Vídeo-Hermenèutica o la interpretació del comportament humà en seqüències d'imatges

01/2010 - **Telecomunicacions, Electrònica i Informàtica.**

L'hermenèutica, definida com l'art d'interpretar un missatge, es va centrar durant molts segles en teoritzar el procés d'interpretació de textos escrits, sobretot bíblics. L'aparició a finals del s. XIX dels primers enregistraments de seqüències d'imatges va ampliar el domini de recerca, interessant durant el s. XX a filòsofs com Heidegger: en aquest nou format de comunicació humana (visual), el text cinematogràfic es convertia en un joc interpretatiu on el llenguatge visual s'articulava en una xarxa de múltiples lectures. L'hermenèutica de seqüències d'imatges, o Vídeo-Hermenèutica, implica explicar el valor subjectiu i social del comportament humà observat en seqüències d'imatges i, en general, de tot contingut multimèdia. La vídeo-hermenèutica no només s'interessa pel que passa en un vídeo, sinó per entendre quin significat té el que està essent descrit, quin missatge ens transmet com a observadors.

L'avatar parlant identifica la situació reproduïda per streaming i la descriu amb paraules.

Aquest anàlisis del comportament humà en seqüències d'imatges també s'ha modelitzat en termes computacionals dins del camp de les Ciències de la Computació. I això ha sigut possible gràcies als avenços tècnics i de hardware: en particular, a l'abaratiment dels costs de les càmeres, que va comportar una expansió del seu ús per a la vídeo-vigilància. I aquest ús ha provocat la necessitat d'analitzar automàticament i en temps real el comportament humà observat en milions de càmeres.

Aquests factors han generat importants aportacions científico-tècniques també en les àrees de la Visió per Computador i de la Intel·ligència Artificial: un recull dels treballs més recents en aquest àmbit es troben en un número especial del International Journal of Pattern Recognition and Artificial Intelligence, el qual és introduït en l'article descrit a les referències. A grans trets, es poden categoritzar 4 graus de complexitat en els sistemes de visió artificial, com les fites que s'han assolit en les dues darreres dècades.

El primer pas, bàsic i crític a la vegada, correspon a la detecció i seguiment del (i) moviment. Sense detecció no hi ha interpretació, i sorgeixen molts problemes en aquest sentit: saturacions, ombres, camuflatge, moviment del fons, ocultacions... La següent tasca correspon al reconeixement de les (ii) accions realitzades pels agents detectats, per exemple caminar, ajupir-se o córrer. El tercer grau implica modelitzar les (iii) activitats, definides com accions més interaccions i reaccions. Les activitats determinen si, per exemple, dos agents s'apropen, giren, es persegueixen.... Per últim, la categoria amb més càrrega contextual correspon als (iv) comportaments, que situen les activitats en una escena concreta; per exemple, si dos agents s'apropen ràpidament entre ells (activitat), l'escena és un carrer i un agent és detectat com humà i l'altre com a vehicle, s'interpretarà un possible perill d'atropellament.

Aquesta gradació ens indica com es va aprofundint en la semàntica de cada imatge per reduir la incertesa i l'error del procés interpretatiu i millorar així la utilitat semàntica de l'explicació del comportament observat. És així com es va donant cada cop més importància a la identificació, per una banda, de l'escena on es produeix el moviment i de les seves regions semànticament més rellevants, com ara una vorera, una parada d'autobús o un pas de vianants, i per altra banda d'aquells objectes amb els que els agents detectats poden interactuar, com ara bosses, bicicletes, cadires, portes, finestres,... Per últim, s'estan dissenyant estratègies que incorporen altres elements semàntics que poden aparèixer en el vídeo i que són addicionals a la informació que hi ha a la imatge, com ara l'àudio o l'aparició de paraules escrites.

Les tendències més prometedores en aquest sentit indiquen que el futur de la vídeo-hermèutica es troba, entre d'altres, a Internet. Per una banda, degut a l'enorme varietat de comportaments humans i d'escenes, moltíssim més alta que en la vídeo-vigilància, una estratègia emergent consisteix en utilitzar les imatges i vídeos que els usuaris pugen a la xarxa per modelitzar aquesta varietat. Per altra banda, s'estan dissenyant sistemes que milloren l'anàlisi de no només el contingut multimèdia generat per empreses audiovisuals i altres indústries culturals, sinó del contingut streaming generat per plataformes com YouTube.

Ha passat més d'un segle des de què pioners com Muybridge i Marey van fer els primers enregistraments de moviments humans amb finalitats d'anàlisi; en aquella època una de les aplicacions més demandades era poder determinar la distribució més òptima del material bèl·lic que portava un soldat, per reduir-li la fatiga durant jornades molt llargues. En els nostres dies, l'aplicació més prometedora a llarg terme és el disseny de programes d'anotació automàtica de vídeo per aplicar indexacions més informatives i eficients d'arxius multimèdia, per refinar els resultats dels motors de cerca i per, en definitiva, trobar ràpidament el contingut semàntic desitjat en un volum de dades cada cop més infinit.

Jordi González

Centre de Visió per Computador

"Video Analysis and Understanding for Surveillance Appilcations". Wang, Liang; Wu, Qiang; Li, Ming; Gonzalez, Jordi; Geng, Xin. International Journal of Pattern Recognition and Artificial Intelligence, 23 (7): 1221-1222 NOV 2009.