

ESTIMASI KEDALAMAN PADA CITRA DENGAN *CONDITIONAL RANDOM FIELD*(CRF) DAN *STRUCTURED SUPPORT VECTOR MACHINE* (SSVM) MENGGUNAKAN FITUR FFT

Derry Alamsyah, Muhammad Rachmadi

¹Prodi Teknik Informatika STMIKGlobal Informatika MDP

²Prodi Sistem Informasi STMIKGlobal Informatika MDP

^{1,2}Jl. Rajawali no. 14 Palembang

E-mail :derry@mdp.ac.id, rachmadi@mdp.ac.id

ABSTRAK

Citra Kedalaman merupakan citra yang dapat digunakan untuk segmentasi objek, 3D Model dan visi pada robot. Penentuan nilai kedalaman dari sebuah citra warna memiliki tantangan tersendiri dan bidang aktif penelitian. Citra kedalaman dapat diestimasi dengan menggunakan model stokastik CRF. Selanjutnya model CRF digunakan sebagai model untuk pembobotan pada Model SSVM. Penggunaan CRF-SSVM untuk estimasi citra kedalaman didukung ekstraksi fitur domain frekuensi yaitu FFT. Hasil yang ditunjukkan dengan menggunakan pendekatan ini sebesar 69.04%

ABSTRACT

Depth Image is an image that can be used for object segmentation, 3D Model and vision on robot. The determination of the depth value of a color image has its own challenges and active field of research. The depth image can be estimated using the stochastic model of CRF. Furthermore the CRF model is used as a model for weighting on the SSVM Model. The use of CRF-SSVM for depth image estimation supported feature frequency domain extraction i.e. FFT. The result shown by using this approach is 69.04%

Keywords: *Supapixel, FFT, CRF, SSVM*

1. Pendahuluan

Citra kedalaman merupakan citra satu kanal dengan nilai jarak sebagai nilai intensitasnya. Citra ini disebut juga sebagai citra 3D. Citra kedalaman dapat digunakan pada *mobile robot* [1], segmentasi objek pada citra [2]-[6] dan visualisasi 3D [7][8]. Citra kedalaman dapat diperoleh melalui sensor, seperti Kinect ataupun ASUS Xtion. Kedua sensor menggunakan sinar inframerah sebagai pemberi informasi jarak. Sensor memberikan hasil akurasi nilai jarak yang baik pada citra kedalaman [1]-[3]. Selain melalui sensor citra kedalaman dapat diperoleh dengan cara estimasi [9]-[17].

Estimasi citra kedalaman dapat dilakukan dengan menggunakan satu citra (*single image*) [9]-[14] dan beberapa citra (*multiple image*) seperti video [15]-[17], dimana baik satu ataupun beberapa citra, estimasi citra dilakukan melalui citra RGB. Estimasi citra kedalaman menggunakan satu citra dapat dilakukan melalui proses pixel per pixel (*pixelwise*) [9][10], pengelompokan pixel [11][12], dan ekstraksi fitur [13][14].

Fitur dalam citra dapat dibagi menjadi tekstur atau intensitas warna, bentuk dan frekuensi. Estimasi pada citra kedalaman yang memanfaatkan fitur tekstur dilakukan dengan mempelajari pola warna setiap pixel. Pola warna ditentukan berdasarkan pola warna tetangga dengan menggunakan *Convolutional Neural Network*(CNN)[9][10]. Pola tersebut digunakan sebagai fitur untuk proses estimasi. Selanjutnya estimasi dilakukan dengan menggunakan

pendekatan stokastik seperti *Markov Random Field*(MRF) [12]. Sebagai perbaikan dari MRF, proses estimasi stokastik dapat dilakukan dengan *Conditional Random Field* (CRF) [9]-[11].

Selanjutnya, estimasi citra kedalaman dengan menggunakan fitur bentuk dapat dilakukan dengan menentukan fitur-fitur tertentu sebagai ciri dari suatu jarak [14]. Selain itu fitur gradien pada citra juga dapat digunakan sebagai fitur untuk mengestimasi citra kedalaman [13]. Seperti halnya pemanfaatan fitur tekstur untuk estimasi citra kedalaman, fitur bentuk juga menggunakan proses stokastik dalam menentukan nilai jarak pada citra kedalaman [13][14].

Sementara itu, Fitur frekuensi merupakan fitur sering digunakan untuk pengenalan objek, sebagai contoh sistem pengenalan wajah [18]. Lebih lanjut, penggunaan fitur frekuensi pada sistem pengenalan ekspresi wajah [19]. Selain itu, fitur ini juga digunakan untuk pengenalan biometrik yaitu pada pengenalan sidik jari [20].

Fitur frekuensi merupakan fitur yang didapat dengan mengubah domain citra spasial ke domain frekuensi, dalam hal ini disebut transformasi. Jenis-jenis trans-formasi pada citra adalah transformasi fourier, sinu-soidal, walsh-hadamard, slant, dan wavelet. Fitur ini dapat digunakan untuk estimasi citra kedalaman, sebagai contoh [21] menggunakan transformasi wavelet dalam penelitiannya. Lebih lanjut, [21] menggunakan MRF sebagai metode untuk estimasi citra kedalaman.

CRF merupakan metode untuk memodelkan permasalahan stokastik. Metode ini merupakan lanjutan dari metode MRF. Hal tersebut memungkinkan pe-ningkatan performa sistem untuk mengestimasi citra kedalaman [9]. Permasalahan estimasi citra kedalaman dapat digolongkan kedalam permasalahan klasifikasi, dalam hal ini klasifikasi wilayah berdasarkan jarak. Salah metode untuk klasifikasi adalah *Support Vector Machine* (SVM). Metode ini memiliki kemampuan yang baik dalam mengklasifikasi objek. Kemampuan SVM dalam mengklasifikasikan objek dapat dimanfaatkan untuk menyelesaikan permasalahan klasifikasi terstruktur atau yang bersifat stokastik. Hal tersebut dicapai dengan menggabungkannya dengan model stokastik yaitu CRF. Sejatinya SVM digunakan untuk menyelesaikan masalah non stokastik, oleh karena itu perlu dilakukan perubahan SVM. *Structured Support Vector Machine* (SSVM) merupakan bentuk SVM yang dapat digunakan untuk permasalahan stokastik, dimana metode ini didukung oleh CRF dalam penyelesaian masalahnya [22].

Dalam prosesnya estimasi citra kedalaman dapat dilakukan dengan terlebih dahulu mengelompokkan sejumlah pixel. Pengelompokan ini ditujukan untuk mengurangi proses estimasi pixel dan mendukung performas sistem dalam mengestimasi citra kedalaman [11][12].

Estimasi citra kedalaman pada domain frekuensi merupakan masalah yang masih sedikit diselesaikan [21]. Untuk menyelesaikan permasalahan tersebut, penelitian ini menggunakan *Conditional Random Field* dan *Structured Support Vector Machine* (CRF-SSVM) sebagai metode untuk estimasi. Sementara itu ekstraksi fitur yang digunakan adalah transformasi fourier berupa *Fast Fourier Transform* (FFT) pada kelompok pixel. Pengelompokan pixel dilakukan dengan menggunakan Superpixel.

II. Studi Literatur

2.1. Superpixel SLIC

Superpixel merupakan gabungan dari beberapa pixel untuk membentuk pixel baru. Pengelompokan pixel ini didasarkan pada metode *gradient ascent* atau yang disebut sebagai SLIC superpixel. Pengelompokan ini diinisiasi berdasarkan warna dan kedekatannya. Pengelompokan dilakukan menggunakan metode k-means pada data berdimensi 5, yaitu warna dan posisi pixel [23]. Pusat dari superpixel diinisiasi pada regular grid dengan step pada persamaan (1).

$$R = \lceil \frac{1}{\sigma} \rceil$$

Selanjutnya setiap pixel dikelompokkan berdasar-kan kedekatan dengan superpixel terdekat

melalui jarak yang ditunjukkan pada persamaan (2).

$$d(x_n, S_j) = \|I(x_n) - I(S_j)\|_2 + \frac{\beta}{R} \|x_n - \mu(S_j)\|_2 \quad (2)$$

Dimana merupakan parameter kepadatan. Prosedur ini terus diperbaharui hingga nilai dari konvergen.

2.2. Fast Fourier Transform (FFT)

Transformasi Fourier merupakan transformasi yang mengubah citra yang berada pada domain spasial ke domain frekuensi. Transformasi ini merupakan transformasi berupa sinyal sinus. Trasformasi fourier dinyatakan pada persamaan (3), sementara transformasi baliknya dinyatakan pada persamaan (4).

Transformasi Fourier 1D:

$$F(u) = \sum_{x=0}^{N-1} f(x) \cos\left(\frac{2\pi ux}{N}\right) - i \sin\left(\frac{2\pi ux}{N}\right) \quad (3)$$

Invers transformasi fourier 1D:

$$f(x) = \frac{1}{N} \sum_{u=0}^{N-1} F(u) \cos\left(\frac{2\pi ux}{N}\right) + i \sin\left(\frac{2\pi ux}{N}\right) \quad (4)$$

Sementara itu, Fitur dari transformasi fourier didapatkan melalui 2 jenis nilai, yaitu spektrum pada persamaan (5) dan sudut pada persamaan (6).

Spektrum (*Magnitude*):

$$\frac{|F(u)|}{N} = \frac{1}{N} \sqrt{Re(u)^2 + Im(u)^2} \quad (5)$$

Sudut (*Angle*):

$$\theta(u) = \tan^{-1}\left(\frac{Im}{Re}\right) \quad (6)$$

2.3. Conditional Random Field (CRF)

Pemodelan stokastik dapat dinyatakan dalam model CRF. Model ini mempertimbangkan setiap nilai bobot berdasarkan nilai dari node sebelum dan saat $(f_k(y_t, y_{t-1}, x))$. Bobot yang dibentuk pada model merupakan nilai bobot terbaik. Bobot tersebut menentukan seberapa besar pengaruh dari node pasangan kejadian $f_k(y_t, y_{t-1}, x)$, dengan fitur/data observasi x . Nilai $f_k(y_t, y_{t-1}, x)$ berada pada rentang $[0, 1]$. Dimana 1 untuk yang terjadi dan 0 untuk tidak terjadi. Model CRF dibentuk dengan menggunakan persamaan (7) dan (8) [24].

Model CRF:

$$(1) p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^T \exp\left\{\sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x)\right\} \quad (7)$$

Normalisasi Faktor:

$$Z(x) = \sum_y \prod_{t=1}^T \exp\left\{\sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x)\right\} \quad (8)$$

2.3. Structured Support Vector Machine (SSVM)

SSVM merupakan bentuk terstruktur dari metode SVM. Meskipun terstruktur model yang digunakan masih merupakan model linier [22][25]. Persamaan model SSVM dinyatakan oleh persamaan (9). Persamaan tersebut merupakan persamaan yang digunakan untuk mengklasifikasikan data ().

Persamaan Klasifikasi (Classifier):

$$f_{ssvm}(x) = \operatorname{argmax}_y \langle w, \phi(x, y) \rangle$$

Pembentukan Model (Training Phase):

Dalam membentuk persamaan klasifikasi dibutuhkan bobot terbaik. Pembentukan bobot ini disebut sebagai fase pelatihan. Untuk mengestimasi nilai bobot terbaik dilakukan dengan memenuhi persamaan optimasi (10) dengan syarat persamaan (11).

Optimasi bobot:

$$\min_{w, \xi} \frac{\lambda}{2} \|w\|^2 + \frac{1}{n} \sum_{i=1}^n \xi_i$$

(10)

dengan syarat,

$$\langle w, \psi_i(y) \rangle \geq L(y_i, y) - \xi_i \quad \forall i, \forall y \in \bar{y}$$

(11)

Dimana

$$\psi_i(y) := \phi(x_i, y_i) - \phi(x_i, y)$$

(9)

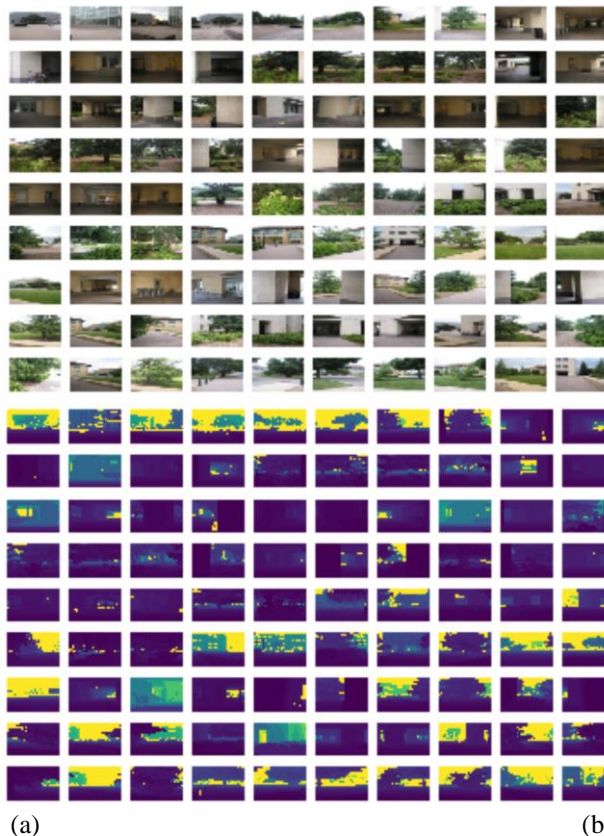
(12)

Bobot terbaik adalah bobot yang memiliki nilai minimum pada persamaan (10).

III. Metodologi Penelitian

3.1. Pengumpulan Data

Data pada penelitian ini merupakan dataset dari Make 3D Dataset pada Cornell University. Dataset terdiri dari 90 citra warna dan 90 citra kedalaman ditunjukkan pada Gambar 1. Jarak terdekat pada citra kedalaman di-tunjukkan oleh warna biru (gelap) sementara terjauh ditunjukkan oleh warna kuning (terang) serta hijau untuk menengah.



Gambar 1. a. Citra warna (RGB image) dan **b.** citra kedalaman (Depth Image)

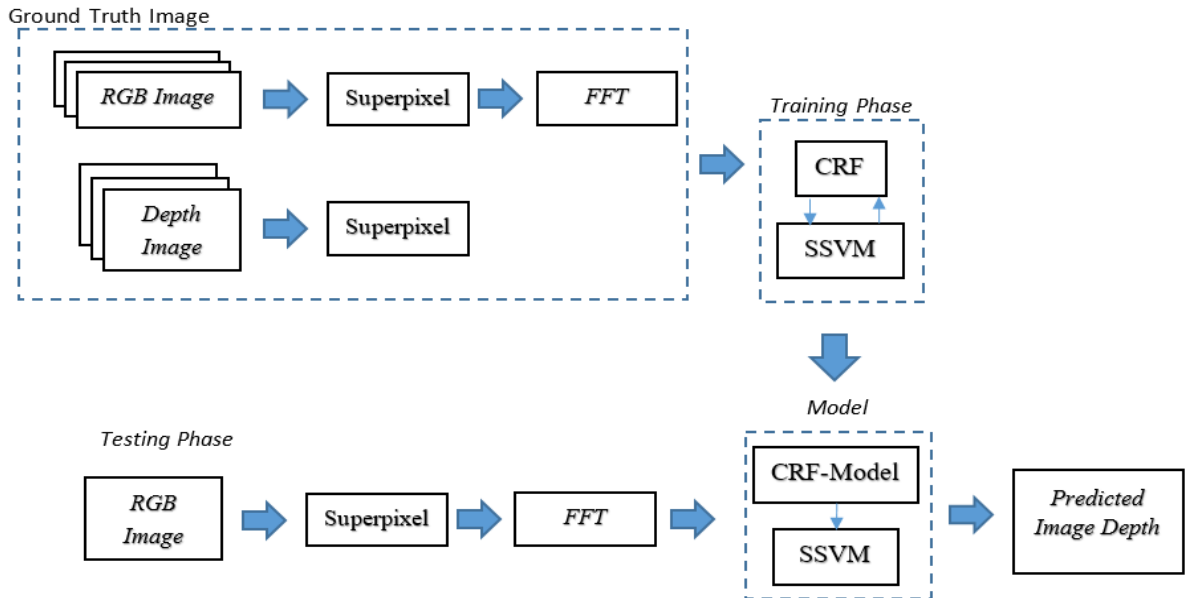
3.2. Rancangan Pendekatan

Pendekatan untuk estimasi citra kedalaman dilakukan dengan dua fase (phase). Pada setiap fase menggunakan superpixel dan FFT sebagai pra proses dan ekstraksi fiturnya. Fase pertama adalah

fase latih dan fase kedua adalah fase uji. Fase latih menghasilkan model CRF-SSVM, sementara fase selanjutnya menghasilkan citra kedalaman. Alur kedua fase ditunjukkan oleh Gambar 2.

Fase Latih dibagi kedalam: Pra-proses, Ekstraksi Fitur, Pelatihan Model. Pada bagian pra-proses, kedua citra (warna dan kedalaman) disegmentasi menggunakan superpixel. Selanjutnya hasil segmentasi citra warna dilakukan ekstraksi fitur berupa nilai rata-rata spektrum. Masing-

masing fitur dipasangkan dengan label yaitu nilai jarak dari citra kedalaman tersegmentasi. Tahap selanjutnya bobot dilatih berdasarkan model CRF. Bobot ini kemudian digunakan untuk membentuk model SSVM.



Gambar 2. Rancangan Pendekatan

Fase Uji dibagi kedalam: Pra-proses, Ekstraksi Fitur, Estimasi Citra Kedalaman. Pada pra-proses hanya dilakukan pada citra warna. Hasil pra-proses berupa citra warna tersegmentasi. Selanjutnya dilakukan ekstraksi fitur pada tiap-tiap segmentasi. Kedua tahap ini serupa dengan kedua tahap awal fase latih. Selanjutnya citra kedalaman didapat dengan menggunakan model CRF-SVM yang didapat dari fase latih.

Hasil performa CRF-SSVM lebih baik dibandingkan dengan *Structured Perceptron* yaitu 69.04% berbanding 48.9%. Hasil keduanya ditunjukkan oleh Tabel 1.

3.3. Evaluasi

Evaluasi dilakukan pada citra kedalaman hasil estimasi dan citra kedalaman dataset (*ground truth*) dengan menggunakan rumusan (13).

$$acc = \frac{\sum_{i=1}^n 1_c}{n}$$

(13)

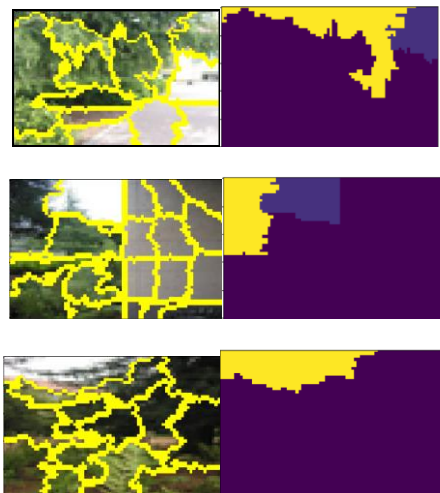
Dimana merupakan label pada dataset dan merupakan label yang diberikan hasil estimasi. Sementara banyaknya label pada setiap citra dinyatakan oleh *n*.

IV. Hasil dan Pembahasan

Hasil estimasi citra kedalaman ditunjukkan oleh Gambar 3. Performa estimasi citra kedalaman dengan menggunakan CRF-SSVM dibandingkan dengan metode *Structured Perceptron* dengan keduanya sama-sama menggunakan FFT sebagai fiturnya.

Tabel 1. Hasil

Metode	Data Latih (%)	Data Uji (%)
CRF-SSVM	65.98	69.04
Structured Perceptron	43.7	48.9





Gambar 3. Citra Input Tersegmentasi (kiri) dan Estimasi Citra Kedalaman (Kanan)

Performa yang baik pada pendekatan yang diajukan, menunjukkan bahwa fitur fourier dari ekstraksi fitur FFT dapat memberikan ciri yang unik untuk Model CRF-SSVM. Sementara untuk metode estimasi *structured perceptron* informasi ciri tersebut belum mampu memberikan akurasi yang baik.

Hasil yang baik tersebut ditunjukkan pada gambar 3 dimana untuk objek terdekat berupa tanaman dan bangunan dinyatakan dengan warna ungu gelap dan yang terjauh ditunjukkan dengan warna kuning terang.

V. Kesimpulan

Pendekatan CRF-SSVM sebagai model estimasi citra kedalaman dengan menggunakan ekstraksi fitur FFT menghasilkan akurasi yang cukup baik sebesar 69.04%. Fitur fourier dari FFT memberikan informasi ciri yang baik pada model CRF-SSVM dalam estimasi Citra Kedalaman.

Referensi

- [1] T. Hamedani and A. Harati, "Multi Scale CRF Based RGB-D Image Segmentation Using Inter Frames Potentials", *International Conference on Robotics and Mechatronics (ICRoM)*, IEEE, Pp. 920-925, 2014.
- [2] S.S. Mirkamali and P. Nagabhushan, "RGBD Image Segmentation". *Iranian Conference on Machine Vision and Image Processing (MVIP)*, IEEE, Pp. 41-44, Iran, 2015.
- [3] A.C. Mueller and S. Behnke, "Learning Depth-Sensitive Conditional Random Fields for Semantics Segmentation of RGB-D Images", *International Robotics and Automation (ICRA)*. IEEE. Hong kong, 2014.
- [4] W. Zhang and M. Li, "MRF and CRF Based Image Denoising and Segmentation", *International Conference Digital Home (ICDH)*. IEEE, Pp. 128-131, China, 2014.
- [5] Y. Qi, G. Zhang, Y. Qali, and Y. Li, "Object Segmentation Based on Gaussian Mixture Model and Conditional Random Fields", *Information and Automation (ICIA)*, IEEE, Pp. 900-904, China, 2016.
- [6] F. Liu, G. Lin, R. Qiao, and C. Shen, "Structured Learning of Tree Potentials in CRF for Image Segmentation", *IEEE Transactions on Neural Networks and Learning Systems*, 2017.
- [7] J. Feng, Y. Wang, and S.F. Chang, "3D Shape Retrieval Using a Single Image Depth Image from Low-Cost Sensors", *Winter Conference on Application of Computer Vision (WACV)*, IEEE. USA, 2016.
- [8] J. Rock, T. Gupta, J. Thorsen, J. Y. Gwak, D. Shin, and D. H., "Completing 3D Object Shape from One Depth Image", *Computer Vision and Pattern Recognition (CVPR)*, IEEE, USA, 2015.
- [9] H. Tian, and Y. Hua, "Depth Estimation with Convolutional Conditional Random Field Network", *Neurocomputing*, Vol. 214, Pp. 546-554, 2016.
- [10] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields", *IEEE T. Pattern Recognition and Machine Intelligent*, 2015.
- [11] M. Liu, M. Salzmann, X. He, "Discrete-Continuous Depth Estimation from a Single Image", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Pp. 716-723, 2014.
- [12] B. Liu, S. Gould, D. Koller, "Single Image Depth Estimation from Predicted Semantic Label", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Pp. 1253-1260, 2010.
- [13] H. Cheng, C. Tseng, C. Hsin and S. Wang, "Single Image 3-D Depth Estimation for Urban Scenes", *IEEE Image Processing (ICP)*, Pp. 2121.2013.
- [14] A. Saxena, S. H. Chung, A.Y. Ng, "3-d Depth Reconstruction from a Single Still Image", *International Journal of Computer Vision*, Vol. 76. Pp. 53-69, 2007.
- [15] G. Zhang and T. T. Wong, "Consistent Depth Maps Recovery from a Video Sequence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31. Pp. 974-988.2009.
- [16] S.H. Raza, O. Javed, A. Das, H. Sawhney, H. Cheng, and I. Essa, "Depth Extraction from Video Using Geometric Context and Occlusion Boundaries". *British Machine Vision Conference (BMVC)*, 2015.
- [17] K. Karcsch, C. Liu, and S. B. Kang, "Depth Extraction from Video Using Non-Parametric Sampling", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 36. Iss. 11. Pp. 2144-2158, 2012.
- [18] Y. Rangaswamy, K.B. Raja, and K.R. Venugopal, "FRDF: Face Recognition using Fusion of DTCWT and FFT Features". *Procedia Computer Science. Elsevier*. Vol. 54. Pp. 809-817.2015.
- [19] D. Zhang, D. Ding. J. Li and Q.Liu, "PCA Based Extracting Feature Using Fast Fourier Transform for Facial Expression

- Recognition”, *Transaction on Engineering Technologies, Springer*. Pp. 413-424.2015.
- [20] G. Aguilar, G. Sanchez, K. Toscano, M. N. Miyatake, and H. P. Meana, “Automatic Fingerprint Recognition System Using Fast Fourier Transform and Gabor Filters”. *Cientifika*. Vol. 12. Pp. 9-16, 2008.
- [21] W. Y. Lee, C.Y. Li, and J.Y Yen, “Integrating Wavelet Transformation with Markov Random Field Analysis for Depth Estimation of Light-Field Images”, *IET Computer Vision*. Vol. 11. Iss. 5. Pp. 358-367.2017.
- [22] R.P. Rangkuti, A.J. Mantau, V. Dewanto, N. Habibie, and W. Jatmiko, “Structured Support Vector Machine Learning of Conditional Random Fields”, *International Conference Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, Pp. 548-555, Indonesia, 2016.
- [23] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC Superpixels”, *EPFL Technical Report 149300*, 2010.
- [24] OOC.Sutton, and A. McCallum, “An Introduction to Conditional Random Fields”, *Foundation and Trend® in Machine Learning*, Vol. 4, Pp. 267-373, 2011.
- [25] N.Shah, V. Kolmogorov, and C. H. Lampert, “A Multi-Plane Block-Coordinate Frank-Wolfe Algorithm for Training Structural SVMs with a Costly Max-Oracle”, *CVR*, 2015