

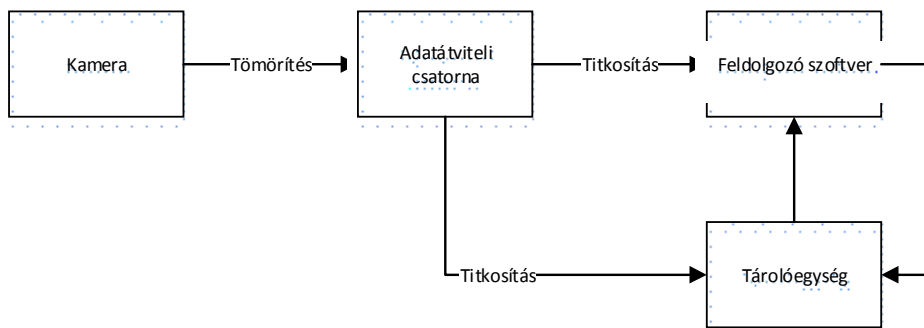
ARCFELISMERŐ RENDSZEREK GYAKORLATI PROBLÉMÁI

Bevezetés

Az arcfelismerő rendszerek – legyen szó egy beléptetéshez használt eszközzől, vagy éppen térfigyelő rendszerről – alapvetően négy részből épülnek fel. Ezek a kamera, az adatátviteli csatorna, a feldolgozó szoftver, valamint a tárolóegység. A kamera feladata a szoftver számára értelmezhető képet alkotni, majd azt a továbbításhoz betömöríteni mivel a nyers videojel a kamera felbontásától függően másodpercenként több száz megabyte méretű is lehet.

Az adatátviteli csatorna biztosít kapcsolatot a kamera, a tárolóegységek, valamint a feldolgozó szoftver között. Kapacitását úgy kell megválasztani, hogy az egész rendszertől megkívánt minőségű felvételeket az elvárt sebességgel legyen képes továbbítani. Amennyiben a csatorna kapacitása adott, úgy a többi paramétert kell oly módon megválasztani, hogy ne lépjék túl a kapacitás által biztosított lehetőségeket. Bizonyos alkalmazások (például térfigyelő rendszerek) esetében, főleg ha azok interneten keresztül továbbítanak adatokat, alapvető elvárás, hogy az adatok titkosítás után kerülhessenek csak továbbításra. A tárolóegység biztosítja a feldolgozandó, valamint a már feldolgozott anyagok tárolását, archiválását.

Végül a feldolgozó szoftver feladata az arcfelismerés és esetlegesen más egyéb intelligens funkciók (például mozgásérzékelés, forgalomszámlálás, irány felismerés, stb.) megvalósítása. A különböző szoftverek más-más algoritmusok segítségével dolgozhatnak, azonban az elvárt végeredmény azonos – felismerni és azonosítani a célszemélyt az arca alapján.



1. ábra Az arcfelismerő rendszerek elemei

Egyes rendszerekben, mint például egy biometrikus beléptető megoldás esetén, ezek az elemek fizikailag egymáshoz közel, gyakran egy tokozásban kapnak helyet. Más felépítésű rendszereknél azonban komoly fizikai távolságok is lehetnek a kamera és a feldolgozó szoftver között. Minél kisebb ez a távolság, annál nagyobb sávszélesség

biztosítható alacsony költség mellett az adatátviteli csatornának. Ígéretes fejlesztés a magyar Intellio gyártó kezdeményezése, ahol az intelligens algoritmusokat a kamerákba helyezik el, emiatt jelentős átviteli sávszélességet és szerver oldali számítási kapacitást megspórolva. Terjedelmi keretek miatt ezt a megoldást részleteiben nem tárgyalom, azonban a legtöbb megállapítás erre a technológiára is igaz.

Alapvetően a választott szoftver által használt algoritmus függvénye az, hogy milyen minőségű képre van szükség a sikeres felismeréshez, azonban mindegyikre igaz, hogy egy bizonyos minimumnak meg kell felelni annak érdekében, hogy az algoritmus egyáltalán fel tudja dolgozni a kapott információkat.

A gyakorlatban sok esetben a vezetők és döntéshozók által tévesen elképzelt valóság és az ezek alapján történt tenderkiírások nem veszik figyelembe a fizikai és technológiai korlátokat. Jelen dolgozatban bemutatásra kerül néhány olyan szempont, amelyek rávilágítanak arra, hogy miért lehetnek sikertelenek az arcfelismerő rendszerek és milyen megfontolásokat kell megtenni a kívánt eredmény eléréséhez.

Kamera [1] [2]

Napjainkban elsősorban az IP kamerák azok, amelyek a jövőbe mutató technológiát képviselik. Számos olyan paraméterrel rendelkeznek, melyek helyes megválasztása elengedhetetlen, amennyiben sikerrel kívánják ezeket alkalmazni arcfelismerés céljára. Az arcfelismerés szempontjából legfontosabb paraméterek az alábbiak.

- Felbontóképesség: a kamera vízszintes és függőleges felbontását értjük alatta. Például egy full HD felbontású kamera képe 1920 (vízszintes) x 1080 (függőleges) képpontból áll, ami nagyjából 2 MP felbontásnak felel meg.
- Érzékenység: a legnehezebben meghatározható kamera paraméter, mivel számos olyan tényezőtől függ, amelyek jelentősen befolyásolják a kép minőségét. Emiatt a gyakorlati alkalmazásoknál az az információ, hogy egy kamera 0,1 lx érzékenységű, nem elegendő a döntéshez. A gyártók által megadott érzékenységet befolyásoló tényezők, melyek a prémium gyártók kivételével legtöbb esetben nem ismeretesek:
 - Megvilágítás színhőmérséklete.
 - A tesztábra fényvisszaverő képessége.
 - A mérés során használt objektív. Sok gyártó igen jó minőségű, alacsony F-stop értékű objektívvel teszteli a kameráit. Például egy F-Stop = 0,7 a beeső fény mennyiség 40%-át, míg egy F-Stop = 1,4 értékű objektív mindössze a 10%-át engedí át.
 - Egyéb mérési körülmények, például bekapcsolt AGC (Automatic Gain Control) vagy Super dynamic képjavító megoldások.
- Egyéb felszerelések (pl. infravörös megvilágító).
- Az eszköz pozicionálása (szerelési magasság, dőlésszög).

A CCD kamerák működési elvéből következően a nagyobb pixelsűrűség kisebb pixelenkénti fényérzékenységgel jár. Ez azt jelenti, hogy alacsonyabb fényerősség esetén drasztikusan romolhat a képminőség.



2. ábra: 3 MP kamera F1,4; forrás: [2]



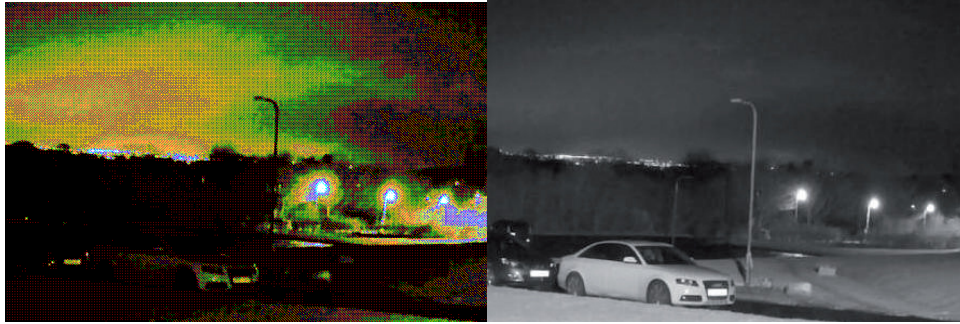
3. ábra: 1,3 MP kamera F1,4; forrás: [2]

A felhasznált objektív képes növelni a hatásosan megfigyelt távolságot, és távolabbi arcfelismerést is lehetővé tenni, azonban ez minden esetben képletvágással jár, ugyanis a nagyobb nagyítás azt eredményezi, hogy az eredetileg vizsgált területből csak egy kisebb részletet láthatunk. Emiatt egy nagyobb terület teljes megfigyelése mindenképpen több vagy jobb minőségű kamerát igényel. Figyelembe kell venni, hogy az optikák is rendelkeznek felbontással, azaz meghatározott, hogy egy objektív által alkotott képben mi az a minimális távolság két pont között, amit még különálló képpontként képes leképezni. Egy adott felbontású kamerára csak azonos, vagy nagyobb felbontási képességű objektívet szabad rakni.

Ennek számítását a $\frac{\text{CCD vízszintes elemzése}}{\text{CCD vízszintes mérete}} = \text{Minimális felbontási képesség igény}$ képlet segítségével végezhető, amit a katalógusok vonalpár/mm-ben tüntetnek fel.

Fontos körülmény, hogy a kamerák, akár csak az emberi szem, nem látnak ugyanolyan jól különböző fényviszonyok esetén. A képek feldolgozásához fény szükséges, azonban a kameráknál a CCD-k spektrális érzékenysége miatt mód nyílik a látható fény túli tartományban is feldolgozni az információkat. Szakmai tapasztalatok alapján a 780 nm körüli NIR¹³⁵ fény alkalmazása a legmegfelelőbb ilyen célokra. Ugyanakkor a megvilágítás mértéke a kamerától való távolság függvényében négyzetesen csökken, így éjszaka nem, vagy csak nehezen érhető el ugyanakkora azonosítási távolság, mint napközben – illetve annak elérése jelentős megvilágítást igényel, mely energiafelhasználás nem utolsósorban az üzemeltetésekor jelentkezik komoly többletköltséggént. Az alábbi két képen D1 felbontású (720x576) kamerakép látható 0,05 lx érzékenységu kamerával az első esetben a normál kép, a másodikban bekapcsolt infra megvilágítással.

¹³⁵ Near Infrared Light – A látható fényhez legközelebbi infravörös tartomány



4. ábra: Normál kép; forrás: [2]

5. ábra: Infravilágítás bekapcsolva; forrás: [2]

További tényezők, mint az időjárás, egy nagyobb eső, vagy a köd szintén rontják a felismerési képességet. Az eszköz telepítési magassága és vertikális dőlésszöge is befolyásolja a megfigyelhető terület méretét valamint az felismerés biztonságát. Alapvetően azonban elmondható, hogy a kamera teljes látóteréből csak egy kisebb szelvény az, ahol arcfelismeréshez megfelelő minőségű képet képes alkotni.

Adatkapcsolat, tároló kapacitás [1]

A kamera által készített képek pixelekből állnak (egészen pontosan legfeljebb a felbontás által meghatározott mennyiségű pixelből, ami azt jelenti, hogy hiába kerül telepítésre egy 5 MP felbontású kamera, ha sávszélesség korlátok miatt a továbbított képet le kell csökkenteni 1,3 MP értékre). Beszélhetünk natív felbontásról, amikor az érzékelő felület minden egyes képpontja egy képpontnak felel meg a létrehozott képen, vagy annál kisebb felbontásról, amikor több érzékelő képpont alkot egy meta képpontot. Minden pixel rendelkezik színtelítettséggel és világosságjellel. Egy teljes képkockát a felbontás által meghatározott mennyiségű pixel alkot, ahol minden pixel az előbbieken leírt tulajdonságokkal bír. Az FPS érték azt mutatja meg, hogy másodpercenként hány képkockából épül fel a kép (az emberi szem 24 FPS-től érzékeli folyamatosnak a mozgóképet, de arcfelismerési célokra 10 FPS is elég). Ezek alapján a tömörítetlen kép mérete úgy határozható meg, hogy $[\text{Pixel adatméret} \times \text{felbontás}]$, az igényelt másodpercenkénti sávszélességet pedig $[\text{Képméret} \times \text{FPS}]$.

Egy olyan eszközben, ahol a kamera és a feldolgozó egység is fizikailag egy helyen van, ez nem jelent problémát, mert az eszközök belső sávszélessége a több gigabitet is eléri másodpercenként és a gyors feldolgozás miatt ritkán szükséges az adatok folyamatos tárolása, elég őket memóriába tölteni, majd törölni onnan). Amikor azonban a felvételeket földrajzilag távoli helyen kell feldolgozni és tárolni, jelentős problémák lépnek fel, mivel egyrészt nem áll rendelkezésre megfelelő sávszélesség, másrészt az adatátviteli csatorna kiépítése és üzemeltetése komoly költséget jelentenek. E problémák megoldására jöttek létre a különböző tömörítő algoritmusok. Ezeknek két fajtája ismert, a veszteséges és a veszteségmentes.

Veszteséges tömörítési algoritmusnál az alkalmazás szempontjából szükségtelen információkat eliminál, amelyek megléte nem befolyásolja a végeredményt. Erre jó példa a színmélység, vagy a felbontás észlelési határon belüli csökkentése. Veszteségmentes tömörítés esetén az algoritmus többnyire megvizsgálja a beérkező adatokat statisztikai

szempontból, és az ismétlődő mintázatokat, vagy nagyobb, megegyező adatszoportokat kicseréli valamely kisebb helyigényű adatra, például egy olyan képnél, ahol 200 darab egymás mellett levő, azonos adattartalmú pixel található, letárolásra kerül az egy pixel információja és a darabszáma. A veszteségmentes tömörítés típus rendelkezik egy olyan elméleti határral, amely alá nem csökkenhet egy adott fájl mérete, bizonyos helyzetekben még a tömörítés nélküli adatfolyam méretét is meghaladhatja.

Annak függvényében, hogy a feldolgozó szoftver milyen képminőséget és képi információt igényel, meghatározható, hogy mekkora felbontású kamerát és milyen tömörítési eljárást kell alkalmazni, figyelembe véve természetesen a rendelkezésre álló sávszélességet és tárolókapacitást is.

A tárolókapacitás meghatározásánál is kalkulálni kell a sávszélesség szükséglettel, mivel a tárolóeszköznek képesnek kell lennie fogadni és rögzíteni a beérkező adatokat. Egyes rendszerekben a további terhelést jelent, hogy egy időben legyen képes kiszolgálni a lokális és távoli megjelenítés, visszajátszás, adminisztrálás és archiválás feladatait. Az ilyen rögzítő megoldások pentaplex néven ismeretesek.

Feldolgozó szoftver

A feldolgozó szoftver alatt értünk minden olyan intelligens funkciót, amely a pentaplex működésen túlmutató szolgáltatásokat valósítanak meg. Ilyen funkciók például az arcfelismerés, szituáció elemzés, irányfigyelés, tárgyfigyelés, ember és forgalomszámlálás, stb.

Az arcfelismerő algoritmusok az arc különböző paramétereit vizsgálják, jellemzően bizonyos pontokat keresnek meg, és az alapján építik fel az arc geometriáját, amelyet egy kódban tárolnak, majd ezzel a kóddal vetik össze a látott képeket. A geometriai elemzéshez szükséges egy minimális méretű és felbontású kép. Jelenleg a szakmában elfogadott elvárás az, hogy egy szemből érkező személy esetén a két szem között a képen legalább 60 pixel (ennél léteznek már jobb aránnyal rendelkező algoritmusok is) látható legyen. Ez biztosítja, hogy megfelelő méretű és minőségű kép álljon rendelkezésre a szoftvernek az azonosítás elvégzéséhez.

Különbséget kell tenni az együttműködő és az együtt nem működő célszemélyek közt. Az előbbi esetre jó példa egy biometrikus beléptető rendszer, ahol a felhasználó célja a belépés és ennek érdekében együttműködik a rendszerrel, azaz jól pozícionálja az arcát.

Az együtt nem működő célszemély lehet például egy térfigyelő rendszer által megfigyelt ember, aki egyébként a megfigyelés tényéről nem tud, emiatt természetesen viselkedik, nem néz a kamerába, nem pozícionálja az arcát. Egy másik példa, amikor a rosszindulatú elkövető szándékosan takarja el az arcát vagy fordul el a kameráktól. Például a bostoni merényletnél a körözött elkövetők utólag felismerhetők voltak a térfigyelő rendszer felvételein. Ez utóbbi esetekben a szoftvernek jóval több képkocka közül kell kiválasztania egy olyat, amely alkalmas az azonosítás elvégzésére, vagy képesnek kell lennie más szögekből is felismerni a személyt.

Alkalmazási példa

Az egy kamera által 10 FPS mellett igényelt sávszélesség, és 24 óra felvételhez elegendő tárterület az alábbiak szerint alakul H.264-10 (10 bites mintavételezés) kódolás esetén:

Felbontás	1080p	5MP
Sávszélesség (Mbit/s)	2,46	5,98
Tárhely (GB)	26,5	64,6

6. ábra H.264-10 kódolás által igényelt sávszélesség és 24 órás tárhely; forrás: IP Video System Design Tool 7

Látható, hogy már egy kamera is jelentős sávszélességet és tárhelyet igényel.

Az alábbi példákban a Széll Kálmán tér méretarányos képére elhelyezett, 4 méter szerelési magasságú kamerák által belátott terület kerül bemutatásra. A kamerák párhuzamosan helyezkednek el. A szimuláció az IP Video System Design Tool 7 programmal készült. A szimulációt a kamerától 60 méterre elhelyezkedő 2 méter magas referenciaszemélyre végeztük el. A kamerák H.264-10, magas minőségű tömörítést használnak. Az objektívek úgy lettek megválasztva, hogy a meghatározott távolságokban (referenciátávok) a kamera látótere 10 méter széles legyen. Fontos megjegyezni, hogy ennél szélesebb látószög kisebb képkivágású, következésképpen kisebb fókusztávolságú (és nagyítású) objektívet igényel. Ebben az esetben természetesen az a terület, ahol az eszköz képes olyan képet alkotni, ahol megvan a 60 pixel a két szem között, jelentős mértékben csökken. A szimuláció időjárási- és fényviszonyok tekintetében ideálisnak tekinthető környezetben készült. Az alábbi táblázatok szemléltetik a felhasznált kamerák paramétereit:

Kamera megnevezése (MP)	Felbontás	Formátum, képarány
1,3 MP	1365×1024	1/3" 4:3
3 MP	2048×1536	1/3" 4:3
5 MP	2600×1950	1/3" 4:3

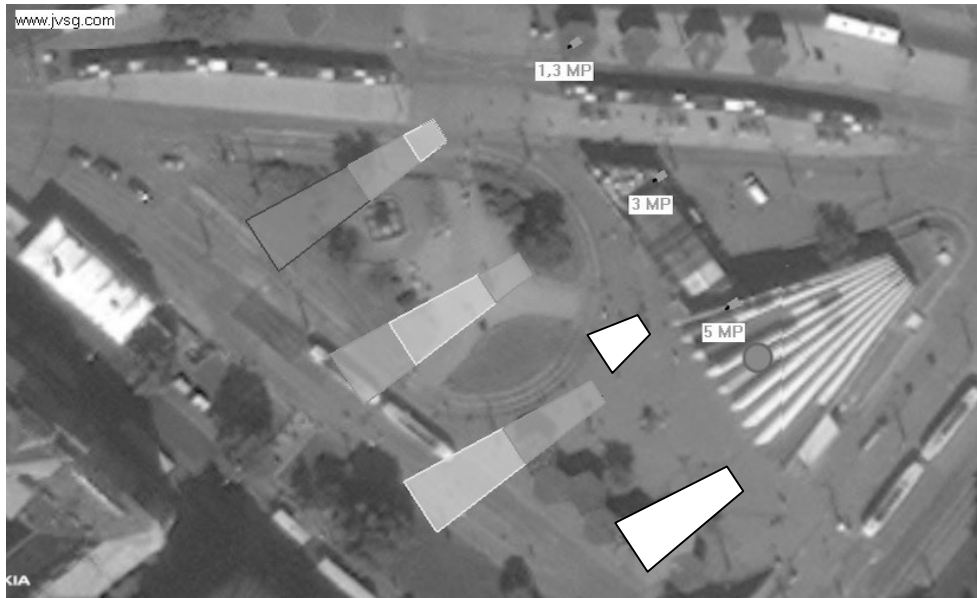
1. táblázat A szimulált kamerák paramétereit

Referenciátáv (m)	Fókusztaáv (mm)	Látószög (fok)
60	28,9	9,5

2. táblázat A szimulált távolságokhoz rendelt fókusztáv és látószög

Az arcfelismerésre alkalmas tartományt fehér színnel tüntettük fel.

Fontos megjegyezni, hogy amennyiben az algoritmusnak ennél nagyobb pixelszámra van szüksége, úgy az arcfelismerési sávok ennek megfelelően csökkenni fognak.



7. ábra Arcfelismerési zónák 60 méter referenciatáv esetén

Az 5-4. ábrán jól látszik, hogy az 1,3 MP felbontású kamera használatánál nincs olyan tartomány, ahol az előírt 60 pixel eredmény realizálható. 3 MP felbontásnál már elérhető a kívánt cél, 5 MP-es kameráknál ez a terület mintegy duplájára nő. Ezekből az eredményekből kiszámítható, hogy a Széll Kálmán tér, képen látható részletét mintegy 600 darab 5 MP felbontású kamerával lehetne teljesen lefedni arcfelismerés minőségű felbontással.

A kamerák képalkotási képességét a megvilágítás jelentősen befolyásolja, önmagában teljes sötétségben, vagy rendkívül gyenge környezeti fényben az sem lát. Ennek megfelelően a minden segédeszköz nélkül alkalmazott kamera nem tud a nap 24 órájában a szoftver számára felhasználható minőségű képet előállítani. Általában véve elmondhatjuk, hogy ha az ideális körülményekhez képest vizsgáljuk a képminőséget, az éjszaka legsötétebb részén valamilyen minimális szintről indul a görbe, majd ahogy világosodik, a görbe felfut. A legfényesebb órákban, erős napsütés mellett, szintén kialakulhat egy minőségcsökkenés, tekintetbe véve, hogy a túl erős fény elvakíthatja a kamerát, amennyiben a dinamikataromány nem elég széles. Ezek után a délután és alkony folyamán a hajnalhoz hasonló, közel inverz folyamat zajlik le, majd az érzékelési képesség visszasüllyed az éjszakai környezeti megvilágítás által biztosított szintre.

Az éjszakai megvilágítást egy rendszer esetében sokféleképp lehet megoldani. Amennyiben a közvilágítás megfelelő erősségű fényt vet a vizsgált területre, teendő nincs. A valószínűbb helyzet azonban az, hogy a kamerának valamilyen saját megvilágításra van szüksége. Itt figyelembe kell venni, hogy a megvilágítás mértéke a távolság arányában négyzetesen csökken, ezért sok esetben a megvilágítást nem a kameránál kell elhelyezni, hanem a referenciatávolság közelében.

A másik fontos paraméter a megvilágító fény hullámhossza, több szempontból is. Egyrészt a CCD kamerák fényérzékenysége nem egyezik meg az emberi szemével – bizonyos hullámhosszokon, beleértve a NIR sávot, jelentősen érzékenyebb, másrészt a különböző hullámhosszú fényvel történő megvilágítás különböző minőségű képet eredményez az emberi test fényvisszaverő képessége miatt, melyet az alábbi ábra szemléltet:



8. ábra Normál és NIR megvilágítás közti különbségek¹³⁶

A képen a felső sorban különböző szögből, színösszetétellel és erősséggel érkező megvilágítás látható, az alsó sorban pedig ugyanazon megvilágítások mellett NIR fényvel történő megvilágítás hatása. Jól látható, hogy a NIR megvilágítás mellett bármilyen más megvilágítás szinte érzékelhetetlen, a képminőség pedig egyenletes.

Összefoglalás

Egy kamera által belátott terület csak igen kis része alkalmas automatikus arcfelismerésre, amit jelentősen befolyásol a napszak és az időjárás is. Ebből fakadóan nem javasolt konkrét kamera paraméterek kiírása tenderekben úgy, mint felbontás, érzékenység, látószög, mivel azok önmagukban nem nyújtanak információt arról, hogy a megfogalmazott arcfelismerési biztonság teljesül-e vagy sem. Ezzel szemben célszerűen olyan, a konkrét területre meghatározott magas szintű igények leírása javasolható, amelyek biztosítják a megfogalmazott eredmények elérését. A példában szereplő Széll Kálmán tér kapcsán érdemes lehet megadni, hogy nappali megvilágításnál a tér körvonalától számított 2 méteres sávban legyenek az algoritmus által felismerhetők az arcok az esetek 75%-ában, ezen belül és kívül 10-10 méteres sávban elegendő az alak és mozgásfelismerés.

Az esetek döntő többségében sincs szükség mindenhol arcfelismerésre. Jól megválasztott tervezéssel azonosíthatók azok a véges számú kulcs területek, ahol ténylegesen szükséges az arcfelismerés. Ilyenkor elegendő lehet, ha arcfelismerés a megközelítési útvonalakon van, magát a területet elég olyan – egyébként szélesebb látószöggel bíró, ugyanakkor jó minőségű képet adó – eszközökkel megfigyelni, amelyek lehetővé tesznek valamilyen más tulajdonság (például ruházat vagy mozgás) alapján történő

¹³⁶ Forrás: http://what-when-how.com/wp-content/uploads/2012/06/tmp75272_thumb.png; (Letöltés ideje: 2014.05.12.)

azonosítást. Ilyen esetekben, például bűncselekmény elkövetésekor az elkövetőt ruházata alapján vissza lehet keresni az érkező személyek között, mivel az elkövető mozgását visszakövetve megtalálható az arcfelismerésre alkalmas pont, ahol már azonosítható az elkövető.

Irodalomjegyzék

- [1] T. Levente, CCTV Magyarul, BM Nyomda, 2004.
- [2] AXIS Communications, „A summary report from Axis and Raytec regional test nights,” 2012.