

FACE TRACKING DAN DISTANCE ESTIMATION PADA REALTIME VIDEO MENGGUNAKAN 3D STEREO VISION CAMERA

Edy Winarno¹, Agus Harjoko²

¹ Program Studi Teknik Informatika FTI Universitas Stikubank (UNISBANK) Semarang
Jl. Tri Lomba Juang No. 1 Semarang 50241 Indonesia

² Jurusan Ilmu Komputer dan Elektronika FMIPA Universitas Gadjah Mada Yogyakarta
Gedung SIC Lt.3 FMIPA UGM Sekip Utara Bulaksumur Yogyakarta 55281 Indonesia
e-mail: edywinarno3@gmail.com , aharjoko@ugm.ac.id

Abstrak

Face Tracking dan Distance Estimation pada realtime video menggunakan 3D stereo vision camera yang diajukan dalam paper ini adalah sebuah sistem deteksi wajah dan pengukuran estimasi jarak obyek wajah yang terdeteksi menggunakan 3D stereo vision camera. Dalam penelitian ini dikembangkan sistem untuk deteksi wajah menggunakan Haar Cascade Classifier dan untuk pengukuran estimasi jarak wajah dengan kamera menggunakan proyeksi gambar 2D menjadi 3D. Data 3 dimensi pada stereo vision kamera yang digunakan dapat direkonstruksi menggunakan proyeksi 2 Dimensi dari 2 buah titik kamera pada stereo vision camera. Implementasi deteksi wajah (face tracking) dan estimasi jarak pada realtime video menggunakan stereo vision camera yang diusulkan dapat bekerja untuk mendeteksi setiap obyek wajah manusia dengan baik, dan mampu memberikan estimasi jarak antara obyek wajah yang ditangkap dengan stereo vision camera secara riil. Deteksi wajah dan estimasi jarak wajah yang optimal adalah pada kisaran jarak 51-200cm, dengan deteksi wajah dan estimasi jarak yang ideal adalah pada posisi frontal view. Dari percobaan yang dilakukan dapat dihasilkan sebuah sistem tracking wajah yang robust dan dapat diketahui akurasi perhitungan estimasi jarak dibandingkan dengan jarak riil wajah mencapai 94.74 %.

Kata Kunci : *Face Tracking, distance estimation, realtime, stereo vision, 3D.*

1. PENDAHULUAN

Salah satu topik utama yang menjadi bahan penelitian di bidang komputer dewasa ini adalah penelitian di bidang pengolahan citra dan *computer vision*. Pengolahan citra dan *computer vision* merupakan sebuah penemuan di bidang komputer yang digunakan untuk menghasilkan suatu sistem yang hampir mendekati dengan sistem visual manusia pada umumnya. Dengan perangkat input *image capture* seperti kamera dan *scanner* sehingga memungkinkan penelitian di bidang pengolahan citra dan *computer vision* untuk dijadikan sebagai sistem visual. Salah satu bidang yang menggunakan pengolahan citra dan komputer visi yang saat ini banyak dikembangkan adalah sistem deteksi wajah manusia atau *face tracking*. Sistem deteksi wajah merupakan sebuah metode untuk mengenali wajah dari manusia baik pada gambar diam ataupun gambar bergerak seperti *video realtime* yang diambil dari sebuah kamera atau *webcam*. Pendeteksian wajah biasanya menggunakan algoritma yang mampu mengklasifikasikan wajah manusia.

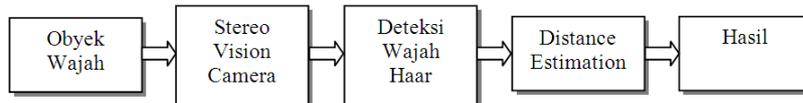
Beberapa penelitian tentang deteksi wajah telah banyak dilakukan sebelumnya, diantaranya penelitian tentang deteksi *image* wajah yang dilakukan oleh Yang [1], penelitian tentang *face tracking* menggunakan *skin colour model* yang dilakukan oleh Zhao dkk [2]. Martinez-Gonzalez dkk [3] melakukan penelitian tentang *realtime face detection* menggunakan *neural networks* dan *skin colour*. Penelitian tentang *Face Tracking* menggunakan *Illumination Invariant Multi-pose* dilakukan oleh Wei dkk [4]. Dari beberapa penelitian tentang deteksi wajah atau *face tracking* yang telah banyak dilakukan, belum banyak dilakukan penelitian menggunakan model untuk melokalisasi atau menentukan estimasi jarak obyek terhadap kamera atau *webcam* yang digunakan.

2. USULAN SISTEM

Metode yang diusulkan dalam penelitian ini adalah sebuah sistem deteksi wajah (*Face Tracking*) dan *Distance Estimation* menggunakan 3D *stereo vision camera*. *Stereo vision camera* digunakan sebagai *video capture* untuk menangkap obyek wajah dan untuk perhitungan estimasi jarak wajah. Data wajah merupakan *video realtime* hasil tangkapan kamera stereo vision. Metode deteksi wajah yang digunakan adalah menggunakan algoritma Haar Cascade Classifier. Untuk penghitungan dalam penentuan estimasi jarak obyek wajah terhadap kamera digunakan metode rekonstruksi menggunakan proyeksi 2 dimensi menjadi 3 dimensi dari 2 buah titik lensa pada stereo vision camera. Bahasa pemrograman yang digunakan adalah Visual C++ yang diaplikasikan menggunakan *library* dari openCV [5]-[7].

A. Perancangan Sistem

Sistem deteksi wajah (*Face Tracking*) dan *Distance Estimation* menggunakan 3D *stereo vision camera* yang dibuat seperti ditunjukkan pada Gambar 1.

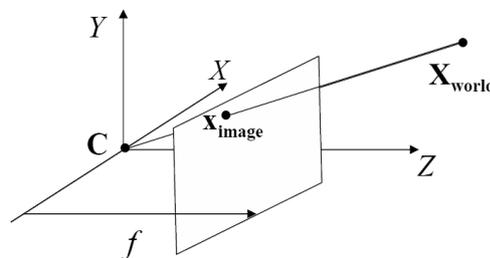


Gambar 1. Diagram Sistem Deteksi Wajah dan Distance Estimation

Sistem pengenalan wajah dan *distance estimation* yang diusulkan ini menggunakan beberapa proses dimulai dari pengambilan wajah secara *realtime* menggunakan *stereo vision camera*. Dengan *stereo vision camera* ini akan dihasilkan data yang diproses oleh 2 buah titik lensa kamera untuk dilakukan proses akuisisi data dan diumpungkan ke dalam bagian deteksi wajah menggunakan *Haar Cascade Classifier* [8]. Estimasi jarak obyek wajah terhadap *stereo vision camera* ditentukan oleh perhitungan yang dilakukan pada bagian *distance estimation*. Pada bagian *distance estimation* ini dilakukan perhitungan jarak menggunakan proyeksi dari gambar 2 dimensi menjadi 3 dimensi menggunakan 2 buah titik lensa pada *stereo vision camera*. Output yang dihasilkan dari sistem deteksi wajah dan *distance estimation* pada *realtime video* menggunakan *stereo vision camera* ini adalah sebuah hasil deteksi wajah yang ditampilkan pada video secara *realtime* dan disertai dengan perkiraan jarak obyek wajah secara *realtime* terhadap kamera yang menangkapnya.

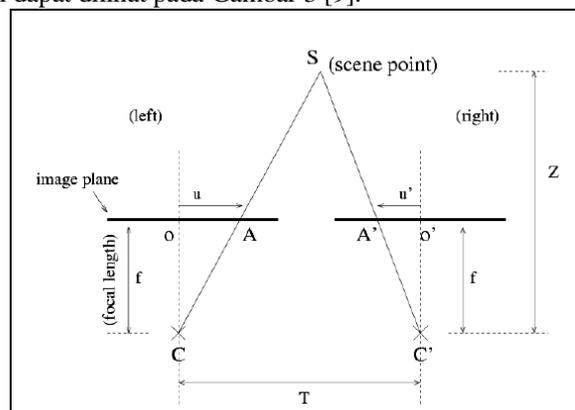
B. Stereo Vision Camera

Stereo vision camera yang digunakan dalam penelitian ini adalah sebuah kamera 3D *stereo vision* yang memiliki resolusi maksimum 800 x 600 dilengkapi 2 buah lensa dengan jarak kedua lensa 6 cm dan memiliki nilai focus minimal 10 cm. Fungsi dari *stereo vision camera* ini adalah menangkap data obyek wajah yang akan diproses menggunakan deteksi wajah *Haar Cascade Classifier*, kemudian dilakukan proses penghitungan estimasi jarak obyek wajah terhadap kamera. Obyek 2 dimensi hasil tangkapan *stereo vision camera* ini kemudian akan diproyeksikan menjadi data 3 dimensi sehingga dapat diukur estimasi jarak obyeknya. Penggunaan *stereo vision* ini akan memudahkan dalam perhitungan estimasi jarak obyek yang dicari, karena perhitungan estimasi jarak tidak akan dapat dilakukan menggunakan *single vision camera*. Model *single vision camera* hanya dapat menangkap gambar 2 dimensi seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Model Single Vision Camera

Pada penelitian ini menggunakan metode pengambilan gambar yang dilakukan oleh 2 buah titik lensa supaya diketahui proyeksi gambar dari 2 sudut yang berbeda, sehingga bisa ditentukan perhitungan untuk mengetahui estimasi jarak obyek yang dicari. Model pengambilan gambar pada *stereo vision camera* yang digunakan pada penelitian ini dapat dilihat pada Gambar 3 [9].



Gambar 3. Model pengambilan gambar pada *stereo vision camera*

Pada gambar di atas diketahui bahwa 2 buah lensa camera (C, C') menangkap gambar obyek pada titik yang sama yaitu pada titik S . Lokasi titik pada 2 posisi gambar dinotasikan dengan A dan A' . T adalah jarak antara 2 buah kamera, sedangkan selisih jarak antara garis A dan A' dengan garis sumbu normal kamera dinyatakan dengan U dan U' . Untuk menentukan jarak Z yang merupakan jarak antara kamera dengan obyek yang ditangkap dapat diperoleh dari persamaan 1 di bawah ini [9].

$$Z = f \frac{T}{U - U'} \quad (1)$$

Nilai Z yang diperoleh dalam persamaan di atas digunakan dalam penelitian ini untuk menentukan estimasi jarak wajah yang diukur dari kamera yang digunakan.

C. Deteksi Wajah menggunakan Algoritma Haar

Metode deteksi wajah yang diusulkan dalam penelitian ini adalah deteksi wajah menggunakan algoritma Haar Cascade Classifier. Dasar utama untuk sistem deteksi obyek menggunakan Haar Cascade Classifier adalah dengan sistem pengklasifikasi menggunakan nilai-nilai intensitas piksel dan menggunakan perubahan nilai kontras antara kelompok piksel segi empat yang saling berdekatan [8]. Varians kontras antara kelompok piksel digunakan untuk menentukan daerah yang terang dan daerah gelap. Dua atau tiga kelompok piksel yang berdekatan dengan varians kontras akan membentuk fitur seperti Haar. Fitur Haar seperti yang ditunjukkan pada Gambar 4 digunakan untuk mendeteksi gambar atau obyek yang akan diklasifikasi. Dalam penelitian ini fitur Haar digunakan untuk mendeteksi obyek bentuk wajah.

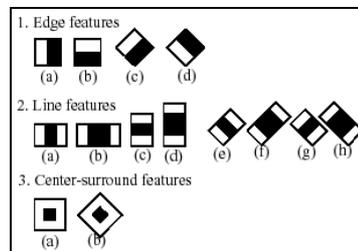
Haar mempunyai algoritma seperti di bawah ini:

1. Setiap fitur Haar terdiri dari dua atau tiga bagian persegi panjang yang berwarna "hitam" dan "putih".
2. Nilai dari fitur Haar merupakan perbedaan antara jumlah tingkat nilai piksel abu-abu di dalam wilayah persegi panjang hitam dan putih :

$$f(x) = \text{Sum}_{\text{black rectangle}}(\text{pixel gray level}) - \text{Sum}_{\text{white rectangle}}(\text{pixel gray level})$$

3. Dibandingkan dengan nilai-nilai piksel baku, fitur Haar dapat mengurangi atau meningkatkan jumlah variabel *in-class* atau *out-of-class*, sehingga membuat klasifikasi lebih mudah.

4. Persegi panjang pada fitur Haar dapat dihitung dengan cepat menggunakan "Integral Image" [10].
 Fitur algoritma Haar dapat dilihat pada Gambar 4.

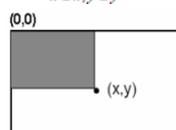


Gambar 4 Algoritma Haar

Fitur yang digunakan dalam pengklasifikasi tertentu ditentukan oleh Gambar 4 (1a, 2b, dll), posisi dalam area dan skala (skala ini tidak sama dengan skala yang digunakan pada tahap deteksi, meskipun dua skala dikalikan). Sebagai contoh, dalam kasus fitur baris ketiga (2c) respon dihitung sebagai perbedaan antara jumlah piksel gambar di bawah persegi panjang yang mencakup seluruh fitur (termasuk dua garis putih dan garis hitam di tengah) dan jumlah piksel gambar di bawah garis hitam dikalikan dengan 3 untuk mengkompensasi perbedaan dalam ukuran bidang. Jumlah nilai-nilai piksel atas wilayah persegi panjang dihitung dengan cepat menggunakan gambar integral.

Persegi panjang sederhana yang merupakan fitur dari suatu gambar yang dihitung menggunakan representasi gambar, disebut *integral image*. *Integral Image* adalah array yang berisi jumlah nilai-nilai intensitas piksel yang terletak pada kiri piksel dan di atas piksel pada lokasi (x,y) inklusif. Jadi jika $A[x,y]$ adalah gambar asli dan $AI[x,y]$ adalah *integral image* maka *integral image* dapat dihitung seperti ditunjukkan pada persamaan di bawah ini dan diilustrasikan pada Gambar 5.

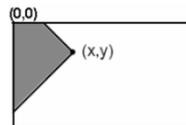
$$AI[x,y] = \sum_{x' \leq x, y' \leq y} A(x',y') \quad (2)$$



Gambar 5 Gambar Integral

Fitur yang diputar 45 derajat, seperti fitur garis yang ditunjukkan dalam Gambar 4(1e), memerlukan representasi intermediate yang disebut gambar yang diputar integral. Gambar integral yang diputar dihitung dengan mencari jumlah dari nilai-nilai intensitas piksel yang terletak pada sudut 45 derajat ke kiri dan ke atas untuk nilai x dan ke bawah ini untuk nilai y. Jadi jika A [x,y] adalah gambar asli dan AR [x,y] maka gambar integral dihitung seperti pada persamaan di bawah ini sebuah diilustrasikan dalam Gambar 6.

$$AR[x,y] = \sum_{x' \leq x, x' \leq x - |y-y'|} A(x', y') \quad (3)$$

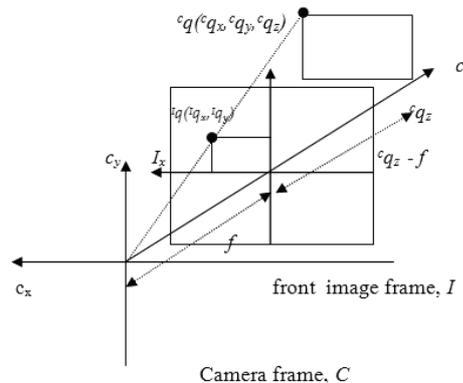


Gambar 6 Gambar Integral yang diputar

Dibutuhkan dua kali untuk untuk menghitung array gambar integral untuk setiap array-nya. Dengan menggunakan gambar integral dan mengambil perbedaan antara 6 sampai 8 elemen array dapat membentuk dua atau tiga persegi panjang yang terhubung. Fitur skala apapun dapat dihitung dengan sangat cepat dan efisien. Untuk menghitung fitur dari berbagai ukuran yang lain juga memerlukan cara yang sama meskipun hanya dua atau tiga piksel.

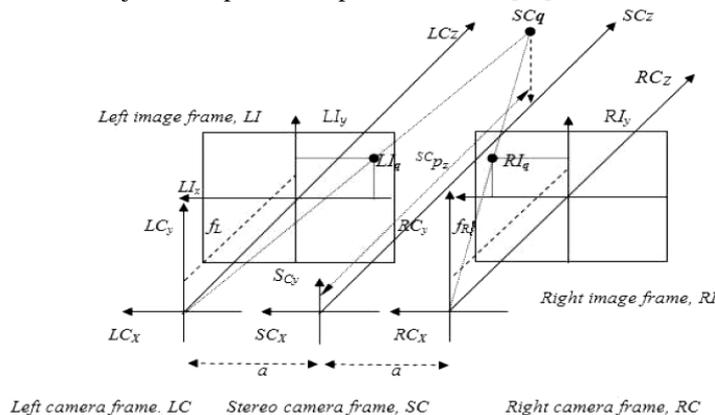
D. Distance Estimation

Distance estimation merupakan pengukuran estimasi jarak antara obyek wajah yang telah terdeteksi menggunakan algoritma Haar dengan stereo vision camera yang digunakan. Jika diumpamakan titik q merupakan titik obyek wajah yang ditangkap oleh kamera, titik Iq(Iqx, Iqy) pada frame gambar adalah proyeksi dari titik Cq(Cqx, Cqy, Cqz) pada frame camera. f adalah panjang fokus lensa stereo vision camera yang digunakan. Proyeksi titik pada frame gambar dapat dilihat pada Gambar 7 [11].



Gambar 7. Proyeksi titik di depan image frame.

Dalam model pencitraan stereo, titik 3 dimensi pada kamera stereo diproyeksikan di sebelah kiri dan kanan image frame. Dengan proyeksi ini titik 3 dimensi dapat ditentukan. Model stereo vision camera yang digunakan dalam sistem deteksi wajah ini dapat dilihat pada Gambar 8 [11].



Gambar 8. Model proyeksi pada stereo vision camera

Stereo vision camera digunakan untuk mendapatkan posisi obyek wajah yang bergerak pada gambar, kemudian dilakukan proses perhitungan untuk memperkirakan jarak obyek wajah yang bergerak. SCq merupakan titik 3D yang akan dicari estimasi jaraknya, RIq_x dan LIq_x adalah posisi image pada kamera kanan dan kamera kiri. a merupakan jarak antara titik tengah dengan masing-masing kamera, f adalah focus (*focal length*) yang sudah ditentukan pada spesifikasi kamera stereo vision. Titik 3D dalam frame kamera stereo yang dinotasikan sebagai SCq dapat direkonstruksi dengan menggunakan proyeksi titik 2D titik menggunakan rumus [11]:

$${}^{SC} \mathbf{q} = \begin{bmatrix} {}^{SC} q_x \\ {}^{SC} q_y \\ {}^{SC} q_z \end{bmatrix} = \frac{2}{{}^{RI} q_x - {}^{LI} q_x} \begin{bmatrix} \frac{1}{2} a ({}^{RI} q_x + {}^{LI} q_x) \\ a {}^{RI} q_y \\ f a \end{bmatrix} \quad (4)$$

3. HASIL DAN PENGUJIAN

Sistem deteksi wajah (*face tracking*) dan *distance estimation* pada realtime video menggunakan *stereo vision camera* ini dilakukan dengan mengambil data beberapa wajah secara acak baik laki-laki maupun perempuan menggunakan *stereo vision camera*. Hasil tampilan deteksi wajah dan *distance estimation* ditampilkan pada 2 gambar layar untuk menampilkan hasil deteksi wajah, dan pada layar yang lain untuk menampilkan estimasi jarak obyek wajah yang diambil. Hasil dari *face tracking* dan *distance estimation* menggunakan *stereo vision camera* dapat di lihat pada Gambar 9.



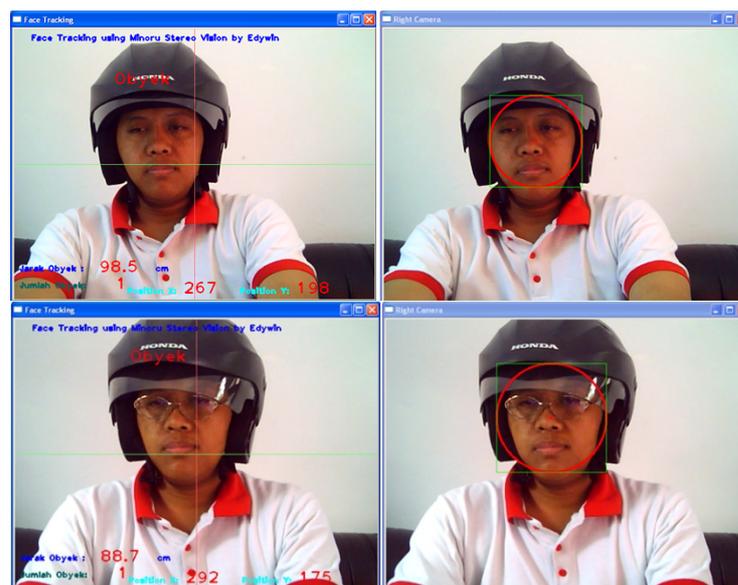
Gambar 9. Hasil deteksi wajah dan *distance estimation* dalam berbagai posisi

Pada gambar 9 di atas terlihat bahwa sistem deteksi ini dapat mendeteksi dan memberikan estimasi jarak wajah dalam berbagai posisi. Deteksi wajah dan *distance estimation* dapat bekerja secara optimal jika posisi wajah dalam posisi frontal view dan seluruh wajah tertangkap oleh kamera. Posisi terdekat wajah agar bias dikenali dan diukur estimasi jaraknya adalah pada posisi minimal seluruh wajah masuk ke dalam frame kamera. Hasil maksimum *tracking* wajah masih dapat dideteksi dalam jarak maksimum sampai dengan lebih dari 2 meter. Hasil deteksi wajah berdasarkan perbedaan jarak dapat dilihat pada contoh Gambar 10 .



Gambar 10. Hasil deteksi wajah dan *distance estimation* berdasarkan jarak

Pada proses pengujian pada *tracking* wajah dan *distance estimation* ini juga dilakukan dengan menggunakan beberapa penutup atau penghalang obyek wajah. Deteksi wajah dan *distance estimation* tetap masih dapat mendeteksi wajah dan estimasi jarak selama titik wajah masih dapat tertangkap kamera dengan jelas. Contoh dari beberapa hasil pendeteksian menggunakan penghalang dan penutup dapat dilihat pada contoh yang ditunjukkan pada Gambar 11.



Gambar 11. Hasil deteksi wajah dan *distance estimation* menggunakan beberapa penghalang atau penutup.

Untuk membuktikan hasil dari sistem *tracking wajah* dan *distance estimation* pada *realtime video* menggunakan *stereo vision camera* ini dilakukan beberapa pengujian untuk membandingkan kinerjanya. Pengujian dilakukan dengan menggunakan sampel 10 obyek wajah yang berbeda. Pengujian dilakukan dengan

mengambil gambar wajah dari berbagai jarak yang berbeda untuk mengetahui seberapa jauh target wajah bisa dideteksi. Nilai 1 menunjukkan obyek dapat terdeteksi, nilai 0 menunjukkan obyek tidak dapat terdeteksi. Hasil pengujian dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1 Pengujian Deteksi Wajah

Obyek	Distance (cm)					
	0-50	51-100	101-150	151-200	201-250	>251
1	1	1	1	1	1	0
2	0	1	1	1	1	0
3	1	1	1	1	1	0
4	1	1	1	1	0	0
5	1	1	1	1	1	0
6	1	1	1	1	1	0
7	0	1	1	1	1	0
8	1	1	1	1	0	0
9	1	1	1	1	1	0
10	1	1	1	1	0	0

Dari hasil yang ditunjukkan pada tabel diatas dapat diketahui bahwa deteksi wajah dan estimasi jarak terdekat bisa dideteksi pada jarak minimal sekitar 50 cm, tergantung pada lebar dan tinggi wajah seseorang yang dapat ditangkap oleh kamera. Untuk jarak maksimal yang dapat dideteksi dan diketahui estimasi jaraknya adalah pada posisi maksimal sekitar 250 cm. Posisi ideal dan terbaik untuk tingkat akurasi deteksi wajah dan estimasi jarak adalah pada rentang jarak antara 51-200 cm. Pada rentang jarak ini semua obyek wajah dapat dideteksi dan diukur jaraknya dengan baik.

Untuk mengetahui akurasi estimasi jarak yang dihasilkan oleh sistem, dilakukan perbandingan pengukuran antara pengukuran riil dan hasil estimasi yang ditunjukkan oleh sistem. Hasilnya dapat dilihat pada Tabel 2.

Tabel 2 Perbandingan estimasi jarak obyek

Obyek	Real Distance (cm)						
	60	75	100	125	150	175	200
1	58.5	75.1	94.3	118.6	137.6	162.7	185.7
2	57.6	74.8	93.2	117.9	138.9	161.5	185.9
3	58.2	75.6	92.8	119.4	137.5	165.5	186.3
4	58.6	74.3	94.5	118.5	136.2	163.9	183.9
5	57.7	74.1	94.3	117.8	136.5	167.3	186.2
6	58.9	75.3	93.7	119.3	137.3	165.4	185.4
7	57.3	74.7	94.2	117.2	135.7	164.9	185.3
8	59.4	74.8	94.7	118.7	138.2	162.4	182.6
9	58.7	75.3	92.9	118.6	135.6	164.7	182.8
10	57.5	74.8	94.6	119.1	135.3	163.9	185.3
Average	58.24	74.88	93.93	118.51	136.88	164.22	184.94
%	97.06	99.84	93.92	94.80	91.25	93.84	92.47

Dari percobaan yang telah dilakukan dengan membandingkan antara jarak riil obyek yang diukur secara manual (*real distance*) dengan hasil yang ditunjukkan pada sistem ini menunjukkan bahwa akurasi dari pengukuran berbagai jarak antara 60cm-200cm menghasilkan akurasi sebesar 94.74%. Akurasi tertinggi pada pengukuran dengan jarak 75cm yaitu sebesar 99.84 %.

4. KESIMPULAN

Hasil dari implementasi deteksi wajah (*face tracking*) dan estimasi jarak pada *realtime video* menggunakan *stereo vision camera* yang usulkan dapat disimpulkan bahwa teknik deteksi wajah dan *distance estimation* ini dapat bekerja untuk mendeteksi setiap wajah obyek manusia dengan baik, dan mampu memberikan estimasi jarak antara obyek wajah yang ditangkap dengan *stereo vision camera* yang digunakan.

Deteksi wajah dan estimasi jarak wajah maksimal dan optimal adalah pada jarak 51-200 cm, dengan deteksi wajah dan estimasi jarak yang ideal adalah pada posisi *frontal view*. Untuk hasil akurasi perhitungan estimasi jarak wajah dilakukan dengan membandingkan estimasi jarak yang ditunjukkan pada sistem dengan jarak riil obyek wajah. Akurasi pengukuran estimasi jarak mencapai 94.74%

DAFTAR PUSTAKA

- [1] Yang, M. (2002) *Detecting faces images: A survey*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24 no. 1, pp.34-58
- [2] Zhao, X., Hui, Y.,(2009) *Face Tracking Based on Fusion Skin Color Model and Optical Flow Algorithm*, International Conference on Wireless Networks and Information Systems Issue Date:December 2009 pp. 89-92
- [3] Martinez-Gonzalez, A.N., Ayala-Ramirez, V., (2011) *Real Time Face Detection Using Neural Networks* Mexican International Conference on Artificial Intelligence, December 2011 pp. 144-149 2011
- [4] Wei, W.,Zhang, Y., Lin, Z., (2009) *Illumination Invariant Multi-pose Face Tracking* Shaanxi Key Laboratory of Speech & Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
- [5] Robert L., *OpenCV 2 Computer Vision Application Programming Cook Book*, Packt Publishing, Birmingham, UK, 2011.
- [6] Low, A., *Introductory Computer Vision and Image Processing*, McGraw-Hill , UK, 1991.
- [7] Jahne, B., Haubecker, H., Geibler, P., *Handbook Of Computer Vision And Applications*, Volume 2, Signal Processing and Pattern Recognition, 1999, Accademic Press.
- [8] Philip I. W., John F., *Facial Feature Detection Using Haar Classifier*, Cosortium for Computing Sciences in College, JCSC, 2006.
- [9] Hendriks, E. A., Redert, P. A., *Converting 2D to 3D: A Survey*, Information and Communication Theory Group (ICT) Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, the Netherlands, 2005.
- [10] Nixon, M, S., Aguado, A, S., *Feature Extraction and Image Processing second edition.*, Academic Press, 2008.
- [11] Budiharto, W., Santoso, A., Purwanto D., Jazidie, A., *Multiple Moving Obstacles Avoidance of Service Robot using Stereo Vision*, TELKOMNIKA, Vol.9, No.3, December 2011, pp. 433-444 e-ISSN: 2087-278X (p-ISSN: 1693-6930)