



Sintesis *Visual Speech* Bahasa Indonesia pada Beberapa Karakter Virtual 3 Dimensi Menggunakan Metode *Radial Basis Function (RBF)* Untuk Mendukung Produksi Film Animasi

Aripin^{1*}, Hanny Haryanto²

^{1,2} Fakultas Ilmu Komputer Universitas Dian Nuswantoro, Jl. Imam Bonjol 207 Semarang, 50131

*email: arifin@dsn.dinus.ac.id

Abstrak. Penelitian ini bertujuan untuk mengimplementasikan hasil sintesis visual speech Bahasa Indonesia pada beberapa karakter virtual 3D menggunakan metode *Radial Basis Function (RBF)* untuk mendukung produksi film animasi khususnya adegan dialog pada beberapa karakter animasi. Metode *RBF* digunakan untuk proses transformasi ruang dari satu karakter ke beberapa karakter lain yang memiliki struktur wajah yang berbeda. Penggunaan metode ini bertujuan agar animasi wajah dari satu karakter dapat ditransformasikan ke karakter wajah lain. Sintesis visual speech merupakan proses kompilasi antara visualisasi gerakan mulut yang disinkronkan dengan suara pengucapan rangkaian fonem berdasarkan input teks Bahasa Indonesia. Untuk memproduksi film animasi, seorang animator harus menganimasikan gerakan mulut dan ekspresi wajah saat karakter virtual berbicara. Seorang animator membutuhkan waktu dan sumber daya yang sangat besar untuk membangun animasi berbicara apalagi pada adegan dialog yang melibatkan beberapa karakter wajah yang berbeda-beda. Hasil penelitian ini dapat membantu animator dalam memproduksi film animasi wajah dengan lebih efisien dan efektif. Seorang animator hanya memerlukan desain wajah beberapa karakter virtual dan teks untuk dialog. Setiap karakter virtual diberikan teks dialog yang dapat dituliskan dengan menggunakan text editor seperti notepad. Selanjutnya karakter virtual 3D dapat memvisualisasikan dialog tersebut. Untuk mengevaluasi aplikasi ini, kami melakukan pengujian subyektif yang dilakukan pada 30 orang pemakai yang menilai kesesuaian antara ucapan dan gerakan mulut berdasarkan teks dialog. Hasil pengujian subyektif dihitung dengan menggunakan metode *Mean Opinion core (MOS)*. Hasil perhitungan metode *MOS* adalah 4.0143 dengan range nilai 1 s.d. 5 yang menunjukkan bahwa tingkat kesesuaian antara ucapan dan gerakan mulut adalah bagus. Hal ini berarti bahwa adegan dialog yang dihasilkan dari karakter virtual 3D dalam penelitian ini mendekati realistik.

Kata kunci : karakter virtual 3D, produksi film animasi, radial basis function, teks Bahasa Indones

PENDAHULUAN

Perkembangan teknologi komputer di Indonesia berdampak terhadap meningkatnya produksi film animasi 3D. Film animasi 3D mempunyai banyak kelebihan dibandingkan film animasi 2D, yaitu lebih menarik, mampu menampilkan grafis yang lebih realistis dan mampu memberi kebebasan yang lebih besar kepada animator untuk merealisasikan imajinasi dan ide-ide kreatif (Orvalho dkk, 2006). Dampak kemajuan teknologi komputer lainnya adalah peningkatan fleksibilitas dan kemampuan futuristik pada sistem interaksi manusia dan komputer (Choe dkk, 2001; H. Prendinger dkk 2004). Di bidang animasi, tuntutan penyajian animasi yang realistis dan luwes semakin tinggi (Deng dkk, 2006).

Beberapa contoh produksi film animasi 3D adalah “*Lord of The Ring*” yang diluncurkan tahun 2001, “*Shrek*”, “*Final Fantasy*” dan “*The Adventure of Tintin*”. Sedangkan film animasi 3D yang diproduksi dari studio animasi di Indonesia dan dibuat versi bahasa Indonesia yang dirilis tahun 2009 adalah “*Meraih Mimpi*” dari versi bahasa Inggris “*Sing to the Dawn*”. Animasi-animasi tersebut diproduksi dengan memerlukan waktu yang lama dan biaya yang sangat mahal, serta memerlukan skill animator yang sangat tinggi. Walaupun demikian, animasi berbicara dari beberapa film animasi tersebut belum menghasilkan animasi berbicara yang realistis.

Wajah manusia merupakan bagian penting dari tubuh agar seseorang dapat mengenal secara visual. Wajah dapat menghasilkan ekspresi yang mencerminkan keadaan mental seseorang. Penganimasian wajah memerlukan tingkat akurasi yang berbeda dan realistis dibandingkan dengan bagian lain dari tubuh manusia (Noh dkk, 2001; Sarah L Taylor, 2012). Gerakan mulut yang sesuai dengan suara yang diucapkan dan ekspresi wajah menjadi suatu hal yang sangat penting dari animasi wajah. Menganimasikan wajah memerlukan ketelitian yang tinggi, ini disebabkan oleh kompleksitas bagian wajah dan masalah *rigging* wajah (Sumner dkk, 2004; H. Prendinger dkk, 2004). *Rigging* adalah pemberian struktur tulang pada animasi 3D agar dapat digerakkan melalui tulang tersebut (A C. Hecker dkk, 2008).

Membangun animasi wajah 3D yang realistik masih menjadi salah satu tantangan besar dalam komputer grafis. Beberapa metode yang digunakan untuk membuat animasi wajah 3D adalah metode *keyframing* dan *motion capture (MoCap)*. Pada metode *keyframing*, animator menentukan *pose* kunci pada frame tertentu. Pada frame-frame selanjutnya, animator menciptakan kurva gerak dengan menggunakan teknik interpolasi. Animator harus dapat menentukan jumlah *keyframe* dengan tepat. Jika jumlah *keyframes* terlalu sedikit atau terlalu banyak akan berpengaruh pada gerakan yang dihasilkan.

Teknologi *motion capture (MoCap)* digunakan untuk menghidupkan karakter (Natasha Kholgade dkk, 2011). Saat ini, penggunaan teknologi *MoCap* untuk menghasilkan animasi wajah 3D yang realistis meningkat tajam. Teknologi



MoCap digunakan untuk merekam gerakan *marker* yang ditempelkan pada aktor. Data gerak yang direkam selanjutnya dapat digunakan untuk menggerakkan karakter virtual. Semakin banyak jumlah *marker* yang digunakan, maka animasi yang diperoleh akan semakin nyata. Animasi wajah yang realistis dapat dibuat dengan menggunakan teknologi MoCap dengan menempatkan penanda di area wajah sesuai dengan struktur wajah (A C. Hecker dkk, 2008). Dalam penelitian ini digunakan 37 (33+4) *marker* untuk menghasilkan animasi wajah yang lebih realistis. Dengan teknologi MoCap, gerakan aktor direkam melalui sensor. Nilai-nilai yang dihasilkan kemudian digunakan untuk membuat gerakan karakter animasi (Dutreve dkk, 2008). Untuk menghasilkan karakter animasi wajah, pada wajah aktor dipasang beberapa *marker* yang dapat ditangkap oleh sensor kamera. Setiap gerakan wajah aktor akan ditangkap oleh sensor sehingga terbentuk database MoCap wajah.

Database MoCap wajah dalam penelitian ini diperoleh dari kamera *motion capture OptiTrack™* yang memiliki resolusi sebesar 480 x 640 pixel dan memiliki kecepatan tangkap sebesar 100 *frame per second (fps)* (Orvalho dkk, 2012). Kami menggunakan OptiTrack ARENA MoCap *Software* untuk merekam gerakan wajah aktor. Database MoCap wajah yang telah diperoleh dari proses ini tersimpan dalam file dengan format C3D yang digunakan sebagai dasar untuk menggerakkan karakter animasi wajah. Selanjutnya, gerakan animasi wajah yang dihasilkan ditransformasikan pada karakter-karakter virtual lain yang memiliki struktur wajah yang berbeda. Proses transformasi menggunakan metode RBF sehingga karakter-karakter virtual lain dapat memvisualisasikan animasi gerakan wajah berdasarkan masukan teks bahasa Indonesia tertentu. Oleh karena itu, hasil penelitian ini dapat mendukung produksi film animasi khususnya pada adegan dialog. Untuk Memvisualisasikan adegan dialog, pada masing-masing karakter virtual diberikan teks berbahasa Indonesia sesuai dengan skenario dialog.

PENELITIAN TERKAIT

Penelitian sintesis *visual speech* untuk bahasa Indonesia dengan menggunakan metode *Syllable Concatenation* (Arifin dkk, 2017). Penelitian ini bertujuan membangun sistem Text-to-Audiovisual Bahasa Indonesia berdasarkan database suara berbasis suku kata. Teknik perangkaian viseme - viseme pada proses sinkronisasi menggunakan metode *syllable concatenation*. Dengan metode ini, viseme-viseme dirangkai dengan mengacu pada bentuk suku kata tertentu dalam bahasa Indonesia. Sistem yang dihasilkan dengan metode ini menunjukkan bahwa animasi visualisasi pengucapan teks lebih halus dibandingkan dengan menggunakan metode berbasis fonem.

Metode *morphing viseme* juga digunakan dalam sintesis *visual speech* untuk bahasa Indonesia (Arifin dkk, 2015). Metode ini digunakan untuk membuat animasi perubahan bentuk dari satu viseme ke viseme lainnya. Dengan cara seperti ini, animasi visualisasi pengucapan fonem yang dihasilkan dari model-model viseme statis menjadi lebih halus.

Penelitian yang menggunakan model-model viseme dinamis untuk melakukan sintesis *visual speech* bahasa Inggris (Sarah L. Taylor dkk, 2012). Gerak artikulator *visual speech* digunakan untuk menghasilkan *gesture visual speech* yang dinamis. Viseme dinamis diaplikasikan untuk menganimasikan pengucapan dengan merangkai unit-unit viseme. Hasil penelitian menunjukkan bahwa tampilan animasi pengucapan yang dinamis sehingga lebih akurat dan menarik secara visual.

Berdasarkan penelitian-penelitian di atas, kami mencoba menerapkan hasil penelitian sintesis *visual speech* bahasa Indonesia pada beberapa karakter virtual dengan menggabungkan metode Radial Basis Function (RBF) sehingga dapat menghasilkan tampilan adegan dialog untuk produksi film animasi.

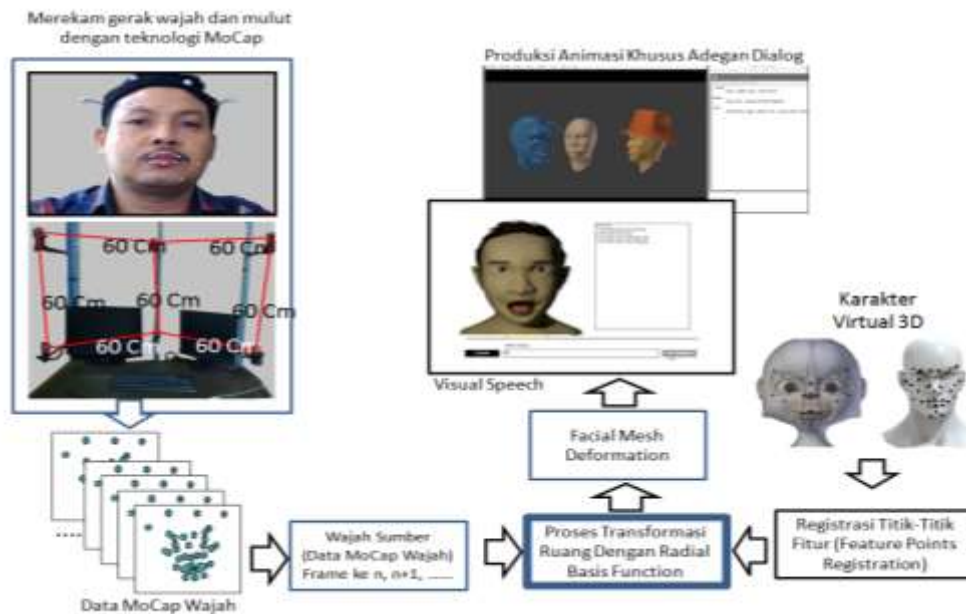
METODE PENELITIAN

Beberapa tahapan penelitian dilakukan untuk menghasilkan animasi berbicara pada beberapa karakter animasi. Tahapan penelitian terdiri dari perekaman wajah aktor yang telah dipasang *marker-marker* dengan teknologi MoCap sehingga terbentuk database MoCap wajah, membuat karakter animasi wajah dan mempersiapkan bone-bone wajah, pemetaan titik *marker* dari data MoCap wajah pada *face template* sehingga diperoleh data MoCap wajah dalam format FBX, melakukan *retargeting* ke beberapa karakter animasi dan penerapan hasil penelitian pada produksi film animasi pada adegan dialog. Gambar 1 menampilkan tahapan-tahapan penelitian secara garis besar.

Facial Motion Capture

Salah satu peralatan yang diperlukan dalam proses perekaman ini adalah *OptiTrack Camera tipe VR100:R2* yang berjumlah 6 buah. Formasi kamera disusun menyerupai busur lingkaran dengan kisaran sudut 120°. Masing-masing kamera dipasang secara orientasi dan *landscape*. Tiga kamera disusun diatas kepala dan tiga kamera disusun setinggi dada. Jarak kamera dengan model sepanjang 60 cm dan kamera OptiTrack yang digunakan seperti ditampilkan pada Gambar 2(a) dan (b).

Selain peralatan di atas, kami juga menggunakan *WebCam* untuk merekam data suara dan video. Data suara direkam ke dalam file dengan format *.wav* dan kami juga menggunakan perangkat lunak *Audacity* untuk melakukan peng-editan suara. Setelah peralatan *MoCap* selesai dilakukan, langkah selanjutnya adalah mempersiapkan *OptiTrack ARENA motion capture Software* agar dapat digunakan untuk merekam data. Ada beberapa langkah yang dilakukan adalah sinkronisasi dan kalibrasi *OptiTrack Camera* serta pembentukan *face template*. Proses sinkronisasi ini bertujuan untuk mengatur fokus 6 buah *OptiTrack Camera* pada titik fokus yang sama. Sedangkan, proses kalibrasi bertujuan untuk mengatur kualitas dan jangkauan area tangkap *OptiTrack Camera*.



Gambar 1. Metode Yang Diusulkan



(a) (b) (c) (d)
Gambar 2. Formasi Kamera (a), Kamera *OptiTrack* Tipe *VR100:R2* (b), Pemasangan *Marker* di Wajah Aktor (c), *Data MoCap Wajah* (d)

Pengambilan data dilakukan pada gerakan wajah aktor. Dalam penelitian ini, kami memasang *marker* pada wajah aktor sebanyak 37 yang terdiri dari 33 *marker* yang diletakan di area wajah dan 4 *marker* diletakan diatas kepala. Jumlah dan formasi *marker* ini mengacu pada template yang telah disediakan oleh *OptiTrack ARENA motion capture Software*. Pemilihan ini juga diharapkan dapat menghasilkan gerakan wajah dan mulut yang lebih baik. Gambar 2(c) dan (d) mengilustrasikan pemasangan *marker* diwajah aktor dan data *MoCap Wajah* yang dihasilkan dari proses perekaman.

Data MoCap Wajah

Data MoCap wajah merupakan database yang berisi data-data gerakan bentuk wajah terutama mulut dari seorang aktor yang telah dipasang beberapa *marker* dan ditangkap menggunakan teknologi *MoCap*. Sedangkan aksi yang dilakukan seorang aktor adalah mengucapkan kalimat-kalimat berbahasa Indonesia sejumlah 1592 kalimat. Proses perekaman bertujuan untuk merekam gerakan *Optical marker* yang terdapat pada wajah aktor pada saat aktor mengucapkan kalimat - kalimat berbahasa Indonesia. Proses ini dilakukan untuk menghasilkan datae *MoCap wajah* seperti terlihat pada Gambar 2(d). Dari hasil proses perekaman data ini, terbentuklah data *MoCap wajah* yang tersimpan dalam file dengan format *C3D*. Kami juga menggunakan *webcam* untuk merekam video dan suara adegan aktor saat proses perekaman. Data video dapat digunakan sebagai acuan untuk pengecekan gerakan mulut pada hasil *MoCap wajah*

dengan gerakan mulut aktor yang ada dalam video. Sedangkan, data suara disimpan dalam format .wav dan nantinya digunakan sebagai salah satu data masukan dalam pengembangan sistem ini. Perangkat lunak *Audacity* digunakan untuk melakukan pengeditan suara sehingga sesuai dengan spesifikasi dan kebutuhan dalam pengembangan sistem ini.

Pemodelan Wajah Karakter Virtual

Pemodelan wajah 3D merupakan proses pembentukan objek wajah 3D menggunakan komputer. Wajah yang dimodelkan adalah wajah manusia dan beberapa wajah karakter fantasi yang akan digunakan dalam penelitian ini. Proses pemodelan wajah 3D terdiri dari beberapa tahap, yaitu pembentukan objek dasar, menggunakan metode dalam memodelkan objek 3D, *lighting* dan menganimasikan gerakan objek sesuai urutan proses. Gambar 3 menunjukkan contoh pemodelan wajah 3D pada karakter virtual manusia.



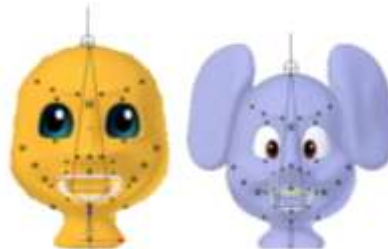
Gambar 3. Pemodelan Wajah 3D Manusia

Ada beberapa metode yang digunakan dalam pemodelan 3D, yaitu NURBS (*Non-Uniform Rational Bezier Spline*), poligon atau *subdivision*. Pemodelan Poligon merupakan bentuk *triangle* dan *rectangle* yang menentukan area permukaan karakter. Tiap poligon menentukan permukaan data dengan menempatkan area poligon sehingga dapat membuat bentuk permukaan. Untuk memperoleh permukaan yang halus, dibutuhkan banyak bidang poligon. Pemodelan NURBS merupakan metode yang paling banyak digunakan untuk membangun sebuah model organik. Ini disebabkan karena kurva NURBS dapat dibentuk hanya dengan tiga titik. Apabila dibandingkan dengan kurva poligon yang memerlukan banyak simpul (*vertices*) yang membuatnya lebih mudah untuk dikontrol. Satu titik CV (*Control Vertex*) dapat mengontrol satu area proses tekstur.

Langkah selanjutnya dalam proses pemodelan adalah *rendering*. Dalam proses *rendering*, seluruh data yang dimasukkan dalam proses pemodelan, yaitu animasi, tekstur, dan *lighting* ditransformasikan menjadi sebuah bentuk output. *Lighting* diperlukan agar model 3D terlihat lebih halus. Ada beberapa *teknik lighting*, yaitu *key light*, *fill light*, *back light*, *kickers light* dan *bounce light*.

Membuat Rigging Pada Wajah (Facial Rigging) Karakter Virtual

Facial rigging merupakan proses pembuatan kendali wajah untuk melakukan pembentukan animasi yang dilakukan oleh animator (Gaurav dkk, 2012; Llaya dkk, 2007). Pada penelitian ini proses *facial rigging* dilakukan pada tiap model dari karakter wajah yang digunakan. Proses pembentukan *facial rigging* berupa pemberian titik fitur pada topeng wajah yang mengacu pada pendekatan *Facial Action Coding System (FACS)* yang digunakan dalam *optitrack motion capture system*. Proses *facial rigging* ini dilakukan secara manual pada tiap model avatar. Hasilnya dari proses *facial rigging* ini berupa data koordinat titik fitur *mesh* wajah dari model avatar yang digunakan, seperti dicontohkan pada Gambar 4.

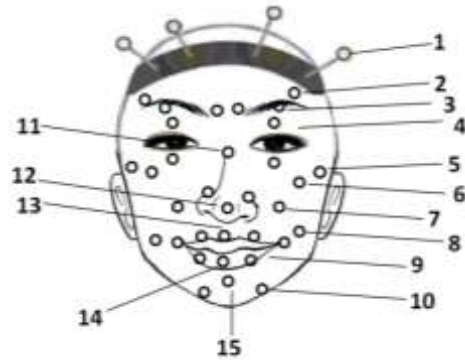


Gambar 4. Facial Rigging pada Karakter Virtual

Dalam proses pemetaan ini, setiap *marker* di data MoCap wajah dipetakan ke masing-masing *rigging* karakter animasi wajah (Kyung Hun Cho dkk, 2013). Pemetaan titik-titik *marker* akan disesuaikan dengan posisi *marker* pada *rigging* tertentu. Model karakter 3D yang dilengkapi dengan *bone-bone* pada titik-titik fitur yang telah ditentukan sebagai kontrol *blendshape* wajah karakter seperti ditunjukkan pada Tabel 1 dan Gambar 5. Model karakter 3D digerakan sesuai dengan data MoCap wajah. Langkah selanjutnya adalah mengkoneksikan data MoCap wajah dengan *bone-bone* pada tiap titik fitur pada wajah karakter.

TABLE I. Area Fitur Wajah Yang Ditempel Marker

No	Face Area	The Number Of Markers	Description
1	Head	4	-
2	Eyebrow	2	Penanda
3	Orbital upper	1	diletakan
4	Eyelids	2	pada wajah
5	Ear	1	bagian kanan
6	Orbital Lower	1	dan kiri
7	Nostril Base & bulge	2	secara simetris
8	Puffer	1	
9	Lip and Mouth	3	
10	Jaw end	1	
11	Nose bridge	1	
12	Nose tip	1	-
13	Upper tip	1	-
14	Lower tip	1	-
15	Chin	1	-



Gambar 5. Area Peletakan Penan pada Wajah

Registrasi Titik-Titik Fitur

Proses registrasi ini dilakukan dengan menyamakan asosiasi penamaan tulang gerak yang terdapat pada *rigging* wajah model dengan rekaman data MoCap (H. M. Kai dkk, 2005). Setelah semua data sudah teregistrasi, proses pengujian dengan pendekatan data MoCap dapat dilakukan.

Pengamatan hasil dan analisis data tangkapan visual dilakukan dengan perbandingan hasil uji coba menggunakan pendekatan tradisional dan hasil uji coba menggunakan pendekatan data MoCap wajah. Hasil tersebut kemudian dijabarkan secara nyata dan dilakukan pembuatan kesimpulan atas hasil pengujian yang dilakukan. Hasil yang didapatkan dalam proses registrasi pada model wajah karakter virtual 3D berupa data koordinat titik fitur model wajah karakter virtual 3D.

Transformasi Ruang Menggunakan Radial Basis Function (RBF)

Radial Basis Functions (RBF) sering digunakan pada aplikasi grafika komputer interpolasi permukaan (Matahari Bhakti Nendya, 2015) dan masalah data interpolasi tersebar (Tamás Umenhoffer, 2012). Pada penelitian ini, RBF digunakan sebagai transformasi ruang. Untuk melakukan hal ini, didefinisikan terlebih dahulu dua buah ruang dengan dua buah himpunan titik fitur. Misalkan S_0 sebagai himpunan titik fitur sumber, T_0 sebagai himpunan titik fitur target, dan N adalah ukuran himpunan. Setiap titik pengendali $\vec{s}_i \in S_0$ memiliki hubungan dengan $\vec{t}_i \in T_0$. Setelah dilakukan pelatihan dengan dua buah himpunan titik fitur tersebut, RBF dapat melakukan transformasi posisi dari ruang sumber ke ruang target dengan menggunakan persamaan (1).

$$F(\vec{s}_j) = \sum_{i=1}^N \vec{w}_i \cdot h(\|\vec{s}_j - \vec{s}_i\|) \quad (1)$$

dimana

$$h(\|\vec{s}_j - \vec{s}_i\|) = \sqrt{\frac{\|\vec{s}_j - \vec{s}_i\|^2}{\|\vec{s}_j - \vec{s}_i\|^2 + SC_j^2}} \quad (2)$$

dan

$$SC_j = \min_{j \neq i} (\|\vec{s}_j - \vec{s}_i\|) \quad (3)$$

Pelatihan jaringan terdiri atas proses perhitungan sistem linier dari persamaan seperti $\vec{t}_j = F(\vec{s}_j)$. Misalkan H adalah sebuah matrik $H_{ij} = h(\|\vec{s}_j - \vec{s}_i\|)$ dan $T_x = (t_1^x \ t_2^x \ \dots \ t_N^x)^t$, t_j^x adalah koordinat x dari \vec{t}_j . Selanjutnya dengan menggunakan kedua persamaan tersebut, persamaan didefinisikan sebagai

$$T_x = H \cdot W_x \quad (4)$$

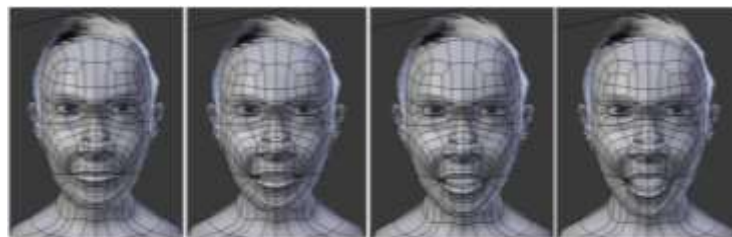
Dimana $W_x = (w_1^x \ w_2^x \ \dots \ w_N^x)^t$. Oleh karena itu, untuk menyelesaikan sistem dihitunglah nilai $W_x = H^{-1}T_x$. Setiap jaringan RBF dilatih untuk setiap sumbu, posisi di ruang target \vec{t} untuk setiap titik \vec{s} dari ruang sumber diperoleh dengan menerapkan transformasi $F(\vec{s})$ yang dihasilkan.

Transformasi ruang menggunakan *RBF* digunakan sebagai cara untuk melakukan transformasi ruang ekspresi wajah antara data MoCap wajah dan target model wajah. Bentuk morfologis wajah manusia secara umum sangat bervariasi dan memiliki perbedaan dengan bentuk morfologis karakter 3D seperti karakter kartun, monster ataupun binatang. Hal ini mengakibatkan pergerakan dari titik fitur yang menggunakan wajah manusia sebagai sumber data

animasi tidak dapat langsung dipergunakan. Ada beberapa hal yang perlu diperhatikan dalam penggunaan data animasi tersebut seperti skala dan orientasi.

Mesh Deformation

Deformasi adalah teknik pengolahan citra yang mengubah gambar dengan menggunakan titik acuan dalam membuat perubahan citra. Sementara, pemodelan mesh merupakan teknik dasar pemodelan yang secara luas digunakan pada banyak perangkat lunak 3D. Pemodelan mesh adalah pemodelan berbasis vertex atau *face by face*, yang hanya mungkin dilakukan pada level vertex. Deformasi *Facial mesh* merupakan upaya untuk merubah bentuk dari *facial mesh* untuk menghasilkan animasi wajah yang berkualitas. Deformasi mesh bertujuan untuk menghaluskan gerakan mulut pada animasi berbicara. Hasilnya adalah model wajah 3D yang dapat menganimasi wajah karakter sama dengan gerakan wajah aktor manusia. Deformasi secara langsung didefinisikan pada permukaan *mesh* wajah yang sering menghasilkan animasi berkualitas. Hal ini mengabaikan anatomi wajah dan struktur otot yang ada. Sehingga fokusnya hanyalah pada pembuatan variasi ekspresi wajah dengan memanipulasi *thin-shell mesh* (Abdul Razzaq dkk, 2015). Kategori ini meliputi proses *morphing* antara model berbeda dan otot semu tersimulasi pada bentuk *splines* (L.-Y. Yu dkk, 2013).



Gambar 6. Visualisasi Perubahan Bentuk Mulut

Untuk mendapatkan hasil secara visual, dilakukan proses deformasi *mesh* dengan bantuan perangkat lunak Blender 3D versi 2.70. Hasil yang didapatkan berupa model karakter virtual 3D yang dapat melakukan ekspresi wajah sama seperti data MoCap wajah yang didapatkan dari hasil perekaman aktor manusia. Gambar 6 menunjukkan visualisasi perubahan bentuk mulut pada karakter virtual 3D. Salah satu metode deformasi yang umum digunakan adalah FFD (*Free Form Deformation*). Metode ini dapat mendeformasi permukaan objek 3D sehingga menghasilkan bentuk kurva bebas. Ketika titik kontrol digerakan, maka simpul-simpul yang dipengaruhi titik-titik kontrol akan bergerak mengikutinya. Deformasi metode ini menghasilkan permukaan objek lebih halus.

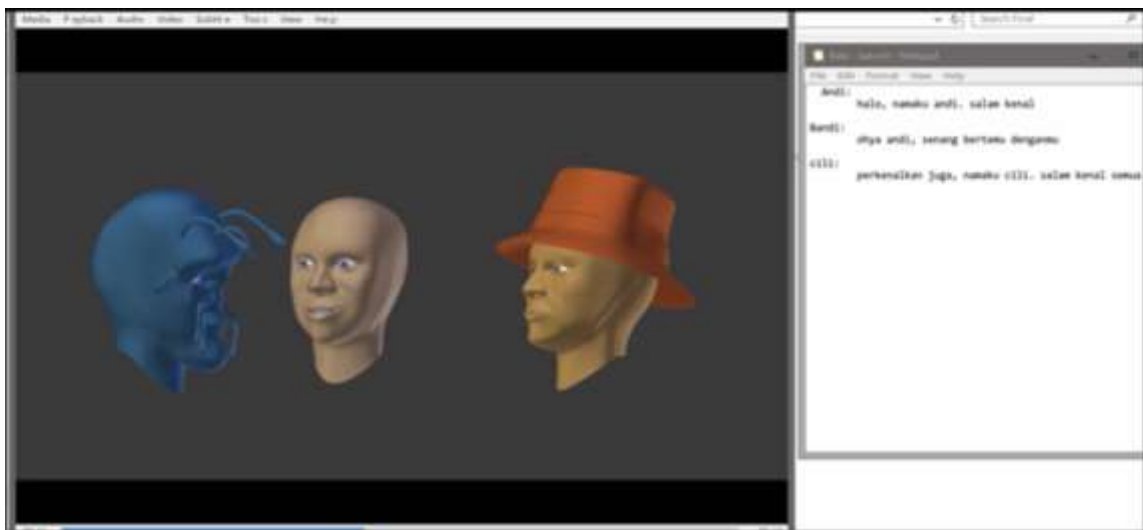
HASIL EKSPERIMEN DAN PEMBAHASAN

Penelitian ini menghasilkan Sistem *visual speech* Bahasa Indonesia berdasarkan teks Bahasa Indonesia yang dimasukkan. Sistem mampu memvisualisasikan gerakan mulut sesuai dengan teks Bahasa Indonesia yang telah dimasukkan seperti yang terlihat di Gambar 5.



Gambar 5. Sistem Visual Speech Bahasa Indonesia yang Dihasilkan

Selanjutnya, data MoCap wajah ditransformasikan pada beberapa karakter virtual yang memiliki struktur wajah yang berbeda. Dari hasil proses transformasi menggunakan RBF diperoleh beberapa karakter virtual yang dapat memvisualisasikan gerakan mulut berdasarkan teks berbahasa Indonesia yang dimasukkan. Tiap karakter virtual diberi nama dan diberikan teks dialog sesuai dengan skenario. Film animasi yang dihasilkan dari penelitian ini dapat menunjukkan adegan dialog dari beberapa karakter virtual seperti yang terlihat di Gambar 6.



Gambar 6. Adegan Dialog Beberapa Karakter Animasi Pada Produksi Film Animasi

Film Animasi wajah yang telah dihasilkan dari penelitian ini diuji oleh 30 responden. Masing-masing responden mengamati dan mengevaluasi kesesuaian gerakan mulut dengan teks dialog yang diberikan. Pada tahap ini, diujikan 7 adegan dialog yang berbeda-beda. Masing-masing responden harus memberikan penilaian dengan beberapa kriteria penilaian seperti yang tertera di Tabel 2. Selanjutnya, hasil penilaian dari masing-masing responden direkap seperti yang disajikan di Tabel 3.

Tabel 2. Kriteria Penilaian MOS

Nilai MOS	Kualitas	Deskripsi
5	Sangat Bagus	Gerakan mulut sangat sesuai dengan pengucapan teks
4	Bagus	Gerakan mulut sesuai dengan pengucapan teks
3	Cukup	Gerakan mulut cukup sesuai dengan pengucapan teks
2	Kurang	Gerakan mulut kurang sesuai dengan pengucapan teks
1	Sangat kurang	Gerakan mulut tidak sesuai dengan pengucapan teks

Tabel 3. Hasil Perhitungan MOS

Adegan Dialog	Jumlah				
	Sangat Kurang	Kurang	Cukup	Bagus	Sangat Bagus
Dialog1	0	4	5	8	13
Dialog2	0	3	7	12	8
Dialog3	0	1	6	13	10
Dialog4	0	0	11	9	10
Dialog5	0	2	5	11	12
Dialog6	0	3	5	12	10
Dialog7	0	1	6	10	13

Berdasarkan pada perhitungan rekapitulasi hasil penilaian oleh responden, selanjutnya dihitung rata-rata penilaian responden terhadap sistem *visual speech* dengan metode MOS. Hasil perhitungan MOS adalah 4,0143 yang menunjukkan bahwa tingkat kesesuaian antara visualisasi gerakan mulut dan teks dialog masing-masing karakter virtual adalah bagus.

KESIMPULAN DAN DISKUSI

Proses transformasi ruang dengan metode *Radial Basis Function (RBF)* telah menghasilkan visualisasi gerakan mulut berdasarkan teks yang diinputkan pada beberapa karakter virtual 3 Dimensi adalah bagus. Berdasarkan hasil pengamatan dan pengujian terhadap karakter-karakter virtual 3D yang dihasilkan, kami dapat mengambil beberapa kesimpulan.

Proses transformasi ruang dengan menggunakan *Radial Basis Function (RBF)* dapat menghasilkan animasi gerakan mulut yang realistis berdasarkan teks bahasa Indonesia pada beberapa karakter virtual 3D. Transformasi ruang RBF merupakan salah satu pendekatan linier yang tidak membutuhkan proses komputasi yang rumit. Pendekatan ini juga dapat memberikan solusi terhadap pembentukan animasi wajah yang sesuai dengan gerakan aktor



dan lebih fleksibel sesuai kebutuhan untuk mengendalikan bone system, blenshape weight dan paramter-paramter lain dalam pembentukan animasi wajah yang spesifik.

Penelitian ini dapat ditingkatkan kekompleksitasannya dengan membangun animasi berbicara yang real time pada beberapa karakter virtual 3D dan menambahkan ekspresi wajah pada karakter-karakter tersebut. Karakter virtual 3D dapat berbicara dan berekspresi seperti marah, kaget, jijik, takut, senang dan sedih sesuai dengan semantik teks yang diinputkan. Dalam pengembangan animasi berbicara yang real time dan berkeekspresi ini, Teknik yang sangat memegang peranan penting dalam proses animasi wajah berekspresi adalah proses kontrol pengendali yang nantinya akan menjadi acuan proses pembangkitan animasi wajah. Semakin baik proses kontrol pengendali dari setiap parameter, maka semakin akurat proses pembangkitan animasi wajah yang berekspresi dan natural.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kepada Kementrian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia yang telah memberikan dukungan dana melalui hibah Penelitian Strategi Nasional Institusi (PSNI) sehingga penelitian ini dapat diselesaikan. Penulis juga mengucapkan terima kasih kepada pengelola Laboratorium *Human Centric System* (HCS) Fakultas Teknologi Elektro Institut Teknologi Sepuluh Nopember (ITS) Surabaya yang telah memberikan fasilitas menggunakan peralatan Motion Capture sehingga proses perekaman data face motion dapat dilakukan.

DAFTAR PUSTAKA

- A C. Hecker, B. Raabe, R. W. Enslow, J. DeWeese, J. Maynard, K. Van Prooijen (2008), *Real-time Motion Retargeting to Highly Varied User-created Morphologies*, In SIGGRAPH 2008, ACM SIGGRAPH 2008 papers, pages 1-11.
- Abdul Razzaq, Zhongke Wu, Mingquan Zhou, Sajid Ali, and Khalid Iqbal (2015), *Automatic Conversion of Human Mesh into Skeleton Animation by Using Kinect Motion*, International Journal of Computer Theory and Engineering, Vol. 7, No. 6, pp. 482-488.
- Arifin, Surya Sumpeno, Mochamad Hariadi, Arry Maulana Syarif (2017), *Development of Indonesian Text-to-Audiovisual Synthesis System Using Syllable Concatenation Approach to Support Indonesian Learning*, International Journal: Emerging Technologies in Learning (iJET), vol. 12 no. 02, ISSN: 1863-0383.
- Arifin, Surya Sumpeno, Mochamad Hariadi, Hanny Haryanto (2015), *A Text-to-Audiovisual Synthesizer for Indonesian by Morphing Viseme*, International Review on Computers and Software, (IRECOS), Vol. 10 No. 11. ISSN: 1828-6003, e-ISSN: 1828-6011, Nopember 2015.
- B. Gaurav, T. Thormählen, H. P. Seidel, and C. Theobalt (2012), *Automatically rigging multi-component characters*, *Computer Graphics Forum*, vol. 31, pp. 755-764, Blackwell Publishing Ltd.
- B. Llaya and J. Popović (2007), *Automatic rigging and animation of 3D characters*, *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp.72.
- Choe, B., Lee, H., & Ko, H.(2001), *Performance driven muscle based facial animation*. The Journal of Visualization and Computer Animation, 12(2) , pp. 67-79.
- Deng, Z., Chiang, P., Fox, P., & Neumann, U. (2006), *Animating blendshape faces by cross-mapping motion capture data*. *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pp. 43-48.
- 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*, ACM.
- H. M. Kai, B. Y. Chen, and M. O. Young (2005), *Motion retargeting and transition in different articulated figures*, in Proc. IEEE Ninth International Conference on Computer Aided Design and Computer Graphics.
- H. Prendinger, M. Ishizuka (2004), *Life-Like Character; Tools, Affective Functions, and Application*, Springer, ISBN : 978-3-540-00867-5.
- Kyung Hun Cho, Xi Chen (2013), *Classifying and Visualizing Motion Capture Sequences Using Deep Neural Networks*, Department of Infomration and Computer Science Aalto University School of Science, Finland.
- L.-Y. Yu, Q. Han, and X. Niu (2013), *An improved contraction-base method for mesh skeleton extraction*, *Multimedia Tools and Applications*, pp.1-14.
- Matahari Bhakti Nendya (2015), *Animasi Ekspresi Wajah pada Karakter Virtual 3D Berbasis Radial Basis Function*, Journal of Animation and Games Studies, Vol. 1 No. 1, ISSN 2460-5662, pp. 67-84.
- Natasha Kholgade, Iain Matthews and Yser Sheikh (2011), *Content Retargeting Using Parameter-Parallel Facial Layers*, Euragraphics / ACM SIGGRAPH Symposium on Computer Animation.
- Noh, J., & Neumann, U. (2001), *Expression cloning*. *Proceedings of the 28th annual conference on computer graphics and interactive techniques*, SIGGRAPH, pp. 277-288.
- Orvalho, V. C., Zacur, E., and Susin, A.(2006), *Transferring Facial Expressions to Different Face Models*, Ibero-American Symposium on Computer Graphics - SIACG. The Eurographics Association,.
- Orvalho, V., Bastos, P., Oliveira, B., and Alvarez, X. (2012), *A Facial Rigging Survey*. In Proc. of the 33rd Annual Conference of the European Association for Computer Graphics- EUROGRAPHICS, May 13-18, Cagliari, Italy, Volume 32, pp. 10-32.



- Sarah L. Taylor, Moshe Mahler, Barry-John Theobald and Iain Matthews (2012), *Dynamic Unit of Visual Speech*, ACM SIGGRAPH Symposium on Computer Animation.
- Sumner, R. W., & Popovic, J.(2010), *Deformation transfer for triangle meshes*. ACM Trans.Graph., 23(3) , pp. 399-405.
- Tamás Umenhoffer and Balázs Tóth (2012), *Facial Animation Retargeting Framework Using Radial Basis Function*, Sixth Hungarian Conference on Computer Graphics and Geometry, Budapest.