

PENJEJAKAN FITUR BERBASIS KOHERENSI TEMPORAL DALAM SISTEM ANIMASI EKSPRESI WAJAH

Samuel Gandang Gunanto, Mochamad Hariadi, Eko Mulyanto Yuniarno

Program Studi Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya

Jalan Arief Rahman Hakim, Kampus Keputih, Sukolilo, Surabaya, 60111

No. Tlp: +62-31-592 7939, 592 3411, No. Hp.: 08562543431

E-mail: gandang6@gmail.com, mochar@ee.its.ac.id, ekomulyanto@ee.its.ac.id

Abstrak

Tingginya permintaan produktivitas industri animasi di Indonesia menuntut adanya perubahan di sektor produksi. Teknologi *motion capture* merupakan penerapan prinsip visi komputer yang mengadaptasi indera mata manusia untuk mengenali fenomena gerakan yang tertangkap kamera dan memetakannya dalam pola gerak virtual. Tulisan ilmiah ini akan membahas metode penjejakan fitur penanda di wajah manusia untuk mendapatkan informasi mengenai ekspresi wajah. Teknik penjejakan menggunakan penerapan prinsip koherensi temporal. Asumsi yang digunakan pada penelitian ini berargumentasi bahwa dengan menggunakan pendekatan koherensi temporal, maka proses penjejakan fitur di citra sekuensial dapat disederhanakan dengan perhitungan nilai kedekatan pada penanda di setiap *frame*-nya. Hasil yang didapat menunjukkan bahwa proses penjejakan fitur yang diusulkan memiliki hasil yang handal untuk menangani banyak *frame*. Komputasi yang digunakan juga sangat efisien dan hemat karena prosesnya tidak memerlukan tahap pembelajaran terlebih dahulu. Kumpulan hasil penjejakan parameter fitur penanda secara sekuensial akan membentuk sebuah basis data ekspresi visual dari wajah manusia.

Kata kunci: penjejakan berbasis penanda, koherensi temporal, animasi ekspresi wajah

Abstract

Temporal Coherence Based Feature Tracking in the Animation System of Facial Expression. High demand on the productivity of the animation industry in Indonesia requires a change in the existing production process. Motion capture technology is the implementation of a computer vision principle to adopt the human eye senses to understand the phenomenon of motion results from a camera and to map the virtual movement patterns. This paper will discuss a method for tracking marker features in the human face to obtain information about facial expressions. The tracking technique is using implementation of temporal coherence principle. This research assumes that by using temporal coherence approach, the tracking process in sequential images can be simplified by calculating similarity on markers in each frame. The result shows that this feature-tracking process have reliable result to handle a lot of frames. The computation used is very efficient and cheap because it does not require a learning process in advance. The precision accuracy of tracking parameters generated a database of good visual expression.

Keywords: marker-based tracking, temporal coherence, facial animation

PENDAHULUAN

Tingginya permintaan produktivitas industri animasi di Indonesia menuntut adanya perubahan dalam proses produksi yang ada.

Kurangnya sumber daya manusia yang ahli di bidang animasi dan mahal teknologi pendukung mendorong munculnya penelitian eksperimental yang mampu menghadirkan

teknologi alternatif serupa yang mampu diterapkan di sektor industri animasi Indonesia.

Teknologi *motion capture* merupakan penerapan prinsip visi komputer yang mengadaptasi indera pengelihatan manusia untuk memahami fenomena visual hasil tangkapan kamera dan memetakannya dalam lingkungan virtual. Tangkapan pergerakan manusia dapat dipetakan dalam model skeletal. Hal ini juga dapat diterapkan pada perubahan gerak fitur wajah yang telah ditandai. Tata letak penanda ditempatkan pada persendian dan otot gerak wajah yang memiliki pengaruh signifikan pada perubahan ekspresi wajah manusia. Hal ini digunakan agar dapat dilakukan penangkapan fitur penanda ekspresi wajah secara optimal sehingga proses pemetaannya secara virtual mampu menggambarkan ekspresi tangkapan yang serupa.

Tulisan ilmiah ini akan membahas proses penjejakan fitur penanda yang ditemukan di wajah manusia untuk mendapatkan informasi ekspresi wajah manusia yang akan dipetakan di model wajah virtual tiga dimensi. Hal ini dipandang penting dikarenakan masih banyak industri animasi di Indonesia yang melakukan pemetaan gerakan ekspresi wajah karakter virtual tiga dimensi secara manual. Proses manual membutuhkan waktu yang lama dalam pengerjaannya sehingga jika diinginkan produktivitas meningkat, maka dibutuhkan sumber daya manusia yang banyak juga. Hal ini akan berimbas pada mahalnya biaya produksi sebuah film animasi Indonesia dibandingkan hasil yang serupa diproduksi oleh industri yang telah menerapkan teknologi pendukung animasi, seperti teknologi *motion capture*. Produksi animasi ekspresi wajah memiliki tingkat kesulitan tersendiri dikarenakan bentuk wajah tangkapan yang berlainan dan karakteristik unik dari karakter virtual tiga dimensi yang

digunakan, seperti halnya karakter kartun bukan manusia.

Penelitian ini menggunakan citra tangkapan wajah dua dimensi sebagai objek utama. Wajah manusia yang ditangkap kamera terlebih dahulu telah diberi penanda fitur menggunakan material reflektif yang peka cahaya. Asumsi penelitian dibuat dengan argumentasi bahwa dengan pendekatan koherensi temporal, penjejakan fitur penanda di citra sekuensial dapat disederhanakan dan menghemat proses komputasi. Tujuan penelitian ini adalah untuk melakukan penjejakan fitur penanda di citra tangkapan kamera secara benar dengan mensintesis prinsip koherensi temporal. Pada akhirnya, hasil penelitian ini juga dapat berkontribusi pada penemuan teknik baru proses penjejakan fitur penanda di citra sekuensial atau video.

ANIMASI EKSPRESI WAJAH BERBASIS KOMPUTER

Animasi ekspresi wajah merupakan aspek penting yang menunjukkan eksistensi kehadiran sebuah karakter di lingkungan virtual tiga dimensi. Penggunaan animasi ekspresi wajah dapat dijumpai di industri *game* tiga dimensi, industri film animasi dan industri pengembangan perangkat lunak interaktif. Meskipun untuk memproduksi animasi ekspresi wajah yang bagus membutuhkan waktu yang lama bagi seorang animator profesional dikarenakan kompleksitas ekspresi wajah, namun hal ini tetap harus dikerjakan dengan baik supaya karakter virtual mampu menampilkan ekspresi yang natural (Parke, 1972:451). Sampai saat ini, pengembangan penelitian animasi ekspresi wajah masih terus dilakukan terutama di dua aspek penting, yaitu yang pertama adalah proses penulangan atau

rigging sebuah ekspresi wajah dan pergerakan antara dua buah karakter virtual atau antara manusia ke karakter virtual tiga dimensi.

Proses *rigging* membutuhkan waktu yang lama jika dikerjakan secara manual oleh seorang animator. Bahkan untuk metode sederhana seperti *shape-blending*, sang animator perlu membuat sebuah bentuk referensi sebagai permulaannya (Dutreve, Meyer & Bouakaz, 2008:1). Pendekatan teknik pengulitan atau *skinning* secara umum membutuhkan definisi manual persendian yang akan terpengaruh oleh pergerakan pada wajah, seperti alis, dahi, pipi, rahang, dan mulut. Jika menggunakan pendekatan simulasi kulit dan otot wajah, animator tetap melakukan intervensi pada penetapan kulit wajah yang berhubungan langsung dengan tulang atau persendian pada tengkorak kepala serta pengaturan pergerakan otot yang memengaruhinya. Hal inilah yang menyebabkan produksi animasi ekspresi wajah tidak dapat digunakan berulang secara langsung ke model wajah lainnya dikarenakan kompleksitas bagian dan keunikan bentuk penulangan yang terjadi antarkarakter.

Aspek yang kedua dari penelitian seputar animasi ekspresi wajah adalah adanya ide yang menyatakan bahwa animasi ekspresi wajah secara natural dapat diproduksi dengan cara transfer pergerakan dari satu wajah model atau wajah manusia hasil tangkapan kamera ke wajah karakter baru lainnya. Hal ini dapat dilakukan dengan menggunakan teknologi *motion capture* yang berupa tangkapan kamera dua dimensi dari gerakan wajah seorang aktor yang dipetakan secara langsung ke wajah karakter virtual. Metode ini memungkinkan pengelolaan memori dan penggunaan sumber daya animasi lebih efisien di sektor industri dikarenakan penyimpanan data hanya dilakukan sekali saja untuk data

pergerakan yang serupa. Selain keuntungan dari teknik ini, kesulitan yang dijumpai juga sangat besar terutama saat penyesuaian bentuk model target yang harus sesuai dengan sang aktor supaya hasil tangkapan gerak yang didapatkan dapat optimal. Perihal karakteristik tersebut, maka dibutuhkan sebuah teknik transfer yang mampu beradaptasi terhadap perubahan ekspresi wajah aktor dan perubahan morfologi model target secara simultan.

Metode mimik yang dikembangkan oleh Hjortsjo pada 1970 merupakan usaha awal pemetaan ekspresi wajah manusia yang didasarkan pada sistem gerak otot wajah. Data yang diteliti berasal dari ekspresi wajah sang aktor manusia dan dipetakan secara visual (Hjortsjo, 1970:1). Pada tahun 1978, Paul Ekman dan Wallace Friesen mengembangkan metode serupa dengan istilah bahasa ekspresi yang sering dikenal dengan sebutan *Facial Action Coding System (FACS)*. *FACS* membagi area wajah menjadi 66 unit aksi atau *Action Unit (AU)* yang mampu merekonstruksi ekspresi wajah secara alami (Ekman & Friesen, 1978:1). Sistem *FACS* sampai saat ini masih merupakan standar yang digunakan oleh animator untuk membuat ekspresi wajah dengan memanfaatkan parameter pengontrol di masing-masing *AU* secara langsung.

Teknik pemetaan gerak untuk membangkitkan animasi ekspresi wajah dapat dilakukan dengan menggunakan sintesis penanda ataupun tanpa adanya penanda pada wajah acuan. Sistem yang dibangun dengan sintesis penanda banyak dikembangkan secara komersial dan menjadi standar teknologi *motion capture* di industri animasi. Keunggulan dari sistem yang berbasis penanda terletak pada kemudahan perhitungan komputasi sehingga tidak memerlukan proses yang panjang dan

mendukung penggunaan proses secara seketika dalam proses produksi animasi.

SISTEM PENJEJAKAN FITUR BERBASIS PENANDA

Sistem *motion capture* berbasis penanda dibangun dengan menjaga kestabilan fitur tangkapan kamera secara temporal. Sobottka dan Pitas menggunakan metode *template matching* untuk melakukan penjeakan fitur penanda sehingga setiap fitur yang terdeteksi pada citra sekuensial akan selalu sesuai dengan basis data yang telah dibangun di awal (Sobottka & Pitas, 1997:1). Meskipun semua fitur dapat terdeteksi dengan baik, proses ini membutuhkan perhitungan komputasi yang sangat banyak. Kecepatan proses sangat ditentukan dari besarnya kapasitas memori dan kompleksitas citra tangkapan kamera. Ko, Kim, dan Ramakrishna mencoba mengembangkan sistem penjeakan fitur yang lebih ringan komputasinya dengan cara mengurangi fitur wajah yang dideteksi, seperti mata, lubang hidung, dan bibir sebagai perwakilan fitur wajah. Proses penjeakan fitur yang terdapat di citra sekuensial dilakukan dengan metode *graph matching*. Teknik pendekatan yang digunakan adalah dengan melakukan perubahan perhitungan yang semula menggunakan nilai keabu-abuan menjadi nilai biner yang berupa nilai intensitas sebuah piksel. Metode ini terbukti mampu menyederhanakan perhitungan komputasi yang dilakukan saat sistem sedang melakukan penjeakan fitur sehingga proses yang dilakukan dapat dipersingkat (Ko, Kim, & Ramakrishna, 1999:1).

Bolkart dan Wuhrer mencoba menggunakan metode *spin image* yang dikembangkan oleh Johnson (Johnson & Hebert, 1997:684) untuk melakukan penjeakan

fitur penanda di citra sekuensial yang pertama dan menggunakannya sebagai acuan dasar penjeakan di rangkaian citra selanjutnya (Bolkart & Wuhrer, 2015:100). Sebuah *spin image* mampu menjelaskan keberadaan posisi piksel dan keterkaitannya terhadap faktor tetangga yang berdekatan. Jika didapati *spin image* yang berlainan, dapat dilakukan perbandingan, penggabungan, dan pada akhirnya dilakukan proses korespondensi antara dua buah *mesh* tiga dimensi. Tes kehandalan sistem dilakukan dengan metode RANSAC (*Random Sample Consensus*) untuk menjamin hasil yang didapatkan baik. Namun, penelitian ini memiliki kelemahan dari sisi komputasi yang rumit dan kompleksitas yang tinggi sehingga tidak memungkinkan diterapkan secara langsung di sektor produksi animasi.

Pengembangan sistem animasi ekspresi wajah dapat dilihat dari dua buah kegiatan yang saling bebas, yaitu proses pengembangan parameterisasi yang terkontrol seperti layaknya antarmuka pengguna dan pengembangan teknik animasi wajah berbasis pada parameter (Parke & Waters, 2008:1). Teknik yang memiliki peran penting DALAM proses animasi wajah adalah parameterisasi yang terkontrol karena teknik ini akan menjadi referensi dalam pembangkitan animasi wajah secara keseluruhan. Semakin baik proses pengontrolan yang dilakukan di setiap parameter, maka proses pembangkitan animasi ekspresi wajah akan semakin alami dan akurat.

Koherensi temporal di proses penjeakan fitur terparameter sangat penting karena faktor ini akan menjadi penentu kehandalan sebuah sistem. Pendekatan model temporal akan mampu mengurangi perhitungan komputasi dan meningkatkan akurasi dalam penjeakan fitur. Setiap penanda akan lebih mudah dikenali secara cepat di citra sekuensial

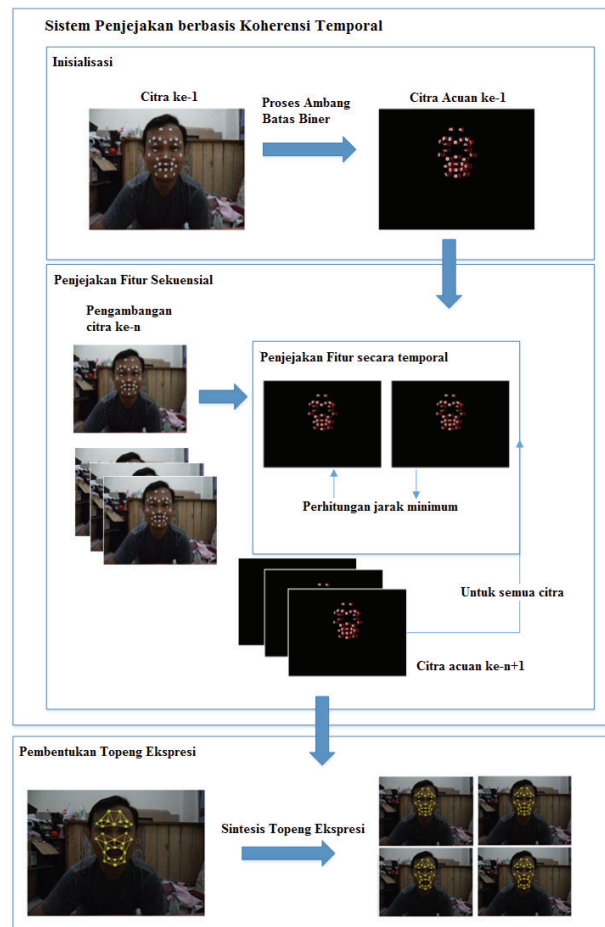
dengan mengandalkan perhitungan jarak minimum terhadap keberadaan fitur serupa di citra sebelumnya. Kecepatan proses yang tinggi ini memungkinkan pendekatan yang dilakukan diterapkan pada sistem produksi secara langsung dan seketika.

PEMBAHASAN

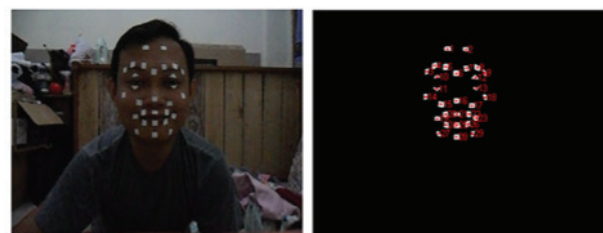
Penelitian ini dilakukan dengan terlebih dahulu melakukan studi literatur dan eksperimentasi tahapan proses secara parsial. Secara keseluruhan tahapan eksperimentasi yang dilakukan dapat dilihat pada Gambar 1. Proses penjejakan diawali dengan tahap inisialisasi parameter wajah dengan metode ambang batas biner. Inisialisasi dilakukan pada citra sekuensial yang pertama guna didapatkan kode penanda di setiap fitur yang muncul secara visual. Penelitian ini menggunakan 29 penanda sebagai acuan parameter ekspresi wajah yang ditangkap di kamera. Setelah semua penanda dikenali dan dinamai, untuk setiap penanda dilakukan proses penjejakan fitur di citra selanjutnya. Aspek kesesuaian penjejakan ditandai dengan nilai jarak yang minimum antara fitur acuan terhadap fitur yang sedang diproses. Proses pembentukan topeng ekspresi dilakukan untuk pembangunan basis data ekspresi wajah manusia yang telah ditanamkan fitur penanda sebanyak 29 titik. Hal ini mempunyai peran penting sebagai salah satu prosedur dokumentasi dan pengujian kehandalan sistem secara visual.

Proses penyederhanaan komputasi dilakukan dengan menggunakan proses ambang batas biner. Pendekatan yang dilakukan pada pengambangan dua dimensi adalah metode *otsu* yang memiliki nilai antara 0 dan 1. Metode *otsu* merupakan proses perhitungan secara berulang untuk mendapatkan nilai penyebaran tingkat piksel sehingga dapat dilakukan pemisahan

fitur secara nyata. Pengontrolan cahaya pada saat proses pengambilan gambar dan penetapan latar belakang objek secara konsisten perlu dilakukan supaya hasil yang didapatkan baik. Dalam penelitian ini nilai pengambangan yang digunakan 0,2824 supaya dapat memisahkan fitur dengan baik. Hasil proses pengambangan dapat dilihat pada Gambar 2.



Gambar 1. Rancangan Penelitian Eksperimentasi

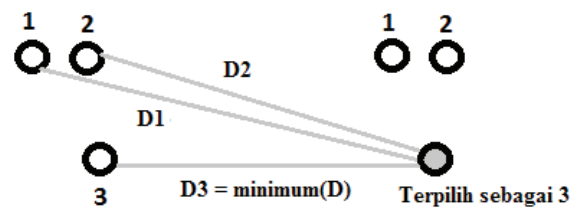


Gambar 2. Proses Ekstraksi Fitur dengan Metode Pengambangan *Otsu*

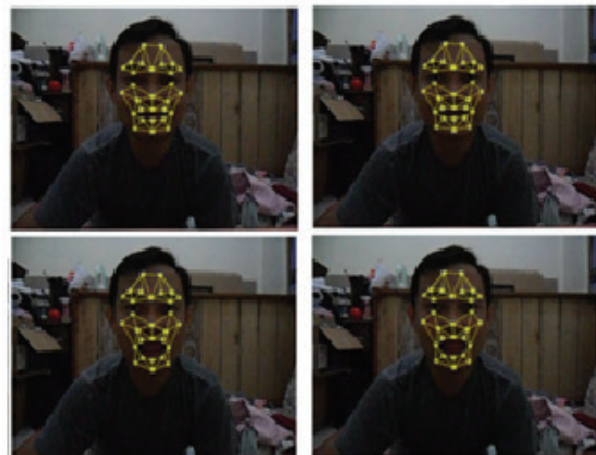
Parameter setiap fitur diproses berdasarkan kesesuaiannya terhadap fitur pada citra acuan. Koherensi temporal sebuah fitur i jika sudah didapatkan fitur acuan dari citra ke-1 dapat dijabarkan dalam tahapan langkah berikut ini:

- a. Pilih fitur di citra terproses sebagai fitur i jika memiliki jarak minimum terhadap fitur i di citra acuan. Perhitungan jarak minimum menggunakan metode *euclidean* untuk setiap titik fitur yang diekstraksi pada citra yang sedang diproses. Jika dijumpai sebuah fitur dengan nilai jarak terminimum terhadap fitur i dibandingkan dengan jaraknya dengan fitur lainnya yang sedang diproses, maka pilihlah fitur tersebut sebagai fitur i di citra tersebut. Visualisasi dapat dilihat pada Gambar 3.
- b. Jika semua fitur sudah dikenali, tentukan citra yang sedang diproses sebagai citra acuan bagi citra selanjutnya. Pada proses ini secara otomatis akan mengubah citra acuan yang digunakan berikut data posisi fitur yang akan dipakai pada proses penjejakan berikutnya.
- c. Lakukan secara berurutan proses a dan b untuk semua citra yang teridentifikasi di kumpulan citra sekuensial. Citra yang diproses bersifat sekuensial sehingga pemrosesan secara berurutan wajib dilakukan untuk menjaga koherensi citra secara temporal.

Setelah semua fitur dikenali dan ditandai di setiap citra sekuensial, pembangkitan topeng ekspresi secara dua dimensi dapat dilakukan sebagai proses validasi akurasi secara visual dan pembangunan basis data ekspresi wajah. Contoh hasil pembangkitan topeng ekspresi dua dimensi dapat dilihat pada Gambar 4.



Gambar 3. Ilustrasi Visual Pemilihan Fitur ke-3



Gambar 4. Sampel Hasil Pembangkitan Topeng Ekspresi Dua Dimensi

Jika dibandingkan dengan hasil metode Chau (Chau, Bremond, & Thonnat, 2011:1), metode yang diusulkan sangat efisien dan memiliki perhitungan komputasi yang ringan dikarenakan tidak adanya tahap pembelajaran pola pada awal pemrosesan. Penjejakan dapat dilakukan secara langsung ke citra sekuensial tanpa proses pendahuluan secara terputus. Nilai akurasi penjejakan secara visual dapat diandalkan dengan capaian yang maksimal.

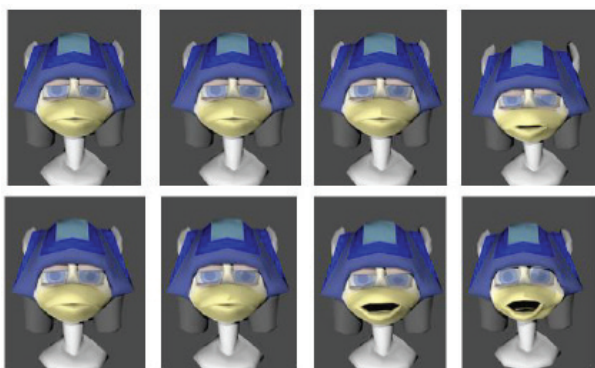
SIMPULAN

Keseluruhan proses dapat dilakukan dengan pengontrolan kondisi alat, ruang, material, dan pencahayaan yang konstan sehingga dapat mengurangi perhitungan komputasi di proses ekstraksi fitur penanda citra. Pada penelitian ini menggunakan nilai ambang batas 0,2824.

Penjejakan fitur berbasis koherensi temporal mampu memproses banyak citra secara berurutan dan memiliki hasil yang baik. Komputasi yang digunakan dalam metode ini sangat efisien dan tidak memerlukan tahapan pembelajaran pola fitur di awal.

Basis data ekspresi wajah yang dibangkitkan dalam bentuk topeng ekspresi dapat digunakan sebagai referensi pembangkitan ekspresi karakter virtual secara sekuensial. Pemetaan dan implementasi hasil penjejakan berbasis koherensi temporal mampu mendukung pembangunan sistem animasi ekspresi wajah secara langsung yang memiliki waktu proses seketika. Hasil salah satu pemetaan ekspresi wajah pada karakter virtual dapat dilihat pada Gambar 5.

Upaya pengembangan sistem dapat dilakukan dengan cara perbandingan metode yang memiliki kesamaan fungsi dan rancangan hasil. Nilai kehandalan dan akurasi dari hasil penelitian ini dapat ditingkatkan pula dengan pengujian sistem menggunakan data visual wajah yang beragam dan memiliki ragam kualitas citra.



Gambar 5. Contoh Hasil Implementasi Pemetaan Ekspresi Wajah

KEPUSTAKAAN

- Bolkart, T., & Wuhrer, S. 2015. "3D Faces in Motion: Fully Automatic Registration and Statistical Analysis". *Computer Vision and Image Understanding*. Vol. 131: 100-115.
- Chau, D. P., Bremond, F., & Thonnat, M. 2011. "A Multi-Feature Tracking Algorithm Enabling Adaptation to Context Variations". *The International Conference on Imaging for Crime Detection and Prevention (ICDP)*. London.
- Dutreve, L., Meyer, A., & Bouakaz, S. 2008. "Feature Points Based Facial Animation Retargeting". *Proceedings of the 15th ACM Symposium on Virtual Reality Software and Technology*.
- Ekman, P., & Friesen, W. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. California: Consulting Psychologists Press, Inc.
- Hjortsjo, C. H. 1970. *Man's Face and Mimic Language*. Malm: Nordens Boktryckeri.
- Johnson, A. E., & Hebert, M. 1997. "Recognizing Objects by Matching Oriented Points". *IEEE International Conference on Computer Vision and Pattern Recognition*. p. 684-692.
- Ko, J. G., Kim, K. N., & Ramakrishna, R. S. 1999. "Facial Feature Tracking for Eye-Head Controlled Human-Computer Interface". *IEEE TENCON'99*. Cheju.
- Parke, F. I. 1972. "Computer Generated Animation of Face. *Proceedings of the ACM Annual Conference*. p. 451-457.
- Parke, F. I., & Waters, K. 2008. *Computer Facial Animation*. 2nd Edition. Massachusetts: AK Peters.

Sobottka, K., & Pitas, I. 1997. "A Fully Automatic Approach to Facial Feature Detection and Tracking". International Conference on Audio and Video-Based Biometric Person Authentication. Switzerland: Crans-Montana.