozgó fehér pontok együttese igen kevés információt tartalmaz ahhoz képest, amennyi egyébként lehetne, habár, mint Cutting és munkatársai kimutatták, a pontok mozgása elegendő információt tartalmaz egy emberi alaknak, az alak nemének, sőt talán az alak identitásának a meghatározásához. Mint nézők észleljük a jelenlévő információt, és ritkán észleljük a hiányát annak, ami jelen lehetne, de nincs jelen, hacsak fel nem hívják rá a figyelmünket. Például egy vágó hamar megtanulja, hogy gyakran időben különböző, közeli vagy távoli eseményeket rakhat egymás mellé úgy, hogy ez rendszerint a közönségben a folytonosság érzetét kelti. Ha fontos, hogy a közönség tudja, hogy időbeli ugrás történt, a vágó megtalálja a módját, hogy jelezze azt.

Másodszor, míg egy mozgó pontokból álló mintázat elegendő információt hordoz az emberi mozgás felismeréséhez, kevés egyéb információt tartalmaz. Kevés információval szolgál arról, hogy a figura dinamikus szimulációs és némileg módosított algoritmusok révén jött létre. A pálcika és a poligonális figurák az előbbivel összehasonlítva jóval több információt kínálnak, de a többletinformáció nem kizárólagosan az emberi mozgásra vonatkozik, egy jelentős része olyan nem emberi alakzatokról és felületekről szól, amelyek az embertől eltérően mozognak.

A számítógépes technikával készült animációk éppúgy, mint a festett animációk, a nézőnek egy szintetikus optikai mintázatot kínálnak. A digitális művész által előállított hullámhossz- és fényerő mintázatnak olyannyira hasonlónak kell lenni a felvevőgép által rögzített mintázathoz, hogy a közönség ne vegye észre a kettő különbségét. A közvetlenül rögzített képek nem puszta szimbólumok, inkább úgy működnek, mint a valóság helyettesítői (avagy szurrogátumai). Vagyis a mozgókép által kínált vizuális mintázatokhoz úgy viszonyulunk, mint a valóságos világ vizuális mintázataihoz. Úgy észleljük azokat, mint ahogyan a természetes mintázatokat. Nem rendelkezünk másfajta képességgel, mint amivel ez utóbbiak észleljük.

Miután megtapasztaltunk egy adott eseményt, a látottakról szabadon elgondolkodhatunk. Gondolkodhatunk róla szimbolikusan, ha efféle késztetést érzünk. Azonban a mozgókép vonzereje, ami megkülönbözteti a puszta szimbolikus kommunikációtól, abban rejlik, hogy észlelő rendszerünkre közvetlenül hat, és az előttünk megjelenő változó fénymintázatot pontosan úgy észleljük, ahogyan a természetes világot. *Tudjuk* persze, hogy „ez csak mozi", de ez a tudás nem semlegesíti sem az érzékszerveinket,

sem az érzelmeinket. Minél realistább egy film, annál nagyobb hatással van az érzékszerveinkre. Másképpen szólva, a mozgóképhez úgy viszonyulunk, mint a természetes világhoz, és mindkét esetben az implicit feltételezés az, hogy a dolgok úgy vannak, ahogyan nekünk megjelennek, hacsak nem értesülünk az ellenkezőjéről. Ezen a ponton releváns az egyediség (specificitás) fogalma, mert bár az optikai mintázat egyetlen esemény képét sem tartalmazza, egyértelmű viszony van egy világbeli esemény és az optikai mintázat megváltozása (vagy átalakulása) között. A viszony valódi és matematikailag kifejezhető. A természetes világ észlelése során ezt a viszonyt rendszerint magától értetődőnek vesszük, azaz az optikai mintázat egy adott egyedi eseményt specifikál. Mivel az egyediség magától értetődő lehet, a természetes világ észlelése során sohasem kényszerülünk rákérdezni arra, hogy valóságos-e, amit látunk. És akkor sem tesszük fel ezt a kérdést, amikor filmet nézve egy olyan optikai mintázattal kerülünk szembe, amely nélkülözi az azt kiváltó fizikai eseményt.

A mozgóképek vagy a mozgóképeken látható jelenetek abban a mértékben, ahogyan pillanatról pillanatra megfelelnek a tárgyak és az események legalapvetőbb észlelésének, valóságosnak tűnnek, annak ellenére, hogy tudhatjuk, a diegetikus világ fikció vagy a fantázia terméke. Abban a mértékben, ahogyan tudatják velünk, hogy kiagyalt, megalkotott, megszerkesztett képek vagy jelenetek, aláássák a realizmusba vetett alapvető „hitünket", és arra késztetnek, hogy annak lássuk őket, *amik*, vagyis konstrukcióknak, s nem azoknak a tárgyaknak és eseményeknek, amelyeket *ábrázolnak*.

A pontfigura a mozgást a legegyszerűbb formában specifikálja. Talán, mint Yves Guiard javasolta, „az emberi mozgás lelkét ragadja meg". Amit a filmkészítők régóta tudtak, most bizonysággá vált: a realizmust illetően a több információ néha kevesebb. A pontfigura a legszélsőségesebb eset: az általa specifikált emberi mozgás benyomását a figura semmiféle részletesebb kidolgozása nem teszi valószerűbbé.

Referenciák

Bodenheimer, B., Shleyfman, A., and Hodgins, J. K., (1999) The Effects of Noise on the Perception of Animated Human Running, Computer Animation and Simulation '99, Eurographics Animation Workshop, Sept. 1999, N. Magnenat-Thalmann and D. Thalmann, eds., Springer-Verlag, Wien, pp. 53-63.

Cutting, James E. and L. T. Kozlowski. (1977) "Recognizing Friends by their Walk: Gait Perception Without Familiarity Cues." Bulletin Psychonomic Society, vol. 9, no.5, pp. 353-356.

Gibson, James J. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin, 1979.

Guiard, Yves (CNRS and University of the Medierranean, Marseilles, France). Statement made during open discussion following Open Topics Panel C, 11th International Conference on Perception and Action, University of Connecticut, Storrs, CT. June 28, 2001.

Hodgins, J. K., O'Brien, J. F., Tumblin, J., (1998) "Judgments of Human Motion with Different Geometric Models", IEEE: Transactions on Visualization and Computer Graphics, December 1998, Vol. 4, No. 4. . Earlier version appeared as Hodgins, J. K., O'Brien, J. F., Tumblin, J., 1997. Do Geometric Models Affect Judgments of Human Motion? Proceedings of Graphics Interface '97, pp 17-25.

Hodgins, J. K., Wooten, W. L., Brogan, D. C. and O'Brien, J. F. Animating Human Athletics. In *Computer Graphics* (Aug. 1995), pp. 71-78. Proceedings of SIGGRAPH 95.

Johansson, Gunnar. (1973) "Visual Perception of Biological Motion and a Model for Its Analysis." *Perception and Psychophysics*, vol. 14, no. 2, pp. 201-211.

Kozlowski, L. T. and J. E. Cutting. (1977) "Recognizing the Sex of a Walker from a Dynamic Point-Light Display." *Perception and Psychophysics*, vol. 21, no. 6, pp. 575-580.

Kozlowski, L. T. and J. E. Cutting. (1978) "Recognizing the Gender of Walkers from Point-Lights Mounted on Ankles: Some Second Thoughts," *Perception and Psychophysics*, vol. 23, no. 5, p. 459.