

Intelligens képtömörítés

Témalabor beszámoló

Antal Áron Gyula
Konzulens: dr Kovács Ilona*

2008. június 6.

Tartalomjegyzék

1. Bevezetés	1
1.1. Az emberi látás	2
1.2. A tömörítés	2
2. Egy kis történelem	3
3. Az elmélet alapjai	4
3.1. A tengely alapú reprezentáció	4
3.2. Az emberi szem kontrasztátviteli függvénye	5
4. Medial-point representation	5
5. Térfrekvenciák	6

1. Bevezetés

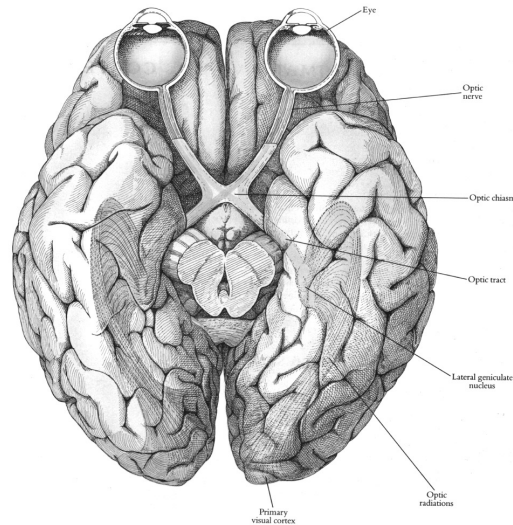
Ha bárki böngészik az interneten, vesz egy DVD-t, vagy akár a tévéhíradásokban szemléli a műholdas, vagy webes kapcsolatot, felfedezheti, sokszor, bár látjuk mi van a képen, de az szaggat, vagy a részletek annyira nem láthatók, hogy az már zavaróak lesznek.

Ez egy kompromisszum terméke. Egy bizonyos tárhely áll rendelkezésre a DVD-n, az internetkapcsolatok nem mindenhol nagy sávszélességen folynak. Ezen okból a filmet, rövid jelenetet minnél kompaktabbra, minnél kisebb helyet foglalóra tömörítik.

*BME-GTK Kognitív Tudományi Tanszék

De vajon nem lehet-e a tárhelyhatár megkötés mellett jobban észlelhető, kevésbé vagy egyáltalán nem lévő zavart elérni?

1.1. Az emberi látás



1. ábra. Az emberi látás szervei

A szembe a pupillán beérkező és a különböző optikai elemeken megtört fényt a retina, egy idegvégződésekkel borított felület alakítja idegi impulzusokká. 150 millió ilyen sejt van a pupillán. Ezt felfoghatnánk a fényképezőgép detektorának pixelszámának (egy jó fényképezőgép 8 megapixeles, ami 8 millió pixelt jelent). Ugyanakkor az agy látóközpontjába már csak ennek 150-ed részényi látóidegvégződés jut. Tehát már az agy ezen nyúlványában, a szemben is megkezdődik az a folyamat ami tömörítésnek nevezhető. Ha minden egyes jelet egyben el szeretnénk juttatni az agyba, a látóidegnek is értelmeszerűen 150-szer nagyobbak kéne lennie, aminek elhelyezése alapvető anatómia változás nélkül megoldhatatlan lenne.

1.2. A tömörítés

Mint láthattuk, az emberi szem is tömörít. Miért ne tehetnénk mi ezt is meg? Vigyázva arra, hogy a *lényeges* információk ne vesszenek el.

Fontos kérdés, hogy mik azok az előbb említett lényeges információk, amiket nem szabad elhagyni. Kutatók kimutatták, hogy az emberi látás szempontjából a színes vagy fekete-fehér kép közötti információvesztés – az előre becsült 3-mal szemben – mindössze 1,5-2-szeres, tehát a látás a színi információkra sokkal

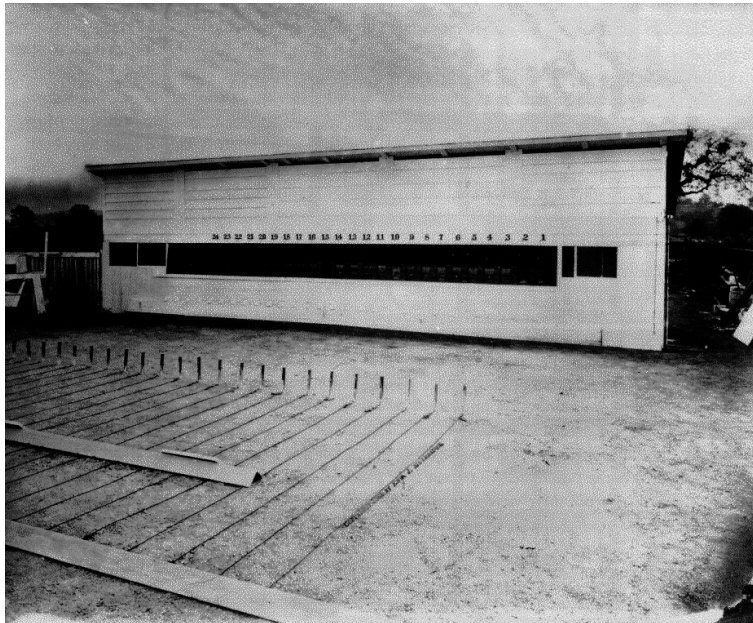
kevésbé érzékeny, mint a fényességre. (Egyes színterek erre építve, sokkal kisebb hagsúlyt fektetve a színességre érnek el kisebb méretet.)

De nemcsak ez lehet egy tömörítési alap. A JPEG például a képen cosinus-transzformációt hajt végre, és így nyert adatot szűri meg. Vagy az MPEG videótömörítés elve a kulcsképkockákkal (keyframe) való operálás. Bizonyos időközönként leírja a mozgóképet a képpel, majd a rákövetkező képkockákon csak a változásokat tárolja.

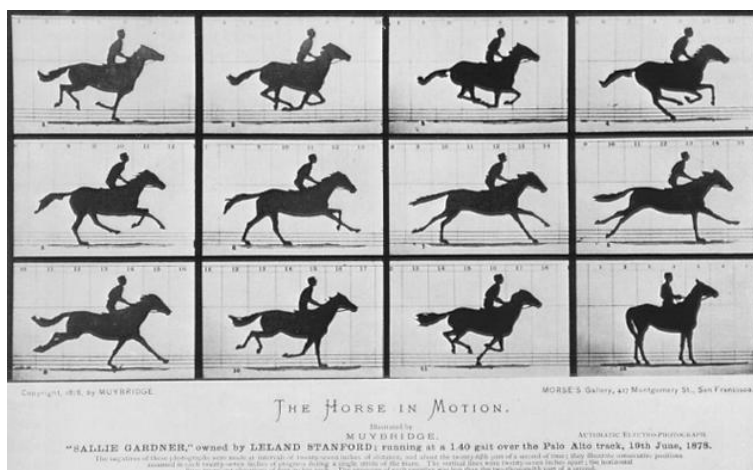
Ezen témabeszámolóban más tömörítési lehetőséget vizsgálok, mely a továbbiakban remélhetőleg hatékonyabban, és sokkal jobban a látás mecsahnizmusának ismeretével tömörít.

2. Egy kis történelem

1872-ben Kalifornia állam akkori kormányzója, Lealand Stanford fogadást kötött 25 ezer akkori dollárban, vajon a vágtazó lónak elhagyja-e mind a négy patája egy időben a földet. Megkérte Eadweard J. Muybridge-t (1830-1904), hogy döntse el ezt a kérdést. Muybridge készített egy arénát, ahol egy ló elfutott, és az elhelyezett fényképezőgépeket kioldotta a ló elhaladása, ezzel először fényképezett egy mozgó tárgyat, annak mozgását kihagsúlyozva. Bebizonyosodott, hogy valóban a levegőben van a ló egy pillanatra, de ami sokkal fontosabb a tudomány szempontjából Muybridge az életét rászánta a mozgás vizsgálatának.



2. ábra. Muybridge arénája



3. ábra. A futó ló [4]

Muybridge művei elkápráztatták korát, így Étienne-Jules Mareyt (1830-1904) is. Orvos volt, diagnosztikára akart használni a képeket. Kronofotográffal a mozgást egy fényérzékeny lemezre fényképezte, különböző szözetekkel.

Ha több esemény van egy kockán, akkor az idői részletesség magas, de a kép elmosódott, így nehezen megállítható a mozgás, ha pedig kevesebbet, ilyenkor jól elkülönülnek az alakok a képen, de a két állapot között eltelt nagyobb idő-különbség már meghatározó lesz. Egyik megoldás sem jó. A következőt csinálta. A vizsgált alak meghatározó pontjaira csillogó lemezeket helyezett, és sötét fekete háttér előtt fényképezte a mozgását. Így bár az alak formája eltűnt, de mozgására lehetett következtetni.

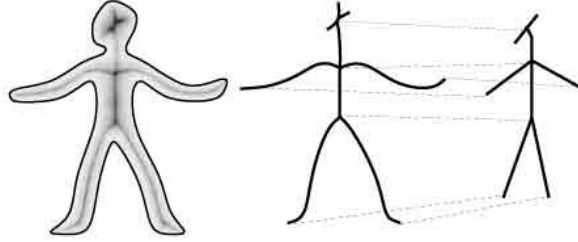
Hogy mi a jelentősége ennek? Felismerhetjük az ilyen intuitív ábrázolásokból is, hogy az emberi szem nem elsősorban az alakot magát vizsgálja, hanem annak mozgását.

3. Az elmélet alapjai

Induljunk ki az előzőekből. Keressünk egy olyan reprezentációját dolgoknak, amely az előzőekhez hasonló eredményt ad, azaz nem magát az alakot írja le, hanem annak vázát, amely igazából végbe megy a mozgásban.

3.1. A tengely alapú reprezentáció

A tengely alapú reprezentáció, vagy más néven *pálcikaemberkéék*. H. J. Blum féle futótűz algoritmussal egy vázat lehet létrehozni egy, már szegmentált képből. A név onnan ered, hogy az alakot mint égő anyagot szemlélve és a határvonalát



4. ábra. Pálcikaemberek [2]

mindenhol egyszerre meggyújtva megnézzük azokat a helyeket, ahol összejárnak a lángok. Ezeket a helyeket megrajzoljuk. Az alak egy vázát kapjuk.

3.2. Az emberi szem kontrasztátviteli függvénye

Egy rendszer kimenetén megjelenő kontraszt változását a térfrekvencia függvényében nevezzük a rendszer kontrasztátviteli függvényének.

Persze a látásra is kiterjeszthető ez. Egyes állatok más és más kontrasztátvitellel rendelkeznek. Sőt az ember élete során is változik ez a görbe. Az azonban elmondható az emberi látásról, hogy sokkal kevésbé érzékeny a nagyon nagy és nagyon kis térfrekvenciákra, kontrasztátviteli-függvényének haranggörbe szerű alakja van, melynek csúcspontja a 3-5 vonal/fok környékén van. Ez az érték az olvasótávolságban levő betű, nem véletlenül.

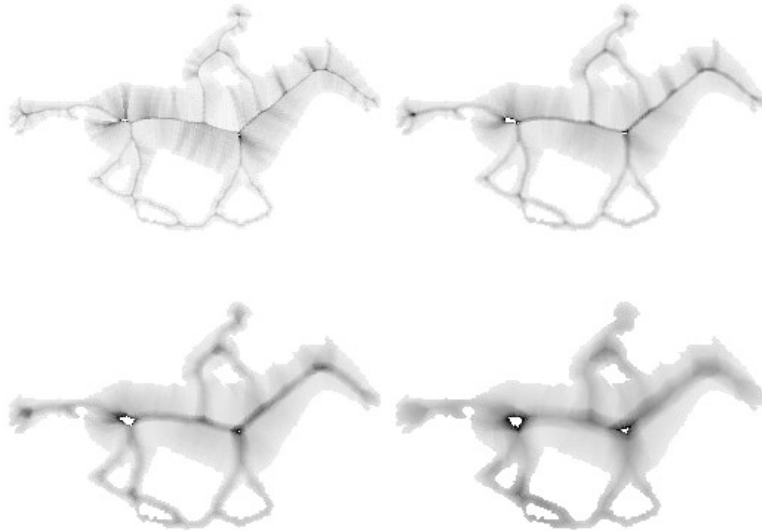
4. Medial-point representation

[1]-ben olvashatunk egy lehetséges továbbfejlesztéséről a pálcikaembereknek. Ez az algoritmus kiemeli a meghatározó pontokat, amelyeket, mint előbb Marey is kiemelt. Az algoritmus a következő:

- Első lépés a kép szegmentációja. A képből kiválasztjuk az alakzatot amit elemezni szeretnénk. Egyszerűbb képek esetén egy egyszerű élkeresés is célravezető lehet.
- Az élhalmazt, mint $B' \subset \mathbb{R}^2$ definiáljuk.
- Legyen $T := \int_{b \in B'} db$
- Minden az alakzat belső p pontjára definiáljuk a legközelebbi élpont távolságát $M(p) := \min_{b \in B'} |p - b|$
- És végül definiáljunk egy D -függvényt a következőképp:

$$D_\varepsilon(p) = \frac{1}{T} \int_{|p-b| \leq M(p)+\varepsilon} db$$

Ez a függvény egy ε paraméterű függvény. Tekintsünk egy képet különböző ε -okkal.



5. ábra. A D_ε különböző ε -ok mellett. Rendre 2, 4, 8 és 16-szorosa a körvonalnak

Az előző függvényhez hozzárendelhetünk egy c küszöbszámot. Ez a küszöbszám túlsordulást okozhat a D_ε függvényben. Ezek a túlsordult pontok meghatározzák az alakzatot.

5. Térfrekvenciák

A következő elgondolásra juttam. A bemenő képet az élmeghatározás előtt térfrekvencia-sávokra osztjuk. És az így keletkező képekre alkalmazzuk a D_ε függvényt, majd minden sávra külön választott c_{f_1, f_2} küszöbszámokkal nyerünk pontokat, és a c -t megfelelően választva még jobb reprezentációt kaphatunk.

Másik megoldás lehet más ε -t választani az egyes frekvencia-sávokra.

Adódik, hogy a különböző téri frekvencia tartományokban levő küszöbindex értéke függjön az emberi kontrasztátviteli-függvény megfelelő értékétől. Ha az ez irányú vizsgálatok bebizonyítják, hogy az jobban látjuk tőle a mozgó lovat, akkor egy lépéssel közelebb kerülünk az agy megértéséhez, és a jobb tömörítélfeljárás elkészítéséhez.



6. ábra. Egy így összállított kép

Zárszó

A fentebb bemutatott eljárás természetesen még nem megvalósított eljárás *egészesítés filmekre*, sőt rövidebb klipekre sem. Csak mankót adhat az ezirányú fejlesztések felé. Hátránya lehet, hogy a kép tömörítése esetleg hosszabb időt vehet igénybe, de a tárolása és elküldése lényesen kisebb helyen, rövidebb idő alatt történhet. Szem előtt tartva, hogy vannak alkalmazások, ahol a kommunikáció sokkal költségesebb, mint egy tömörítés, ott igenis létjogosultsága lesz, egyre gyorsabb asztali gépek esetén pedig akár mindennapjaink része is lehet ez.

Hivatkozások

- [1] *Kovács, I., Á. Fehér and B. Julesz*: Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Research (Special Issue on Recognition)* 38 2323-2333 1998
- [2] *Blum HJ*: A new model of global brain function. *Perspect Biol Medm* 10 381-407 1967
- [3] *R. Sekuler, R. Blake*: Perception, McGraw Hill, 2002
- [4] *Eadweard Muybridge*: *Horses And Other Animals In Motion*, 45 Classic Photographic Sequences, Dover Publications, New York, 1985