



## Estimasi Citra Kedalaman Dengan Conditional Random Field (CRF) dan Structured Support Vector Machine (SSVM)

Muhammad Rachmadi<sup>a</sup>, Derry Alamsyah<sup>b</sup>

<sup>a</sup>Program Studi Sistem Informasi, STMIK Global Informatika MDP, rachmadi@mdp.ac.id

<sup>b</sup>Program Studi Teknik Informatika, STMIK Global Informatika MDP, derry@mdp.ac.id

### Abstract

*Autonomous UAV is typically require a system which is capable of mapping (segmenting) region of a color image, where the segmentation is a depth image. The depth image can be done by estimate the color image. Estimating color images into depth images is an open challenge in the field of computer vision. The problem of depth image estimation is stochastic based problem. Conditional Random Field (CRF) is a model which is capable of mapping stochastic problems. This study uses CRF as a model of structured classification that is Structured Support Vector Machine (SSVM). In the pre-process, the color image process is converted into superpixel form and using Principal Component Analysis (PCA) as its feature extraction. The results of this study show a good accuracy in the segmentation of the region that is 71.25%.*

*Keywords: Conditional Random Field, Structured Support Vector Machine, Depth Image*

### Abstrak

*Autonomous UAV secara khusus membutuhkan sistem yang mampu memetakan (segmentasi) wilayah dari sebuah citra berwarna, dimana segmentasi berupa citra kedalaman. Penentuan citra kedalaman dapat dilakukan dengan cara mengestimasi citra warna. Estimasi citra warna menjadi citra kedalaman merupakan tantangan terbuka dalam bidang visi komputer. Permasalahan estimasi citra kedalaman merupakan jenis permasalahan stokastik. Conditional Random Field (CRF) merupakan model yang mampu memetakan permasalahan stokastik. Penelitian ini menggunakan CRF sebagai model pada klasifikasi terstruktur Structured Support Vector Machine (SSVM). Pada pra-proses citra warna diubah kedalam bentuk superpixel dan menggunakan Principal Component Analysis (PCA) sebagai ekstraksi fiturnya. Hasil dari penelitian ini menunjukkan hasil akurasi yang baik dalam segmentasi wilayah yaitu 71.25%.*

*Kata kunci: Conditional Random Field, Structured Support Vector Machine, Depth Image*

© 2017 Jurnal RESTI

### 1. Pendahuluan

*Unmanned Aerial Vehicle (UAV)* atau yang lebih dikenal dengan *drone* merupakan pesawat berukuran kecil yang tidak memiliki awak. UAV awalnya digunakan pada bidang militer namun pada perkembangannya UAV kini digunakan sebagai teknologi pendukung pada berbagai bidang lain yaitu: Pencarian dan Penyelamatan Korban Bencana; Agrikultur, seperti pengecekan perkebunan; Sains, seperti mengamati pengunungan dan daerah arkeologi yang sulit dijangkau; *Graphical Information System (GIS)*, seperti membuat pemetaan wilayah; Kegiatan Dokumentasi, seperti perfilman; dan lain sebagainya.

Penggunaan UAV dapat dilakukan di wilayah rendah ( $< 150\text{ m}$ ) dan tinggi ( $> 150\text{ m}$ ) dimana penggunaannya masih bersifat manual atau melibatkan manusia sebagai pilot jarak jauhnya. Selain penggunaan manual, UAV dapat juga diterbangkan secara *autopilot* (tanpa awak/bantua) atau yang disebut sebagai *Autonomous UAV*. Penerbangan *autopilot* pada ketinggian lebih dari  $150\text{ m}$  di area terbuka memiliki resiko halangan yang lebih kecil. Sementara itu, penerbangan *autopilot* pada ketinggian rendah di area terbuka memiliki resiko yang lebih besar. Untuk itu UAV membutuhkan sistem yang mampu menentukan adanya halangan atau tidak, sehingga resiko terjadinya

benturan antara UAV dengan objek lain menjadi lebih kecil.

Halangan bagi sebuah *Autonomous* UAV merupakan sebuah atau beberapa objek yang memiliki jarak yang dekat dengan *Autonomous* UAV itu sendiri. Oleh karena itu pengenalan objek sebagai halangan didasari pada jarak dari objek dengan *Autonomous* UAV. Informasi jarak yang diberikan pada sistem *Autonomous* UAV dapat diperoleh dari sensor ataupun citra, dimana keduanya menghasilkan data berupa matriks pemetaan atau yang disebut sebagai citra kedalaman.

Penggunaan sensor untuk memperoleh informasi jarak dapat dilakukan dengan inframerah seperti pada Kinect ataupun Asus Xtion. Kedua sensor tersebut memberikan informasi jarak dalam bentuk citra kedalaman. Hasil citra kedalaman yang diberikan oleh kedua sensor tersebut memiliki akurasi yang baik [1][2][3]. Akan tetapi, penggunaan inframerah sebagai alat bantu untuk mendapatkan informasi jarak memiliki kendala yaitu pada jarak tembak sinar inframerah itu sendiri. Berdasarkan hal tersebut penggunaan sensor inframerah digunakan dalam kondisi ruang tertutup. Selain itu penggunaan sensor tentunya cukup memberatkan UAV itu sendiri, karena harus dipasangkan dengan UAV. Dengan adanya beban tambahan dari sensor akan memberikan pengaruh ke penggunaan daya pada UAV.

Selain dengan sensor, citra kedalaman dapat dilakukan melalui informasi dari citra warna (RGB). Untuk mendapatkan informasi jarak berupa citra kedalaman dapat dilakukan dengan cara mengklasifikasikan nilai intensitas warna pada masing-masing pixel. Pada dasarnya proses klasifikasi pixel berdasarkan jarak memiliki kerangka kerja yang serupa dengan proses segmentasi citra. Segmentasi citra merupakan sebuah proses yang membutuhkan informasi dari lingkungan (pixel) sekitarnya. Jenis klasifikasi seperti ini merupakan klasifikasi pada proses stokastik. Segmentasi citra yang menggunakan jenis klasifikasi citra ini membutuhkan sebuah model stokastik, diantaranya *Markov Random Field* (MRF) [4] dan *Conditional Random Field* (CRF) [5][6]. Performa yang baik ditunjukkan oleh kedua model tersebut dalam melakukan segmentasi citra.

Berdasarkan kemampuannya dalam segmentasi citra, MRF dapat digunakan pada estimasi citra kedalaman [7][8] dengan performa yang baik, yaitu nilai error yang relatif kecil. Perbedaan performa pada [7] dan [8] yaitu error yang lebih kecil dimiliki oleh [8] dipengaruhi oleh penggunaan fitur yang berbeda. Pada [7] menggunakan *visual clues* sebagai fitur, sedangkan [8] menggunakan *superpixel*. Selanjutnya, untuk penggunaan CRF dalam estimasi memiliki hasil yang lebih baik dari penggunaan MRF [9][10]. Fitur *superpixel* kembali digunakan pada [9], sedangkan [10]

lebih memilih menggunakan *raw data*. Hasil yang diberikan pada [9] dan [10] memiliki perbedaan error relatif yang hampir sama dengan [10] lebih unggul dari [9].

Pada [9][10] CRF mampu memberikan performa yang baik dalam estimasi citra kedalaman, dimana keduanya memanfaatkan *Convolutional Neural Network* (CNN) untuk mengklasifikasi. Sementara itu, CRF dapat juga menjadi model pada klasifikasi terstruktur *Structured Support Vector Machine* (SSVM) [11]. Penggunaan klasifikasi terstruktur ditujukan untuk memudahkan proses pembentukan model pada estimasi citra kedalaman. Selanjutnya efisiensi pada fitur dapat dilakukan dengan menggunakan *Principal Component Analysis* (PCA) [12]. Oleh karena itu pada penelitian ini digunakan *Superpixel* dan PCA sebagai fitur untuk berikutnya menggunakan CRF dan SSVM dalam estimasi citra kedalaman.

## 2. Tinjauan Pustaka

### 2.1. Superpixel SLIC

SLIC merupakan metode *gradient ascent* untuk mengenerasi *superpixel* dari pusatnya yang diinisiasi berdasarkan warna dan kedekatan. Secara khusus, SLIC menggunakan algoritma pengelompokan k-means pada data berdimensi 5. Dimensi pada data terdiri dari warna dan posisi pixel [13].

Pusat dari *superpixel* diinisiasi pada *regular grid* dengan *step R*, dimana

$$R = \left\lceil \frac{WH}{K} \right\rceil \quad (1)$$

Selanjutnya setiap *pixel*  $x_n$  dikelompokkan berdasarkan kedekatan dengan *superpixel* terdekat, dengan jarak  $d$  yaitu:

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

Dengan  $\beta$  merupakan parameter kepadatan. Prosedur ini terus diperbaharui hingga nilai dari  $d$  konvergen.

### 2.2. Principal Component Analysis (PCA)

Langkah-langkah ekstraksi fitur dengan PCA [12]:

#### 1. Adjusted Data:

Diberikan data  $x \in R^{m \times d}$  dengan  $d$  adalah dimensi data dan  $m$  banyaknya data, maka *Adjusted Data*  $x_{adj}$  adalah

$$x_{adj} = x - \bar{x} \quad (3)$$

dimana  $\bar{x}$  adalah rerata dari  $x$ .

## 2. Covariance Matrix

$$C = \begin{bmatrix} cov(x_0, x_0) & \dots & cov(x_0, x_n) \\ \vdots & & \vdots \\ cov(x_m, x_0) & \dots & cov(x_m, x_n) \end{bmatrix} \quad (4)$$

dimana  $x_i \subset x$  dengan  $i = 0, 1, \dots, m$  dan  $m = n = d$ .

## 3. Eigenvalues and Eigenvector of Covariance Matrix

$$Cv = \lambda v \quad (5)$$

Dengan  $v = \{v_0, v_1, \dots, v_m\}$  adalah *eigenvector* dan  $\lambda = \{\lambda_0, \lambda_1, \dots, \lambda_m\}$  adalah *eigenvalue*.

## 4. Best Component (Dimensional Reduction)

Pemilihan beberapa nilai eigen (*eigenvalues*) terbaik sebagai fitur, yaitu

$$feature = \{\lambda_0, \lambda_1, \dots, \lambda_{nb}\}$$

dengan  $nb$  adalah Batasan jumlah *eigenvalues* terbaik.

## 5. New Data

Data baru diperoleh yaitu

$$x' = feature \times x_{adj} \quad (6)$$

Data ini juga dapat berlaku sebagai fitur.

## 2.3. Conditional Random Field (CRF)

CRF merupakan sebuah model stokastik yang dibentuk dengan mempertimbangkan nilai sebelumnya (tetangga). Pembentukan model CRF dilakukan dengan menentukan nilai bobot terbaik  $\theta_k$  pada setiap node (kejadian)  $y_t$ . Bobot tersebut menentukan seberapa besar pengaruh dari pasangan kejadian  $f_k(y_t, y_{t-1}, x_t)$  dengan  $x_t$  sebagai fitur/data observasi terkait. Nilai  $f_k(y_t, y_{t-1}, x_t)$  berada pada rentang  $[0, 1] \in \mathbb{Z}$ . Dimana 1 untuk yang terjadi dan 0 untuk tidak terjadi. Selengkapnya model CRF ditunjukkan oleh persamaan (7) dan (8).

Model CRF [14]:

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

Normalisasi Faktor:

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

## 2.3. Structured Support Vector Machine (SSVM)

SSVM merupakan metode klasifikasi yang menggunakan model (pembobotan) untuk melakukan klasifikasi secara linier [11][15]: SSVM dibagi menjadi dua bagian, yaitu:

Persamaan Klasifikasi (Classifier):

$$f_{ssvm}(x) = \operatorname{argmax}_y \langle w, \phi(x, y) \rangle \quad (9)$$

Pembentukan Model (Training Phase):

Diberikan data (himpunan pasangan masukan-luaran/kejadian)  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$  kemudian mengestimasi nilai  $w$  dengan penyelesaian persamaan optimasi:

$$\min_{w, \xi} \frac{\lambda}{2} \|w\|^2 + \frac{1}{n} \sum_{i=1}^n \xi_i \quad (10)$$

dengan syarat,

$$\langle w, \psi_i(y) \rangle \geq L(y_i, y) - \xi_i \quad \forall i, \forall y \in \overbrace{\mathcal{Y}(x_i)}^{=: y_i} \quad (11)$$

dimana

$$\psi_i(y) := \phi(x_i, y_i) - \phi(x_i, y) \quad (12)$$

## 3. Metodologi Penelitian

### 3.1. Pengumpulan Data

Data data yang digunakan adalah Make 3D *dataset* (Cornell Dataset). Data berupa citra warna citra kedalamannya, seperti yang ditunjukkan pada Gambar 1 dan 2. Penggunaan kedua citra ini didasari pada sistem klasifikasi yang bersifat *supevised learning*. Oleh karena itu dibutuhkan citra *ground truth* yang merupakan citra jarak (yang telah diberikan pada Make 3D *dataset*).

Jumlah data yang digunakan pada penelitian ini berjumlah masing-masing 90 citra untuk warna (Gambar 1) dan kedalamannya (Gambar 2). Selanjutnya, data dibagi kedalam dua jenis yaitu data latih dan data uji.

### 3.2. Rancangan Sistem

Pada penelitian sistem yang dibuat mampu memberikan hasil berupa citra kedalaman dengan citra warna sebagai citra masukan. Untuk itu diperlukan 2 proses yaitu proses latih (*Training Phase*) dan proses uji (*Testing Phase*) pada gambar 3. Proses latih menghasilkan model yang siap digunakan untuk memprediksi citra warna masukan menjadi citra kedalaman sebagai luarnya. Proses latih terdiri dari:

*Pra-proses (Segmentasi):* Data berupa citra warna ( $I_{RGB}$ ) dan citra kedalaman ( $I_{depth}$ ) masing-masing diubah kedalam format superpixel ( $I'_{RGB}$ ,  $I'_{depth}$ ). Gambar 1.

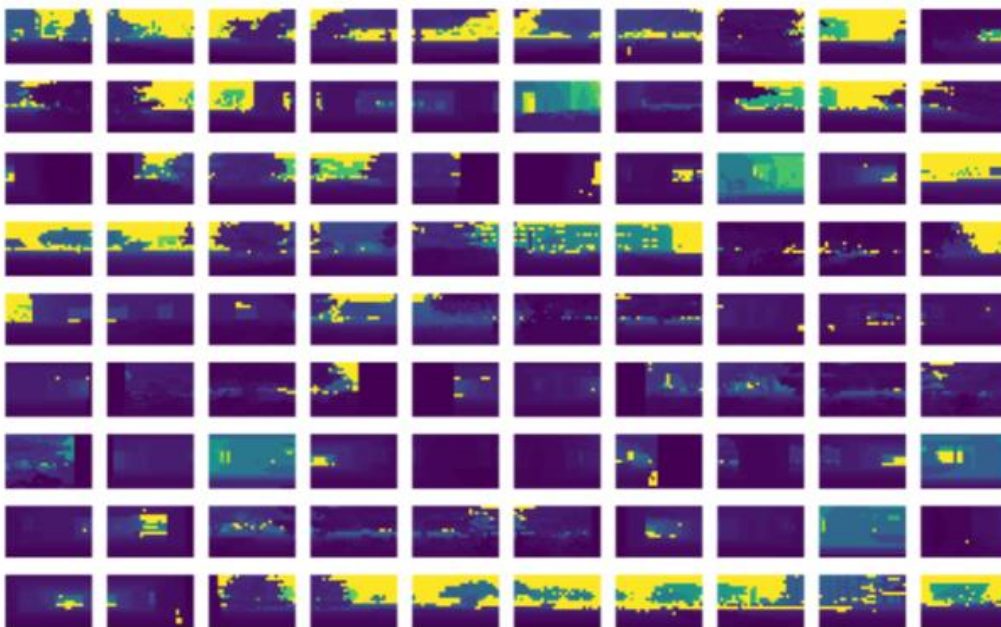
*Ekstraksi Ciri:* Untuk setiap warna pada  $I'_{RGB}$  disatukan kedalam satu vektor ciri  $I_f \in R^3$ . Dengan menggunakan PCA selanjutnya vektor ciri direduksi untuk mendapatkan dua ciri penting yaitu  $f_{pca}(i_f) = i'_f \in R^2$  dimana  $i_f \in I_f$ .

*Data Latih:* Untuk setiap ciri  $i_f$  dipasangkan dengan setiap kelas kedalaman yang didapat dari rerata setiap kelompok superpixel  $\mu'_{depth}$ . Gambar 2.

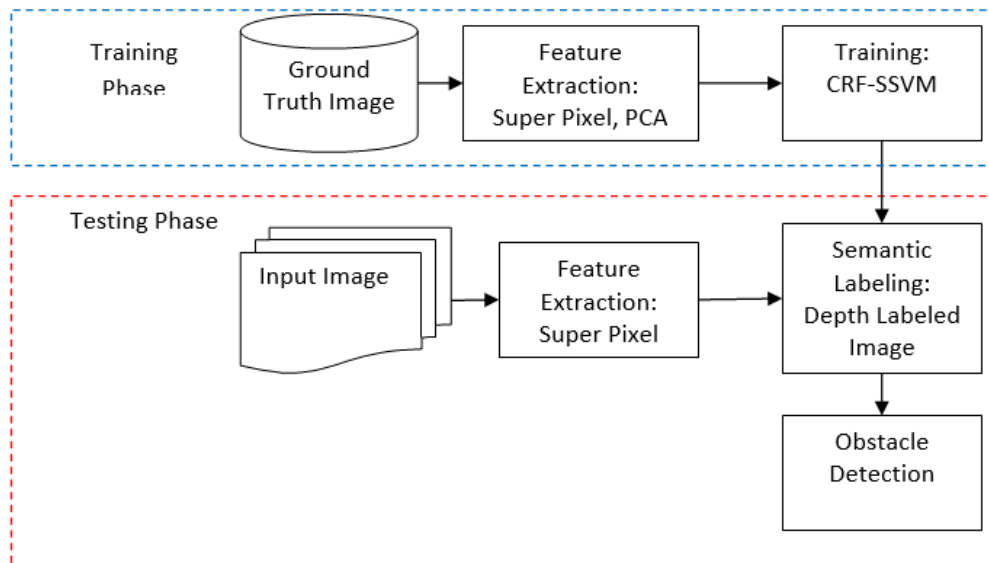
*Pemodelan:* model untuk klasifikasi pada sisteam menggunakan model SSVM dimana setiap bobotnya diinisialisasi berdasarkan model CRF  $w = p(x|y)$  dengan  $x = i'_f$  dan  $y$  sebagai kelas segmentasi. Rancangan sistem terlihat pada Gambar 3.



Gambar 1. Citra Warna



Gambar 2. Citra Kedalaman



Gambar 3. Rancangan Sistem

Selanjutnya sistem menggunakan proses uji untuk mengestimasi citra kedalaman, yaitu melalui:

*Pra-proses (Segmentasi)*: Mengubah format citra warna masukan ( $I_{RGB}$ ) kedalam format superpixel sehingga didapat ( $I'_{RGB}$ ).

*Ekstraksi Ciri*: Didapatkan melalui  $f_{pca}(i_f) = i'_f$  dari data masukan.

*Klasifikasi setiap segmen*: Dengan menggunakan model CRF sebagai penentu bobot dan model SSVM sebagai pengklasifikasi (*Classifier*) maka didapat data baru berupa segmentasi untuk melabelkan wilayah pada  $i'_f$  dengan nilai kedalaman.

### 3.3. Evaluasi

Evaluasi pada sistem dilakukan dengan menghitung nilai akurasi yaitu:

$$acc = \frac{\sum_{i=1}^n 1(y=\hat{y})}{n} \quad (11)$$

## 4. Hasil dan Pembahasan

Data pada penelitian ini terdiri dari 90 citra masing-masing untuk citra warna dan citra kedalaman. Citra kedalaman merupakan citra *ground truth* yang digunakan sebagai kelas label. Pembagian data dilakukan dengan membagi masing-masing citra warna dan kedalaman menjadi 70 citra latih dan 20 citra Uji.

Citra warna dan citra kedalaman dibagi menjadi  $\pm 20$  wilayah *superpixel*. Selanjutnya untuk citra kedalaman data label didapat dari rerata tiap segmen wilayah hasil *superpixel*. Sedangkan ekstraksi fitur didapat dari komponen utama (nilai eigen) dari PCA.

Pembentukan model SSVM terdiri dari nilai bobot dan bias, dimana untuk bobot nilai inisiasi didapat dari model CRF. Pemetaan wilayah segmentasi dilakukan secara berurut untuk setiap wilayah superpixel, yang disebut sebagai proses stokastik. Wilayah pada penelitian ini dibagi kedalam tiga wilayah: dekat, sedang, jauh (Gambar 1). Hasil dari penelitian ini menunjukkan hasil yang cukup bagus yaitu, rerata akurasi tiap gambar uji menunjukkan hasil yang cukup baik yaitu senilai 71.25%, seperti yang ditunjukkan pada Tabel 1. Performa dari CRF dan SSVM dipengaruhi oleh nilai ciri yang diberikan. Penggunaan nilai eigen memiliki tingkat kemiripan untuk fitur warna. Hal ini mengakibatkan tingkat akurasi pada data latih kecil, yaitu sebesar 66.85% dan diikuti dengan akurasi data uji sebesar 71.25%.

Tabel 1. Hasil Penelitian

Metode	Akurasi	Data Latih (%)	Data Uji (%)
<b>SSVM</b>			
PCA (3-components): Rerata		66.2	71.0
PCA (2-components): Rerata		66.1	71.0
PCA (3-components): Eigen		67.5	71.5
PCA (2-components): Eigen		67.6	71.5
Rerata Akurasi		66.85	71.25
<b>Structured Peceptron:</b>			
PCA (3-components): Rerata		43.7	48.9
PCA (2-components): Rerata		43.7	48.9
PCA (3-components): Eigen		25.4	26.1
PCA (2-components): Eigen		67.5	71.5
Rerata Akurasi		45.08	48.85

## 5. Kesimpulan

### 5.1 Simpulan

Dengan menggunakan ekstraksi fitur PCA dan *Supapixel* sebagai pra-proses, maka model CRF dan klasifikasi SSVM mampu memberikan tingkat akurasi estimasi citra kedalaman sebesar 71.25% pada data uji. Sementara itu, untuk data latih tingkat akurasi yang diberikan adalah sebesar 66.85%.

### 5.2 Saran

Pada penelitian selanjutnya dapat menggunakan fitur berbasis tekstur dan frekuensi sebagai alat bantu untuk meningkatkan akurasi estimasi citra kedalaman.

## 6. Daftar Rujukan

- [1] T. Hamedani & A. Harati. 2014. Multi Scale CRF Based RGB-D Image Segmentation Using Inter Frames Potentials. International Conference on Robotics and Mechatronics (ICRoM). IEEE. Hal. 920-925.
- [2] S.S. Mirkamali & P. Nagabhushan. 2015. RGBD Image Segmentation. Iranian Conference on Machine Vision and Image Processing (MVIP). IEEE. Iran. Hal. 41-44.
- [3] A.C. Mueller & S. Behnke. 2014. Learning Depth-Sensitive Conditional Random Fields for Semantics Segmentation of RGB-D Images. International Robotics and Automation (ICRA). IEEE. Hong kong.
- [4] W. Zhang & M. Li. 2014. MRF and CRF Based Image Denoising and Segmentation. International Conference Digital Home (ICDH). IEEE. China. Hal. 128-131.
- [5] Y. Qi , G. Zhang, Y. Qali, & Y. Li. 2016. Object Segmentation Based on Gaussian Mixture Model and Conditional Random Fields. Information and Automation (ICIA). IEEE. China. Hal. 900-904
- [6] F. Liu, G. Lin, R. Qiao, & C. Shen. 2017. Structured Learning of Tree Potentials in CRF for Image Segmentation. IEEE Transactions on Neural Networks and Learning Systems.
- [7] A. Saxena, S. H. Chung, A.Y. Ng. 2007. 3-d Depth Reconstruction from a Single Still Image. International Journal of Computer Vision. Vol. 76. Hal. 53-69
- [8] B. Liu, S. Gould, D. Koller. 2010. Single Image Depth Estimation from Predicted Semantic Label. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hal. 1253-1260.
- [9] M. Liu, M. Salzmann, X. He. 2014. Discrete-Continuous Depth Estimation from a Single Image. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hal. 716-723
- [10] H. Tian, & Y. Hua. 2016. Depth Estimation with Convolutional Conditional Random Field Network. Neurocomputing. Vol. 214. Hal. 546-554.
- [11] R.P. Rangkuti, A.J. Mantau, V. Dewanto, N. Habibie, & W. Jatmiko. 2016. Structured Support Vector Machine Learning of Conditional Random Fields. International Conference Advanced Computer Science and Information Systems (ICACSIS). IEEE. Indonesia. Hal. 548-555.
- [12] Ketelaere, B.D, Mia Hubert, & Eric Schmitt. Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional Data. Journal of Quality Technology. Vol 47. No. 4. Hal 318. 2015
- [13] Achanta, R., Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine S'usstrunk, SLIC Superpixels. EPFL Technical Report 149300. 2010.
- [14] Sutton, C. & Andrew McCallum. An Introduction to Conditional Rancdom Fields. Foundation and Trend® in Machine Learning. Vol. 4. Hal. 267-373. 2011
- [15] Shah, N., Vladimir Kolmogorov, Christoph H. Lampert. A Multi-Plane Block-Coordinate Frank-Wolfe Algorithm for Training Structural SVMs with a Costly Max-Oracle. CVR. 2015