Volume 11 Issue 2 June 2018 ISSN 2088-7051

Jurnal Ilmu Komputer dan Informasi

Journal of Computer Science and Information

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 1-9
DOI: http://dx.doi.org/10.21609/jiki.v11i1.562

SECURING COMMUNICATION IN THE IOT-BASED HEALTH CARE SYSTEMS

Bayu Anggorojati[1] and Ramjee Prasad[2]

[1]Faculty of Computer Science, Universitas Indonesia, Kampus UI, Depok, 16424, Indonesia
[2]Future Technologies for Business Ecosystem Innovation (FT4BI), Aarhus University, Nordre Ringgade 1, 8000 Aarhus, Denmark

E-mail: ramjee@btech.au.dk

Abstract

Rapid development of Internet of Things (IoT) and its whole ecosystems are opening a lot of opportunities that can improve humans' quality of life in many aspects. One of the promising area where IoT can enhance our life is in the health care sector. However, security and privacy becomes the main concern in the electronic Health (eHealth) systems and it becomes more challenging with the integration of IoT. Furthermore, most of the IoT-based health care system architecture is designed to be cross-organizational due to many different stakeholders in its overall ecosystems – thus increasing the security complexity. There are several aspects of security in the IoT-based health care system, among them are key management, authentication and encryption/decryption to ensure secure communication and access to health sensing information. This paper introduces a key management method that includes mutual authentication and secret key agreement to establish secure communication between any IoT health device with any entity from different organization or domain through Identity-Based Cryptography (IBC).

Keywords: *IoT, eHealth, Security, Authentication, ID-based cryptography*

Abstrak

Perkembangan Internet yang cepat dari Things (IoT) dan keseluruhan ekosistemnya membuka banyak kesempatan yang dapat memperbaiki kualitas hidup manusia dalam banyak aspek. Salah satu area yang menjanjikan dimana IoT dapat meningkatkan kehidupan kita ada di sektor perawatan kesehatan. Namun, keamanan dan privasi menjadi perhatian utama dalam sistem Kesehatan Elektronik (eHealth) dan ini menjadi lebih menantang dengan integrasi IoT. Lebih jauh lagi, sebagian besar arsitektur sistem perawatan kesehatan berbasis IoT dirancang untuk menjadi lintas organisasi karena banyak pemangku kepentingan yang berbeda dalam keseluruhan ekosistemnya - sehingga meningkatkan kompleksitas keamanan. Ada beberapa aspek keamanan dalam sistem perawatan kesehatan berbasis IoT, di antaranya adalah manajemen kunci, otentikasi dan enkripsi / dekripsi untuk memastikan komunikasi yang aman dan akses terhadap informasi penginderaan jauh. Makalah ini memperkenalkan metode manajemen kunci yang mencakup saling otentikasi dan kesepakatan kunci rahasia untuk membangun komunikasi yang aman antara perangkat kesehatan IoT dengan entitas dari berbagai organisasi atau domain melalui Identity-Based Cryptography (IBC).

Kata Kunci: *IoT, eHealth, Keamanan, Autentikasi, Kriptografi berbasis ID*

1. Introduction

The convergence of IT and medical world – also known as eHealth – have been transforming the way health care services are delivered. eHealth offers a new means for utilizing health resources, such as information, money, medications, etc, and then help all the relevant stakeholders to utilize those resources more efficiently [1]. For a country that has high population where some of them are living in remote areas, such as Indonesia that consists of thousands of islands, delivering health care services is a big issue especially when the specialist doctors are not well distributed throughout the country [2], [3]. In such situation, tele monitoring of patients' health status by using electronic medical devices that is able to communicate remotely through Internet with the advancement of IoT is a promising solution.

Various IoT health care services and applications have been introduced, such as ECG! (ECG!), glucose level, and blood pressure monitoring, medication management, and a lot more health applications for smart phones, as reported in [4]. Furthermore, many well known companies are developing more products and services within the IoT for health care solutions [4]. It was also reported by McKinsey Global Institute in [5] that the IoT-based health care applications are projected to create about $1.1 - $2.5 trillion in

growth annually by the global economy by 2025 and form the biggest economic impact compared to the IoT applications in other areas. It shows that the IoT in health care has a very bright future, both in terms of benefits for people, technology and economy.

With all those encouraging facts about IoT-based health care solution, there lies big concern about security and privacy. There are many security and privacy challenges pertaining IoT-based health care system, such as physical attack and device vulnerabilities, security in the communication channel and ecosystem (e.g. mutual authentication, key management and cryptographic support), attack on the stored information, etc [1], [6]. On the other hand, an IoT device that transmit patient's health information needs to comply with Health Insurance Portability and Accountability Act (HIPAA). The IoT devices are also considered as having limited power, computation and memory capability which imply that the security mechanism needs to utilize the device's resources efficiently. Furthermore, the architecture of IoT-based health care system in general involves several stakeholders that belong to different organizations with different security domain and policy, which is adding more complexity to the security task.

Based on the circumstances mentioned earlier, it is important to provide key management that supports mutual authentication and secure data transmission between two entities within the IoT-based health care system that belong to different organizations or domains. This paper presents a security scheme based on IBC that supports all the capabilities stated before. The IBC-based scheme is chosen because it is essentially a asymmetric key scheme, which is easier in key distribution and more scalable than the symmetric ones, while it requires no certificate in the practical key distribution like the other asymmetric key schemes, e.g. Rivest, Shamir, and Adelman (RSA) and Elliptic Curve Cryptography (ECC). The scheme provides mutual authentication and key agreement for secure communication between entities across different organizations or domains, and is developed based on variant of Identity-Based Encryption (IBE) that removes the key escrow problem in original IBC which was introduced by Zhaohui Cheng et al. in 2004 [7].

The rest of this paper is structured as follows: Security and privacy challenges especially in the context of IoT-based health care system are reviewed in Section II. The IoT-based health care system architecture that is referred in this work is presented in Section III. The proposed key management, authentication and key agreement scheme is explained in Sec- tion IV. The security and efficiency analysis of the proposed scheme is discussed in Section V. Finally the conclusion and some future works are given in section VI.

2. Methods

Security Challenges and Possible Solutions

This section reviews some of the security challenges on the IoT-based health care system. The challenges consists of two main categories: challenges concerning the inherent nature of IoT which impact the security solution and security challenges related to the IoT system, especially in health care area. Further, some possible solutions of the reviewed challenges are also presented based on some related works.

IoT health devices are embedded with low-speed processors. The central processing unit (CPU) in such devices is not very powerful in terms of its speed. In addition, these devices are not designed to perform computationally expensive operations. That is, they simply act as a sensor or actuator. Therefore, finding a security solution that minimizes resource consumption and thus maximizes security performance is a challenging task [4]. On the other hand, the number of IoT devices has increased gradually, and therefore more devices are getting connected to the global information network. thus, designing a highly scalable security scheme without compro- mising security requirements is another challenge [4].

Medical data contain very sensitive information about patient's health status that must be kept secure and private from any unauthorized people. Hence, hospitals and health care providers are obligated to exchange patients private information securely to comply with HIPAA. With the ubiquitous and pervasive nature of IoT, security breaches and privacy violations are highly possible if the automatic data collection is not verified and managed properly. Patients' sensitive personal and medical information could be a tampered, used or compromised in the absence of having real time monitoring. This will not only cause a threat to infrastructure but has a catastrophic impact of peoples lives. Malicious users could hijack applications and wearable devices taking control of peoples private information and introduce a devastating health and security risks [6].

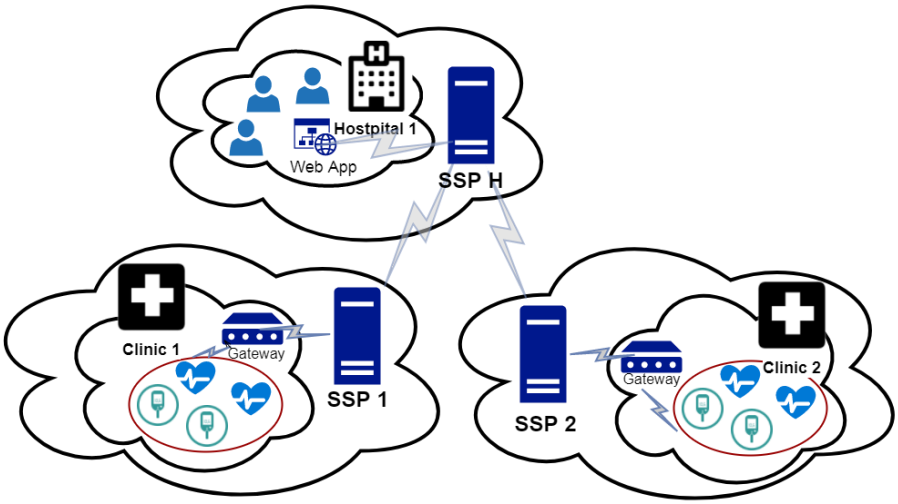


Figure 1. Reference architecture of IoT-based health care across domains

The other challenge is managing credentials and controlling access to applications and patients confidential information. For instance, medical care givers are allowed to access devices in response to patients sensor devices request but the internet connection used may be a public or insecure Wi-Fi network that can be easily tampered to conduct man-in-the-middle attacks. Many authentication techniques could be implemented so that patients are capable to verify and allow medical doctors to access their internally embedded devices. but all of the sudden they lost conscious and they are still desperate to get doctors assistance and guidance. Some IoT healthcare manufacturing companies provide a permanent hard coded password to be used while accessing IoT devices which, the passwords are publicly available in the device manual and would be used to misconfigure the device that introduce risk to patient life [6].

Another challenge is implementing and deploying cryptographic protocols in IoT health cloud correctly. Managing cryptographic keys is crucial but strenuous due to IoT pervasive and continuous capabilities. IoT ecosystem demands the use of concurrent authentication operation with quick real time respose [6].

HIPAA regulation related to Transmission Security Encryption 164.312(e)(2)(ii) mention that entities should implement a mechanism to encrypt and decrypt patient's health information whenever deemed appropriate. Entities shall prepare documentation of the encryption technology that is implemented including policies and procedures, how cryptographic key management are exchanged and restricting access to create and alter cryptographic keys. Furthermore, hardening the con- fidential processes such as managing and sharing keys is also should be audited and enforced [6].

Several security schemes that attempted to solve issues related to the IoT in various applications have been proposed, in which the approaches can also be applied in the health care area. The issue of secure transmission in IoT as required in HIPAA regulation is strongly connected with cryptographic protocols given the fact that IoT consists of constrained-devices (e.g. low computation, memory and power) and big numbers of IoT nodes that has scalability implications. Furthermore, the used for encryption in secure communication needs proper management which is also related to the authentication.

Initially, symmetric key cryptography based scheme was extensively researched due to its small key size that fits the requirement of constrained devices. However, it suffers from major drawback in scalability. In order to address scalability issue, some attempts to introduce Public Key Cryptography (PKC) based scheme in constrained devices have also been made. It have been shown that it is computationally feasible to implement PKC in constrained device, especially by using ECC which require shorter key size compared to RSA based PKC. An example of Public Key Infrastructure (PKI) based encryption implementation in IoT m-health devices was pre- sented in [8].

Yet, traditional PKC requires certificate which consumes bigger memory size and complex to manage. To overcome this shortcoming, a certificateless PKC scheme, known as IBC [9], [10], has been proposed. The basic idea of original IBE is, first there is a central entity called Private Key Generator (PKG) which is responsible to generate some public parameters and a master key that is

kept secret. That master key is then used to generate private keys for all other parties, who trust that particular PKG, given their IDs. Now, the encryption-decryption process can be done in the same manner as the traditional PKC with an exception that the public key can be generated by any entity using a known ID. Due to its benefits, e.g. certicateless and low resource requirements, some IBC based security schemes have also been applied for constrained device, such as Mobile Ad-hoc Networks (MANET) [11]. In IBC, any arbitrary string, such as the identity of participating party in communication, can be used as public key, thus replacing the role of certificate in traditional PKC.

Several IBC schemes have also been proposed for IoT. A key establishment scheme between two communicating entities in which one of them is constrained device and with the help of a proxy is proposed in [12]. An IBC protocol design pattern for Machine-to-Machine (M2M) was proposed in [13]. Federated end-to-end authentication for constrained IoT using IBC and ECC has been suggested by Markmann et al. [14]. Finally, an IBC based authentication scheme for M2M has also been proposed by Shuo Chen et al. [15].

Security Challenges and Possible Solutions

This section reviews possible architecture of IoT-based health care system from several references. Based on the reviews, a reference architecture will be chosen for this work. Furthermore, the chosen reference architecture will be used as a use case scenario for building our proposed security scheme.

According to Gabriel Neagu et al. [16], the sensing services delivered by IoT in general is based on interaction of four entities: sensor owners (SO), sensor publishers (SP), extended service providers (ESP), and sensor data consumers (SDC). The SO might be a private or a public organization, a commercial sensor provider or an individual. In case the SO decides that the data provided by these sensors will be available in the cloud it has to define the access policy to these data that potential SP should implement and potential users should comply with. When a SDC (e.g. government, business organization, academic institution, scientific research community or individual) is interested in accessing data provided by a published sensor, the SP mediates a service agreement between this SDC and respective SO where the SO responsibilities regarding sensing data availability and quality for the requested period of time, as well as their compliance with existing standards are detailed.

Another entity called Sensing Service Provider (SSP) was also introduced in [16] which simplifies the interaction between SDC and other entities in its interests, including SO (for data availability, compliance and quality), SP (for access services to sensor data) and other extended service providers (for value added services). In health care specific scenario, SDC could be a medical institution (e.g. hospital, clinics, etc) or any health care service provider. Depending on available financial, human and technical resources, those SDC in health care may decide to implement IoT-based health care service either as an extension of their existing IT infrastructure or by outsourcing it to specialized providers. In most cases, the second option would be more preferable especially by small medical institutions (e.g. clinics, general practitioners). Finally, the authors in [16] proposed that the SSP is a major actor who interacts with other stakeholders: SO, SP and health care provider as SDC.

On the practical perspective, there can be many different ways of architecture design and deployment model of IoT-based health care system. According to [3], one of the ongoing telehealth pilot project in Indonesia, called tele-ECG!, is carried out in such a way that a well known cardiac hospital becomes the center of the project and it is serving other remotely located health care providers. Remote health care providers that lack of cardiologists may send the ECG! data of their patients to the cardiologists that belong to the referred hospital through tele-ECG! to get their diagnosis pertaining the cardiac issues of the patients' in the remote area.

Combining the proposed model in [16] and the scenario presented in [3], a reference architectural model that will be used to develop our security use case is proposed in Figure 1. The model shown in Figure 1 is a simplified version of the model from [16], in which SP is assumed to be the SSP itself and the SO is part of the health care providers (Clinic 1 and 2), while Hospital 1 is the SDC. It is also assumed that each of the health care provider outsourced the IoT-based health care system to a specialized provider (e.g. SSP 1, SSP2 and SSP H).

Concerning the proposed IBC security scheme which requires PKG, the reference architecture in Figure 1 will have three PKG for each SSP domain, e.g. P KGSSP 1 ,P KGSSP 2 and P KGSSP H. For security reason, PKG is only accessible by entities within its domain. As PKGs, they generate master secret keys and public parameters for each domain. Additionally, P KGSSP 1 and P KGSSP 1 also generates identities and corresponding private keys for all devices in each domain, including S1/S2, device gateways and medical sensors. Besides, the P KGSSP H

TABLE 1.
SUMMARY OF ALL ALGORITHMS IN IBE WITHOUT KEY ESCROW

Algorithm	Input	Output
Setup	1^K: a security parameter	*s*: system's master-key (private) params: system's public parame- ters
Extract	ID: Identity *s* and params	QID: public key dID: private key
Publish	params	tID: sub-private key NID: sub-public key
Encrypt	*m*: plaintext ID, params, NID	*C*: ciphertext
Decrypt	*C*: ciphertext $dID, t,$ params	m': plaintext

generate private keys for all registered users in Hospital 1. It is important to note that the user identity in hospital domain are created in registration process and then the private key is generated accordingly by the P KGSSP H .

Proposed Method

IBE Scheme without Key Escrow

As earlier mentioned, the proposed scheme is developed based on a variant of IBE that is key-escrow free which is adapted from [7]. In the original IBE scheme by Boneh Franklin [10], there are four randomized algorithms involved, namely Setup, Extract, Encrypt, and Decrypt, while another algorithm called Publish is included in the IBE's variant without key escrow. A summary of inputs and outputs for all five algorithms is listed in Table 1, while the detail procedures can be reviewed in [7].

Please note that the Setup algorithm occur fully in the PKG which could happen, for instance in the system initialization. In the Extract algorithm, the PKG receives an input ID from a communicating entity, then after the algorithm is executed in the PKG, QID is published in a publicly available directory while dID is sent to the communicating entity secretly. Finally, the rest of the algorithms (i.e. Publish, Encrypt and Decrypt) happen in the communicating entity, except that NID as one of the results of Publish algorithm is published in a public directory.

Before explaining the other mechanisms in the proposed schemes, i.e. system and device initialization as well as authentication with key agreement, the definitions of the notations used in the proposed scheme is defined in Table 2.

System and Device Initialization

System initialization refers to the process related to IBE when PKG of an SSP! (SSP!) is started, while device initialization refers to the process followed when gateway and constrained device join the SSP!. During system initialization, the most important operation is generating params and master-key, then making params publicly available as explained in the previous section. In addition to that, it is important for the PKG of SSP! to have an identifier that is recognized by everybody(thing) through the Internet. Therefore, we propose the domain name as the primary identity representing the IoT Service Provider (IoTSP) and then the device identity will be appended with this domain name. Having such identifier scheme is beneficial in the lookup process even though the communicating entity is located in different domain.

With regards to the device initialization, there are two important mechanisms need to be performed, i.e. generation and distribution of device identifier and associated private key of the device by the PKG, and then the generation of sub-public and sub-private key pair by the device itself. In principle, distribution of device's identifier and the corresponding private key by PKG can either be done offline and online. Offline method requires configuration of identifier and corresponding private key statically during the flashing time of the device, while online method can be done more dynamically. In this case, online method is chosen and a secure way of delivering device's private key is proposed.

The proposed online device initialization is secured by two symmetric keys, namely KInitReq and KInitRsp, which are one-time randomly generated, i.e. they will be destroyed after device initialization. There can be several ways in obtaining those keys. One practical way is by performing a device registration through web interface. After the registration process, unique device identifier, KInitReq and KInitRsp will be gen- erated for and transfered to the registered device (e.g. they can be loaded to the device by cable data after downloading from PKG). The reason why unique device identifier is generated at this point because it is possible to include more human friendly name into it, such as type of device (gateway, ECG, diabetic sensor, etc) and location of the device (hospital 1 or house 1, etc). Afterwards, the device can request its identifier and corresponding private key securely using Authenticated Encryption with Associated Data (AEAD) [17]. The reason of choosing AEAD being that it is more secure to properly authen-

TABLE 2.
DEFINITION OF USED NOTATIONS

Notation	Definition
s	Master secret key
paramsx	Public system parameter of domain x
IDi	Identity of entity i
Qi	Public key of corresponding entity i
di	Private key of corresponding entity i
Ni	Sub-public key of corresponding entity i
ti	Sub-private key of corresponding entity i
Pm	Plaintext from a message m or a result of decryption
Cm	Ciphertext, a result of encrypting message m
E(k, N, P, A)	AEAD encryption of plaintext P, using key k, nonce N and associated data A
D(k, N, C, A)	AEAD decryption of ciphertext C, using key k, nonce N and associated data A
Eij(m)	ID based encryption of message m using Qj, Nj, and ti
Dij (m)	ID based decryption of message m using Qj, Nj, and ti
Sm	Digest of message m as a result of Message Authentication Code (MAC)

ticate the ciphertext than having simply the encryption, while it works faster than secure implementation of Hash-based Message Authentication Code (HMAC) that requires two keys for encryption and authentication. The detail protocol of secure device initialization is shown in Fig. 2.

Authentication Mechanism with Key Agreement

Fig. 3 illustrates a scenario that using the proposed authentication mechanism with key agreement. In this scenario, a mobile application's user A wants to access sensor B that belongs to an IoTSP domain. For simplicity, user A and sensor B will be referred as A and B respectively from this point onwards. Moreover, A has to go through IoT Server (IoTS) as the entry point to B. It is assumed that the activity of A in this scenario is done by the mobile app (either the mobile app itself or the server that provides API to mobile app), hence it is shown as one entity in Fig. 3. It can also be assumed that practically the entity in each domain (initially) does not have knowledge of system parameters and sub-public key of entities in other domains, therefore a lookup function needs to take place before encryption is performed. Detail of authentication mechanism is explained as follows: (1) First of all, A performs lookup in order to obtain paramsIoT SP and NIoT S by using IDIoT S as input. After successful lookup, it generates QIoT S = H1IoT SP (IDIoT S), where H1IoT SP is included in paramsIoT SP. Note that other paramters in paramsIoT SP are also used for encryption. After-wards, C1 is created by encrypting IDA, IDB, and timestamp T using QIoT S, NIoT S and tA as keys. Here T is used to prevent replay attack. Then, IDA, IDIoT S, and C1 are sent to IoTS; (2) After receiving message from A, IoTS will perform lookup based the received IDA, to obtain paramsM A and NA. After successful lookup, it decrypts C1 using dIoT S, tIoT S and NA to obtain IDA, IDB, and T. After that, T is validated, and IDA is also verified if it similar with the received one. If they are valid the process is continued else, it stops and sends error message to A. After successful validation, a message that contains NB is encrypted as C2 using QA, NA and tIoT S, then C2 is sent to A. Another message that contains paramsM A and NA is encrypted as C3 by using QB, NB and tIoT S and then it is sent to B, informing that A wants to access it; (3) Upon receiving C2, it is then decrypted by A using dA, tA and NIoT S to obtain NB. Afterwards, A generates nonceA, then encrypt it along with IDA using QB, NB and tA as C4 and finally sends it to B; (4) Upon receiving C3 from IoTS, B decrypts it using dB, tB and NIoT S to obtain paramsM A and NA; (5) After receiving C4 from A, B decrypts it using dB, tB and NA to obtain nonceA. B then generates nonceB and use it along with nonceA and IDB to generate shared secret key with A, kBA, using a key derivation function such as HMAC-based Key Derivation Function (HKDF) [18]. After that, IDB and nonceB is encrypted using QA, NA and tB as C5 and a digest S1 is created using a message authentication code, such as HMAC [19], from message that consists of IDB, IDA, and nonceA with key kBA. Then, IDB, IDA, C5 and S1 are sent to A; (6) After C5 and S1 are received by A, C5 is decrypted by using dA, tA and NB to obtain nonceB. After obtaining nonceB, kBA is generated from nonceA, nonce and ID. After that, another S/ is generated the same way as B generated it using newly created kBA, and it is then verified against the received S1. After S1 is verified, another digest S2 is created from IDA, IDB, and nonceA with kBA and then sent to B; (7) After S2 is received, it is then verified by B. After successful verification both A and B will use kAB as they shared secret key.

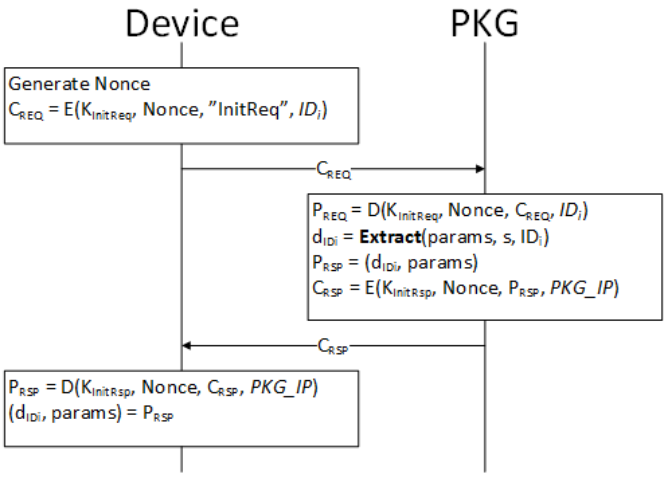


Figure 2. Device initialization protocol in a SSP! Domain

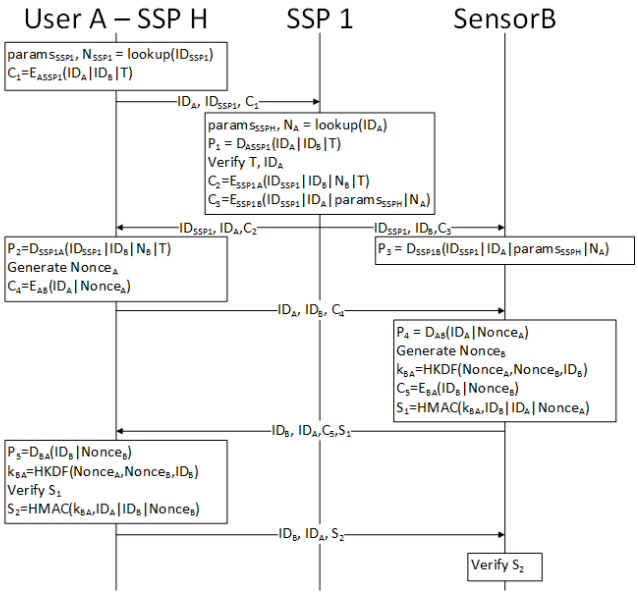


Figure 3. Authentication mechanism with entity in different domain

By the end of this phase, both user A and sensor B are mutually authenticated. They can further communicate securely using symmetric key encryption, such as Advanced Encryption Standard (AES), with kBA that is lightweight than public key encryption, thus more suitable for constrained device.

3. Results and Analysis

In this section, the security capabilities of the proposed scheme is analyzed. First of all, the threat model for the security analysis is presented. Then, the security features of the proposed scheme is analyzed. Finally, the security in terms of mutual authentication is discussed.

Threat Model

There are three types of attacker to be discussed in this section: (1) Oustide Attacker such as eaves-drops on every message transmitted in the system, replays the previous message to receiver, decom-poses the eavesdropped message into pieces, reassembles the pieces into new message, and sends the message to any legal entity, decrypts cipher text if obtain the corresponding key and modifies the decrypted plaintext, and utilizes the

public key of legal entity to forge message.; (2) Compromised Device such as capable of everything the outside attacker could do and utilizes the own secret key shared with MSP to decrypt eavesdropped message or forge message; and (3) Compromised SSP such as capable of everything the outside attacker could do and utilizes the own private key to decrypt eavesdropped message or forge message.

Security Feature of The Proposed Scheme

There are two main characteristics of the proposed scheme: (1) the message is authenticated when it is encrypted: When sender i communicates with receiver j, the sender needs to use ti to encrypt the message, and the receiver needs to use Ni to decrypt the message. Only the correct (ti, Ni) pair could ensure the message is encrypted and decrypted correctly. That means only if the message is encrypted by a legitimate sender i, the receiver could decrypt it by corresponding Ni. So the message is authenticated with the encryption and no more signatures are needed; and (2) the scheme is without key escrow problem: When a receiver j wants to decrypt a message, it needs to use dj and tj. The dj is known to the receiver, and the SSP and the tj is only known to the receiver. So even the SSP is compromised or the private key dj is leaked, the message could still only be decrypted by the receiver because of the tj. So the existence of tj solves the key escrow problem. Whats more, the updating of tj improves the security of the authentication scheme.

Mutual Authentication

A mutual authentication among a user from hospital, a medical sensor, and the SSP1! (SSP1!) can be achieved with the authentication scheme. The ID of hospital user is verified by the SSP1! in step 2 of the authentication mechanism. Only message encrypted by legitimate hospital user could be decrypted by the SSP1! with associated sub-public key NID. Furthermore, the sub-secret key tID ensures that only the legitimate mobile user could make the message authenticated with encryption and only the target sensor could decrypt that message and the other way around.

4. Conclusion

To ensure the security and privacy of IoT-based health care system is a very challenging task. It becomes more challenging due to the fact that IoT is mostly used to connect between patients with medical institutions or among several health care providers that are located across different domains with different trust authority. A scheme based on IBC has been proposed to secure communication in IoT-based health care system across multiple domains. The main contributions include authentication mechanism based on IBE that has key-escrow free feature, mechanism to lookup for IBE system parameters in other domains and to generate shared secret key for secure communication between communicating entities. Security analysis on the threat model, security feature, and mutual authentication has also been presented.

In order to enable verification and add more security on the identity, a cryptographic identity could be used instead of a plain identity, which is still left as an open issue. Furthermore, an extension of the proposed scheme with the extended IoT- based health care system architecture needs to be considered in order to take into account more stakeholders as discussed in the model proposed by [16]. Finally, implementation of the proposed scheme in the prototype or actual IoT system is another future work in order to measure the performance and practical feasibility of it.

References

[1] D. Lake, R. Milito, M. Morrow, and R. Vargheese, “Internet of things: Architectural framework for ehealth security,” Journal of ICT Standardization, vol. 1, no. 3, pp. 301–328, mar 2014.

[2] W. Jatmiko et al., “Developing smart telehealth system in indonesia: Progress and challenge,” in 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Oct 2015, pp. 29–36.

[3] B. Wiweko, A. Zesario, and P. G. Agung, “Overview the development of tele health and mobile health application in indonesia,” in 2016 Inter- national Conference on Advanced Computer Science and Information Systems (ICACSIS), Oct 2016, pp. 9–14.

[4] S. M. R. Islam et al., “The internet of things for health care: A comprehensive survey,” IEEE Access, vol. 3, pp. 678–708, June 2015.

[5] A. Al-Fuqaha et al., “Internet of things: A survey on enabling tech- nologies, protocols, and applications,” IEEE Communications Surveys Tutorials, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[6] S. Alasmari and M. Anwar, “Security privacy challenges in iot-based health

cloud," in 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Dec 2016, pp. 198– 201.

[7] Z. Cheng, R. Comley, and L. Vasiu, "Remove key escrow from the identity-based encryption system," in Exploring New Frontiers of Theoretical Informatics: IFIP 18th World Computer Congress TC1 3rd International Conference on Theoretical Computer Science (TCS2004), J.-J. Levy, E. W. Mayr, and J. C. Mitchell, Eds. Springer US, 2004, pp. 37–50.

[8] C. Doukas et al., "Enabling data protection through pki encryption in iot m-health devices," in 2012 IEEE 12th International Conference on Bioinformatics Bio-engineering (BIBE), Nov 2012, pp. 25–29.

[9] A. Shamir, "Identity-based cryptosystems and signature schemes," in Advances in Cryptology: Proceedings of CRYPTO 84, G. R. Blakley and D. Chaum, Eds. Springer Berlin Heidelberg, 1985, pp. 47–53.

[10] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in Advances in Cryptology — CRYPTO 2001: 21st Annual International Cryptology Conference, J. Kilian, Ed. Springer Berlin Heidelberg, 2001, pp. 213–229.

[11] S. Zhao, A. Aggarwal, R. Frost, and X. Bai, "A survey of applications of identity-based cryptography in mobile ad-hoc networks," IEEE Communications Surveys Tutorials, vol. 14, no. 2, pp. 380–400, Second 2012.

[12] A. Papanikolaou, K. Rantos, and I. Androulidakis, "Proxied ibe-based key establishment for llns," in The 10th International Conference on Digital Technologies 2014, July 2014, pp. 275–280.

[13] F. Corella and K. P. Lewison, "Identity-based protocol design patterns for machine-to-machine secure channels," in 2014 IEEE Conference on Communications and Network Security, Oct 2014, pp. 91–96.

[14] T. Markmann, T. C. Schmidt, and M. Wa¨hlisch, "Federated end-to-end authentication for the constrained internet of things using ibc and ecc," in Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, ser. SIGCOMM '15. ACM, 2015, pp. 603– 604.

[15] S. Chen, M. Ma, and Z. Luo, "An authentication scheme with identity- based cryptography for m2m security in cyber-physical systems," Security and Communication Networks, vol. 9, no. 10, pp. 1146–1157, 2016.

[16] G. Neagu, Preda, A. Stanciu, and V. Florian, "A cloud-iot based sensing service for health monitoring," in 2017 E-Health and Bioengi- neering Conference (EHB), June 2017, pp. 53–56.

[17] D. McGrew, "An interface and algorithms for authenticated encryption," RFC 5116, January 2008.

[18] H. Krawczyk and P. Eronen, "Hmac-based extract-and-expand key derivation function (hkdf)," RFC 5869, May 2010.

[19] H. Krawczyk, M. Bellare, and R. Canetti, "Hmac: Keyed-hashing for message authentication," RFC 2104, February 1997.

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 10-16
DOI: http://dx.doi.org/10.21609/jiki.v11i1.580

IMPROVEMENT METHOD OF FUZZY GEOGRAPHICALLY WEIGHTED CLUSTERING USING GRAVITATIONAL SEARCH ALGORITHM

Imam Habib Pamungkas[1] and Setia Pramana[2,3]

[1]BPS Statistics Indonesia, Jl. Dr. Sutomo 6-8, Jakarta, 10710 Indonesia
[2]Institute of Statistic Jakarta, Jl. Otto Iskandardinata No. 64C, Jakarta, 13330, Indonesia
[3]Medical Epidemiology and Biostatistics Department, Karolinska Instutet, SE-171 77 Stockholm, Sweden

E-mail: setia.pramana@stis.ac.id

Abstract

Geo-demographic analysis (GDA) is a useful method to analyze information based on location, utilizing several spatial analysis explicitly. One of the most efficient and commonly used method is Fuzzy Geographically Weighted Clustering (FGWC). However, it has a limitation in obtaining local optimal solution in the centroid initialization. A novel approach integrating Gravitational Search Algorithm (GSA) with FGWC is proposed to obtain global optimal solution leading to better cluster quality. Several cluster validity indexes are used to compare the proposed methods with the FGWC using other optimization approaches. The study shows that the hybrid method FGWC-GSA provides better cluster quality. Furthermore, the method has been implemented in R package spatialClust.

Keywords: *Clustering, Fuzzy Geographically Weighted Clustering (FGWC), Gravitational Search Algorithm (GSA)*

Abstrak

Analisis geo-demografi (GDA) adalah metode yang berguna untuk menganalisis informasi berdasarkan lokasi, dengan memanfaatkan beberapa analisis spasial secara eksplisit. Salah satu metode yang paling efisien dan umum digunakan adalah Fuzzy Geographically Weighted Clustering (FGWC). Namun, ia memiliki keterbatasan dalam mendapatkan solusi optimal lokal pada inisialisasi centroid. Pendekatan baru yang mengintegrasikan Algoritma Pencarian Gravitasi (GSA) dengan FGWC diusulkan untuk mendapatkan solusi optimal global yang mengarah pada kualitas cluster yang lebih baik. Beberapa indeks validitas cluster digunakan untuk membandingkan metode yang diusulkan dengan FGWC menggunakan pendekatan optimasi lainnya. Studi tersebut menunjukkan bahwa metode hibrida FGWC-GSA memberikan kualitas cluster yang lebih baik. Selanjutnya metode tersebut telah diimplementasikan pada paket R spasialClust.

Kata Kunci: *Clustering, Fuzzy Geographically Weighted Clustering (FGWC), Algoritma Pencarian Gravitasi (GSA)*

1. Introduction

Nowadays geographical data are available and easy to be accessed and getting more attention to be included in the analysis and commonly used to observe people behavior based on their location. Geo-demographic analysis (GDA) is the analysis of spatially geo-demographic and lifestyle data [1]. Geo-demographic analysis explores information based on location, utilizing several spatial analyses explicitly.

Geo-demographic analysis often uses clustering techniques to classify the geo-demographic data into groups, making the data more manageable for analysis purposes [2].

GDA relay on two assumptions: (1) two individuals who live in the same area are more likely to have similar characteristics than individuals selected at random, and (2) two areas can be characterized in terms of their population, using demographics and other measures. Based on these two principles, clustering can be applied to group geo-demographic data and lead to meaningful results [1].

In GDA, fuzzy clustering is commonly used with different approaches such as Bezdek's Fuzzy C-means Clustering (FCM), Gustafson-Kessel, Neighborhood Effect, and Fuzzy Geographically Weighted Clustering (FGWC). FCM is the most popular clustering method because it is easy to use and efficient [2]. FGWC was proposed to

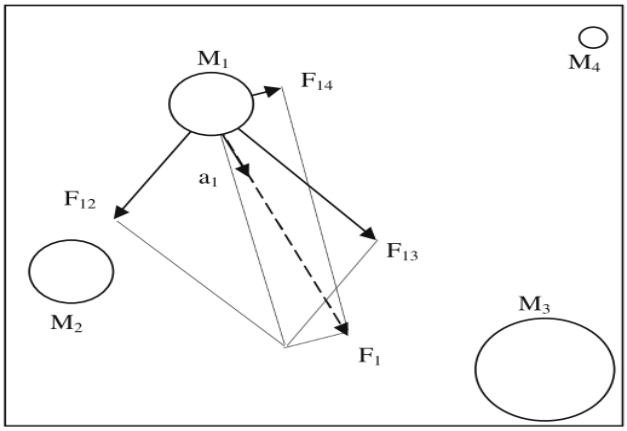


Figure 1. Basic Concept of Object Interaction in GSA [6]

improve FCM on handling spatial data with incorporated geographic and neighborhood data [3]. FGWC was inspired by a hypothesis statement, if we incorporated neighborhood effect to fuzzy clustering, the result will be geographically aware [3].

Similar to the FCM method, FGWC also has limitation in initial phase. The random cluster centroid initialization makes FGWC easily trapped in local optimal solution that effect the cluster quality. Several attempts have been done by using different optimization approach, for example, Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), and Simulated Annealing (SA) [1].

Gravitational Search Algorithm (GSA) is new optimization algorithm which focus in obtaining a global solution. To improve the cluster quality of FGWC, this research aims to integrate the Gravitational Search Algorithm to avoid FGWC falling in local optimal solution providing better clustering result.

Theoritical Background

Fuzzy C-Means (FCM)

Fuzzy C-Means is one of the most popular clustering algorithms aimed to minimize the following objective function based on membership value of the object and distance of the object to the cluster centroids [4]:

$$J_m(\mathbf{U},\mathbf{V})=\sum_{k=1}^{N}\sum_{i=1}^{c}(\mu_{ik})^m\left\|y_k-v_i\right\|^2 \qquad (1)$$

where "y"_"k" is the k-th observation, c is the initial cluster number, m is degree of the fuzziness, "U" is membership matrix that contains membership degree ("μ"_"ik"), "μ"_"ik" membership degree between the i-th and cluster c, "V" is centroid matrix that contains value of cluster centroid, and "v"_"i" is cluster centroid of cluster c.

Membership degree of each object in FCM is changing during the iteration by using the following function:

$$\mu_{ik}=\left(\sum_{j=1}^{c}\left(\frac{d_{ik}}{djk}\right)^{\frac{2}{m-1}}\right)^{-1};\ 1\le k\le N;\ 1\le i\le , \qquad (2)$$

where "d" is the Euclidean distance between data and cluster centroid. The cluster centroid is defined as follows:

$$v_i=\frac{\sum_{k=1}^{N}(\mu_{ik})^m y_k}{\sum_{k=1}^{N}(\mu_{ik})^m};\ 1\le i\le c \qquad (3)$$

Fuzzy Geographically Weighted Clustering (FGWC)

FGWC proposed by Mason and Jacobson is an extension of version of FCM that more geographically aware [3]. This algorithm takes into account basic spatial interaction effect into the model. The adaptation of spatial effect is performed in each iteration of the following membership matrix calculation:

$$\mu'_i = \alpha\times\mu_i+\beta\times\frac{1}{A}\sum_{j}^{n}w_{ij}\times\mu_i, \qquad (4)$$

where 〖 μ '〗 _i is membership value of area i-th, μ _i is old membership value before incorporating spatial effect, and A is scale value to ensure that sum of membership matrix equal 1. The parameters α and β control the membership proportion after and before weighting a+β=1.

$$w_{ij} = \frac{\left(p_i \times p_j\right)^b}{z_{ij}^a} \quad (5)$$

where p_i and p_j are number of population of area i and j, respectively, and z_ij is the distance between the two areas. The other two parameters, a and β tune the effect of the distance and population on the weight and are defined by the users.

Gravitational Search Algorithm (GSA)

GSA is one of population based algorithm [5] developed by [6]. The aim of this algorithm is improving exploration and exploitation of the population based algorithm to reach optimal solution. GSA is naturally inspired by law of motion and Newtonian gravity.

Every object in GSA is called agent and the capability of each agent measured by his mass. Each agent in GSA will interact based on law of gravity. Agent with small capability will move to agent with large capability.

Figure 1 shows that agent M1 is affected by agents M2, M3 and M4. According to the law of gravity, M1 has a resultant force which will make it move towards agent M3. Agent M2, M3, M4 also has resultant fore to each other.

The first step in GSA is randomly generate initial N solutions with m dimension. The agent position is represented as follows:

$$X_i = (X_{i1}, \dots, X_{id}, \dots, X_{im}), \quad (6)$$

In each iteration, the following total force in each agent (F) is evaluated:

$$F_{ij}{}^d(t) = G(t)\frac{M_i(t)M_j(t)}{R_{ij}(t)}\left(x_i{}^d(t) - x_i{}^d(t)\right), \quad (7)$$

$$F_i{}^d(t) = \sum_{j=1, j\neq i}^{N} rand_i F_{ij}{}^d(t), \quad (8)$$

where ⟦x_i⟧ ^d represents the agent position, (t) is gravitation constant at t, M_i (t) is mass of agent i, and R_ij (t) is the euclidean distance between agent.

$$R_{ij}(t) = \left\|X_i(t), X_j(t)\right\|_2, \quad (9)$$

(t) is updated in each iteration using the following function

$$G(t) = G(G_0, t), \quad (10)$$

where G_0is gravity constant.
The agent mass M_i (t) is defined as follows:

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)}, \quad (11)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)}, \quad (12)$$

⟦fit⟧ _i (t) is current fitness value from the solution. The best and worst determined by fitness value. There are two minimization functions to get best and worst:

$$best(t) = \min_{j\in\{1\dots N\}} fit_j(t), \quad (13)$$

$$worst(t) = \max_{j\in\{1\dots N\}} fit_j(t). \quad (14)$$

Whereas the maximization functions:

$$best(t) = \max_{j\in\{1\dots N\}} fit_j(t), \quad (15)$$

$$worst(t) = \min_{j\in\{1\dots N\}} fit_j(t). \quad (16)$$

The acceleration (a) and velocity (v) each agent are defined:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)}, \quad (17)$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) \times a_i^d. \quad (18)$$

The last step is updating the position of each agent x.

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1). \quad (19)$$

Repeat step until maximum iteration or reach stopping criterion.

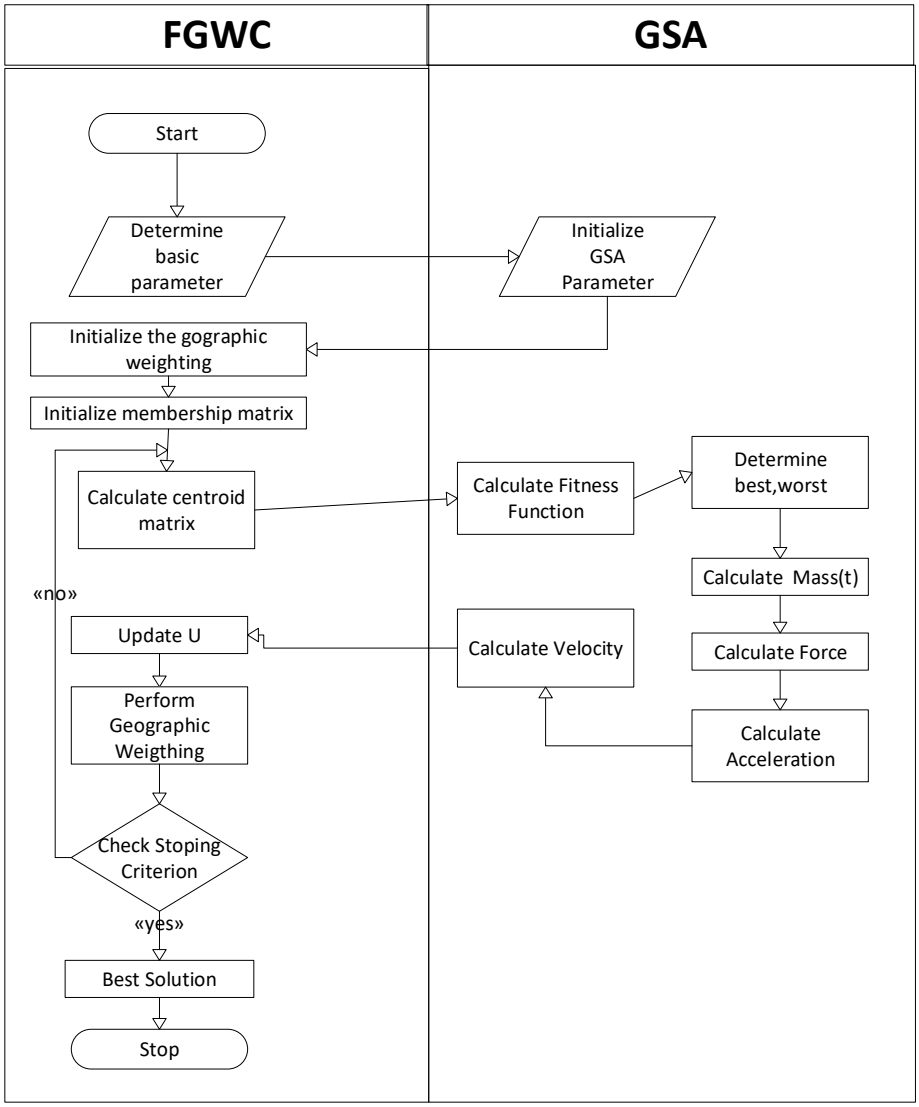


Figure 2 Flow chart of proposed method

Cluster Validity Index

The main problem of cluster validity is finding objective criterion to determine the partition value from the clustering algorithm [7]. In this research we use several cluster validity index. I.e. Partition Coefficient (PC), Classification Entropy (CE), Separation Index (S), Xie Beni Index, and IFV index.

Partition Coefficient measures the average number of relatively degree sharing of each object in membership matrix. The greater value of PC indicate better clustering quality.

$$PC = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{2}, \qquad (20)$$

where "μ" _ij is membership degree of item j in cluster i.

The Classification Entropy (CE) index is used to define the fuzziness of partition in each cluster:

$$CE = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}\log_a(\mu_{ij}). \qquad (21)$$

The Separation Index and Xie Beni Index measure the compactness and the separation of each cluster. The minimum value of Separation Index and Xie Beni Index indicate the better clustering validity.

TABLE 1.
VARIABLES USED IN THE ANALYSIS

Variables	Description
Literacy rate	Proportion of population of certain age group that can read and write Latin letters, letters Arabic, or Other letters
Mean years of schooling	Mean of population
Expected mean years schooling	Average number of years spent by population aged 15 years for formal education
Net enrolment rate primary school	Proportion of population of primary school age that actually attend primary school
Net enrollment rate junior high school	Proportion of population of junior high school age that actually attend junior high school
Primary school Teacher and student ratio	Proportion of number of teacher and student
Primary school student and school ratio	Proportion of number of teacher and student
Junior high school teacher and student ratio	Proportion of number of teacher and student
Junior high school student and school ratio	Proportion of number of teacher and student
Junior high school dropout rate	Proportion of population of junior high school age that dropped out from school
Average monthly expenditure per capita	Average monthly expenditure per capita especially for education

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}\left(\mu_{ij}\right)^{2}\left\|x_j - v_i\right\|^{2}}{N\, min_{i,k}\|v_k - v_i\|^{2}}, \quad (22)$$

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}\left(\mu_{ij}\right)^{m}\left\|x_j - v_i\right\|^{2}}{N\, min_{i,j}\left\|v_j - v_i\right\|^{2}}, \quad (23)$$

where v_i is the centroid of the cluster i.

To validate the cluster fuzziness in spatial data, IFV is implemented as it is stable and robust [8]. Higher IFV index shows better result.

$$IFV = \frac{1}{c}\sum_{j=1}^{c}\left\{\frac{1}{N}\sum_{k=1}^{N}\mu_{kj}^{2}\left[\log_2 c - \frac{1}{N}\sum_{k=1}^{N}\log_2 \mu_{kj}\right]^{2}\right\}. \quad (24)$$

2. Methods

The Improved FGWC using GSA

As mentioned before cluster center initialization in FGWC could fall in local optimal solution easily affecting the clustering results. We propose to minimize the objective function by using GSA to initialize the initial centroid. This is the objective function that will be minimized.

$$J_m(\boldsymbol{U},\boldsymbol{V}) = \sum_{k=1}^{N}\sum_{i=1}^{c}(\mu_{ik})^{m}\|y_k - v_i\|^{2}. \quad (24)$$

Here is step by step of the proposed methods:

Step 1: Determine the basic parameter, number of cluster c, degree of fuzziness m, threshold of error, maximum number of iteration, and some parameter for weighted function.

Step 2: Initialize the GSA parameter such as gravity constant G.

Step 3: Initialize the geographic weighting.

Step 4: Initialize membership matrix.

Step 5: Calculate centroid matrix.

Step 6: Calculate fitness function and start to optimize FGWC using GSA.

Step 7: Update membership matrix

Step 8: Perform geographic weighting to membership matrix

Step 9: Repeat step 5-8 until reach stopping criteria.

For more detail, the step can be seen at Figure 2

We compare the proposed method with the standard FGWC and different optimization approaches such as Particle Swarm Optimization, Artificial Bee Colony and Simulated Annealing, using a case study of Educational Profile of Jawa Tengah Province 2015 published by BPS' Statistics of Jawa Tengah Province, Indonesia. The variables were selected based on research conducted by Bustomi in 2012 [9] which conclude that inequality education in Central Java Province caused by 4 dimensions.

According to the dimension, we choose 11 variables that represent civil participation, education quality and facilities. The details of each variable is presented in Table 1. Hence the dataset used contains 11 variables of 35 regencies in Central Java Province, Indonesia.

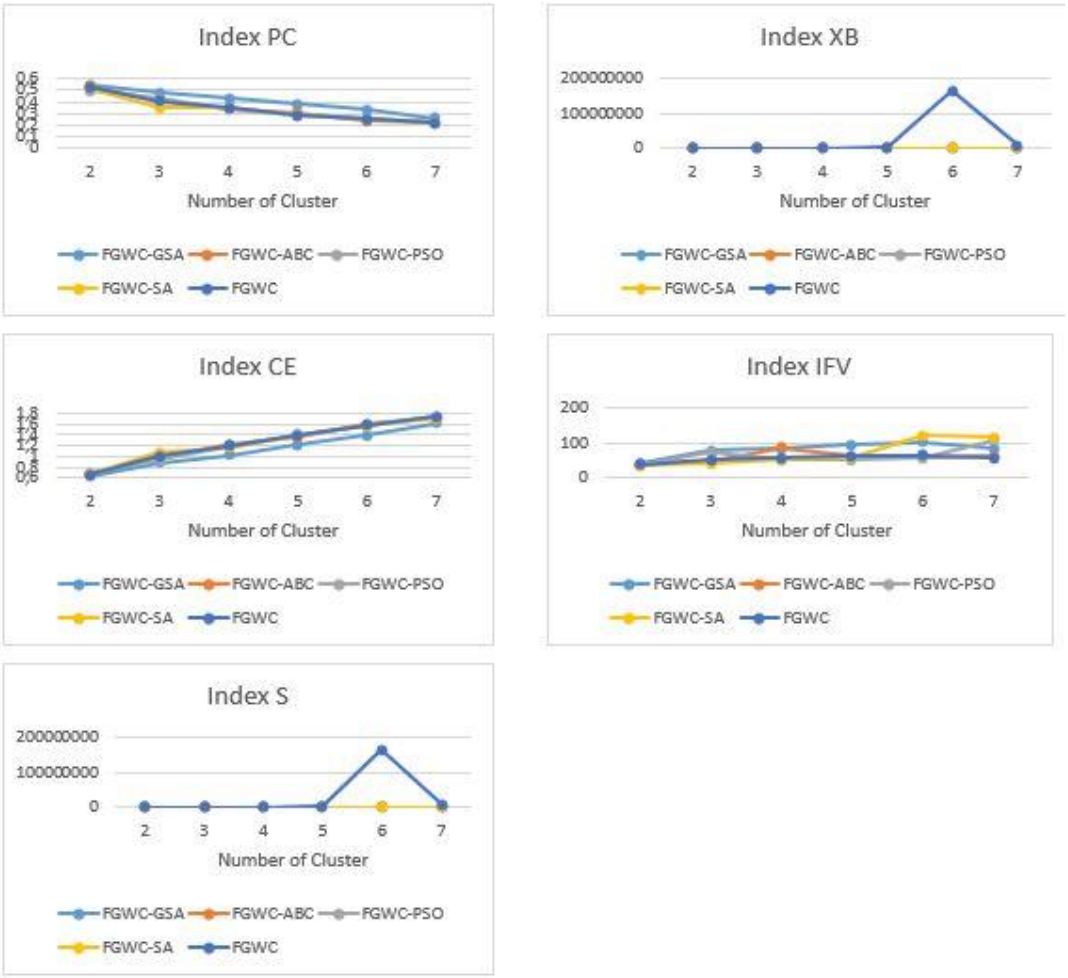


Figure 3. The result of Cluster Validity in different number of clusters obtained from standard FGWC, FGWC-GSA and the other optimization approaches.

The proposed method is implemented in R and now available in CRAN (spatialClust package). The Graphical Interface is also available in FAST [10] which can be accessed through www.stis.ac.id/fast.

3. Results and Analysis

Figure 3 shows the results of different cluster validity indexes from the case study for standard FGWC, FGWC-GSA and other approaches. The x-axis is different number of clusters, and the y-axis is the corresponding validity index.

It can be seen that FGWC-GSA give higher Partition Coefficient Index, IFV index and as Classification Entropy (CE) compared to the other approaches. Furthermore, the FGWC-GSA provide lower Separation Index and Xie Beni Index compared to the others.

In general we observed that FGWC-GSA outperforms FGWC and the other optimization approaches in all validity indexes and all number of clusters.

Based on the results of the analysis, areas in Central Java can be grouped into three clusters based on educational indicators. Cluster 3 contains regencies (such as Semarang and Salatiga) with the high educational quality. While the cluster 1 (e.g., Cilacap, Purbalingga) is the cluster with poor education quality. Cluster 2 consisting medium education quality such as Sragen and Banyumas.

4. Conclusion

In this research, we proposed a new method to avoid local optimal solution that may occur in initial phase of centroid in FGWC, using Gravitational Search Algorithm (GSA) approach. The results show that the proposed method out performs the standard FGWC and its other modification in terms of cluster validity.

References

[1] W. W. Arie, Improvement of Fuzzy Geo-Demographic Clustering using Metaheuristic Optimization on Indonesia Population Cencus, Bandung: ITB, 2014.

[2] A. Mulyanto and R. S. Wahono, "Penerapan Gravitational Search Algorithm untuk Optimasi Klasterisasi Fuzzy C-Means," Jurnal of Intelligent System,vol.1,no.1, pp. 42-47, 2015.

[3] G. A. Mason and R. D. Jacobson, "Fuzzy Geographically Weighted Clustering," in in Proceedings of the 9th International Conference on Geocomputation, 2007.

[4] J. C. R. E. &. W. F. Bezdek, " FCM: The Fuzzy c-meansCLustering Algorithm," Computers & Geosciences, vol. 10, pp. 191-203, 1984.

[5] R. Khadanga and S. Panda, "Gravitational search algorithm for Unified Power Flow Controller based damping controller design," 2011 International Conference on Energy, Automation and Signal, pp. 1-6, 2011.

[6] E. Rashedi, H. Nezamabadi-pour and S. Saryazdi, "GSA: A Gravitational Search Algorithm," Information Sciences, vol. 179, no. 13, p. pp. 2232–224, 2009.

[7] C. Oscar, R. Elid, S. Jose and N. Enrique, "Optimization of the Fuzzy C-Means Algorithm using Evolutionary Methods," Engineering Letters, 20:1, EL_20_1_08, 2012.

[8] L. H. Son, B. C. Cuong, P. L. Lanzi and N. T. Thong, "A novel intuitionistic fuzzy clustering method for geo-demographic analysis2," Expert Syst. Appl., vol. 39, no. 10, p. 9848–9859, 2012.

[9] M. J. Bustomi, "Ketimpangan Pendidikan Antar Kabupaten/Kota dan Implikasinya di Provinsi Jawa Tengah," Economics Development Analysis, 2012.

[10] D.S. M Dalimunthe L. et.al. (2014). FAST: a Web-Based Statistical Analysis Forum. Proceeding Islamic Countries Conference on Statistical Science 13 (pp.185-200)

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 17-26
DOI: http://dx.doi.org/10.21609/jiki.v11i1.581

MULTI-OBJECT DETECTION AND TRACKING USING OPTICAL FLOW DENSITY HUNGARIAN KALMAN FILTER (OFD - HKF) ALGORITHM FOR VEHICLE COUNTING

Muhamad Soleh[1], Grafika Jati[1], and Muhammad Hafizuddin Hilman[2]

[1]Faculty of Computer Science, University of Indonesia, Kampus UI, Depok, 16422, Indonesia
[2]Melbourne School of Engineering, University of Melbourne - Australia

E-mail: muhamad.soleh61@ui.ac.id

Abstract

Intelligent Transportation Systems (ITS) is one of the most developing research topics along with growing advanced technology and digital information. The benefits of research topic on ITS are to address some problems related to traffic conditions. Vehicle detection and tracking are one of the leading steps to realize the benefits of ITS. There are several problems related to vehicles detection and tracking. The appearance of shadow, illumination change, challenging weather, motion blur and dynamic background are significant challenges issue in vehicles detection and tracking. Vehicles detection in this paper using the Optical Flow Density algorithm by utilizing the gradient of object displacement on video frames. Gradient image feature and HSV colour space on Optical Flow Density guarantee the object detection in illumination change and challenging weather for more robust accuracy. Hungarian Kalman filter algorithm used for vehicle tracking. Vehicle tracking used to solve miss detection problems caused by motion blur and dynamic background. Hungarian Kalman filter combines the recursive state estimation and optimal solution assignment. The future position estimation makes the vehicles detected although miss detection occurrence on vehicles. Vehicles counting used single line counting after the vehicles pass that line. The average accuracy for each process of vehicles detection, tracking, and counting was 93.6%, 88.2% and 88.2% respectively.

Keywords: *Intelligent Transportation Systems, Optical Flow Density, Hungarian Kalman Filter, Single Line Counting*

Abstrak

Intelligent Transportation Systems (ITS) merupakan salah satu topik penelitian yang terus berkembang seiring dengan kemajuan teknologi dan informasi digital. Beberapa manfaat yang diperoleh dari penelitian ITS diantaranya adalah untuk mengatasi beberapa permasalahan terkait dengan keadaan lalu lintas. Pendeteksian dan pelacakan kendaraan merupakan salah satu langkah untuk mewujudkan manfaat dari ITS. Keberadaaan bayangan, perubahan iluminasi, perubahan cuaca, *motion blur*, dan *backgound* yang dinamis merupakan tantangan dalam peneteksian dan pelackan objek. Pendeteksian kendaraan pada penelitian ini menggunakan algoritma *Optical Flow Density* dengan memanfaatkan gradient perpindahan objek pada frame video. Fitur gradient image dan ruangwarna HSV pada algoritma *Optical Flow Density* menjamin pendeteksian objek pada kondisi perubahan iluminasi dan perubahan cuaca untuk hasil akurasi yang lebih robust. Algoritma Hungarian kalman filter digunakan untuk pelacakan kendaraan. Pelacakan kendaraan digunakan untuk menyelesaikan permasalahan miss detection yang disebabkan karena motion blur dan background yang dinamis. Hungarian kalman filter mengkombinasikan metode *state estimation* dengan *optimal assignment*. Prediksi posisi objek di masa depan dapat mendeteksi objek walaupun terjadi *miss detection*. Perhitungan kendaraan menggunakan *Single Line Counting* setelah kendaraan berhasil melewati garis tersebut. Rata-rata akurasi untuk masing-masing proses adalah 93,6% untuk pendeteksian, 88,2% untuk pelacakan, dan 88,2% untuk perhitungan kendaraan.

Kata Kunci: *Intelligent Transportation Systems, Optical Flow Density, Hungarian Kalman Filter, Single Line Counting*

1. Introduction

Intelligent Transportation Systems (ITS) is one of the emerging research topics. Along with advances in technology and digital information. Some recent research related to ITS is Video Traffic Surveillance (VTS) [1][2][3][4][5] and Advanced Driver Assistance System (ADAS) [6][7]. VTS is a system that can monitor traffic conditions on the highway by using cameras, lasers, radar, lidar, and

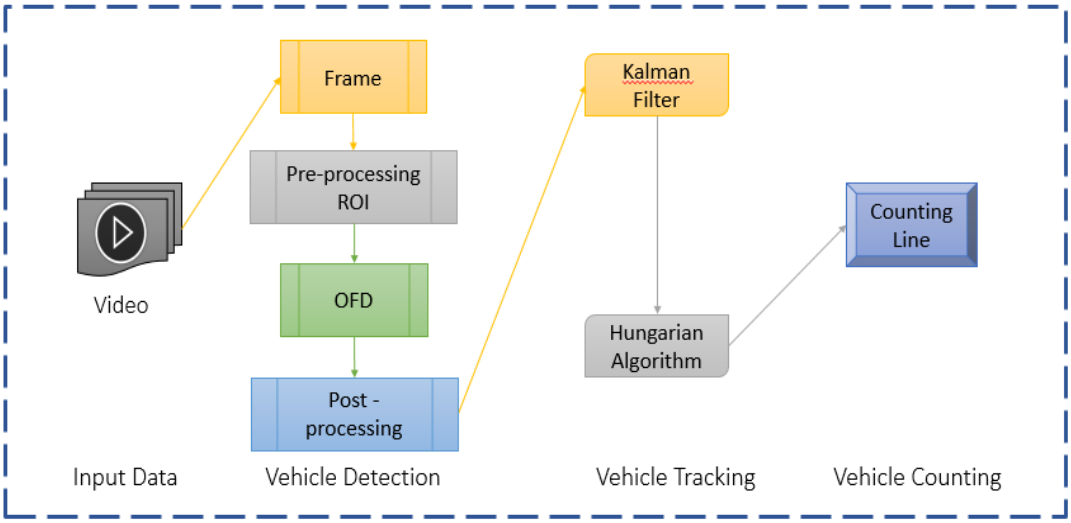


Figure 1 Research methodology frameworkPrediction step:

other equipment [2]. VTS Research using a camera is exciting research because of several advantages. The cost is relatively cheaper; the obtained information shows the best performance, as well as information on the shape, size, and colour of the monitored object, is easily obtained [1][2].

The benefit of VTS Research is to accomplish some problems related to traffic conditions, such as traffic congestion [8], prevention of traffic accidents [9], reduce the risk of human error [10], and many other benefits. This paper aims to solve traffic congestion using adaptive traffic lights. The timing of traffic lights is arranged based on the vehicles density. The number of vehicles can be counted by using two framework approaches. The first framework is vehicle counting based on object detection. The second framework uses three stages in vehicle counting: detection, tracking and counting the number of vehicles.

In 2010, Rashid et al. [12] conducted research using the first framework. Rashid et al. use the time-spatial image (TSI) by generating those pixels of the moving objects that pass the virtual detection line (VDL). Then, the vehicles are counted by detecting the blobs in TSI. The weakness of the detection-based vehicle counting is potential double counting. In 2016, Shiva and Reza [2] conducted research using a second framework approach. Shiva and Reza combine several existing methods to build vehicle counting systems. Detection of the vehicle using active basis model (ABM) method combined with symmetry checking (SC). ABM is used to get sketches from vehicles while SC is used for vehicle shape verification process. Distance similarity measurement (DSM) with the manual threshold has been used to track the vehicle. The single line counting (SLC) is used to count the number of vehicles. Detection of vehicles using ABM works quite well, but tracking and counting of vehicles are not equal to the results of detection. Because the detected vehicle does not occur in every frame, it caused the decreasing accuracy on vehicle tracking and counting.

Both frameworks that have been proposed by [2] and [12] have not been accurate enough to count the number of vehicles especially for Afternoon, Sundown, Night, and Rainy datasets. Some challenges that can not be solved using the previous framework [2] and [12] are the present of shadow, illumination change, challenging weather, motion blur and dynamic background. This paper presents a novel framework for vehicle counting using three stages. Vehicle Detection, Tracking, and Counting using Optical flow density (OFD) - Hungarian Kalman filter (HKF) - Single line counting (SLC). Optical flow density (OFD) utilizes the image gradient feature to detect the movement of objects in the video. OFD is used to solve some of the challenges in object detection, such as illumination change, challenging weather, and Headlamp lights effect.

Hungarian Kalman filter (HKF) is a multi-object tracking method that combines the Kalman filter method with Hungarian algorithm. Kalman filters are used to predict the position of the object, while the Hungarian algorithm is used for data associations for each vehicle. Hungarian Kalman filter (HKF) is used to solve the challenge of object tracking, i.e. motion blur. The target region is blurred due to the motion of the target or the camera [11]. Motion blur can cause the vehicle object to become undetected (miss detection) because the object has a blurring noise. The process of tracking objects using future estimation methods

such as Hungarian Kalman filter is needed to track objects when miss detection occurs. Vehicles are counting using a single line counting. Single line counting will count every vehicle that passes the counting line.

Related Works

Vehicle detection

Vehicle detection in [13] consists of three stages. Pre-processing, Gaussian mixture model, and post-processing. Pre-processing using region of interest (ROI). ROI is an area that became the centre of research attention. The ROI segmentation intended to limit the area to be processed during object detection. Non-ROI area will be considered as a background. Every moving object in Non- ROI area never detected as a vehicle, such as moving leaves of tree or pedestrian.

In [14], Optical flow density is better than Gaussian mixture model for vehicles detection. Optical flow density is one of object detection method based on motion appearance. From video dataset, the input of optical flow density is two consecutive frames. Optical flow density is a movement pattern that indicates the existence of an object in two successive frames caused by the movement of objects or camera.

Optical flow works by using several assumptions:

1. The pixel intensities of an object do not change between consecutive frames. [15]
2. The neighbouring pixel has a similar pixel motion [16].
3. Optical Flow uses vector flow which has magnitude and direction [17]

The image gradient is a vector of the first derivative of a pixel [14]. Using assumption 1, we get the intensity of an image that moves along (dx, dy) mathematically can be written on the equation 1:

$$I(x,\ y,\ t)\ =\ I(x + dx,\ y + dy\ ,\ t + dt) \qquad (1)$$

In the 2D or 3D image the gradient function can be calculated numerically using Taylor series Expansion [14], So the optical flow equation on the intensity change (dx, dy) over time (dt) in the image can be written into the equation 2:

$$fx\ u + fy\ v + ft\ \ = 0 \qquad (2)$$

Where fx and fy are the image gradient functions on the x and y axis whereas u and v are unknown variables. The Farnebäck algorithm is used to obtain the values of u and v. The value of u and v is the vector flow in the image gradient. Gunner Farneback's algorithm works using the principle of polynomial base approximation. The polynomial expansion assumption is used to estimate the approximate value of a neighboring pixel on a two-dimensional function [23]. Considering the quadratic basis of polynomial $1, x^2, y^2, \mathrm{x}, \mathrm{y}, \mathrm{xy}$, the intensity of the pixel value in the image is represented as a polynomial function can be written on the equation 3:

$$f(x)\ =\ x^T A\ x + B^T x + C \qquad (3)$$

Where A is a symmetric matrix, B is a vector and c is a constant. The intensity of the image on moving object along (dx, dy) can be mathematically written as equation 4 and 5. Intensity on frame at time t:

$$f_1(x)\ =\ A_1 x^T x + {B_1}^T x + C_1 \qquad (4)$$

Intensity on the frame at time dt:

$$\mathrm{f_2(x)} = A_1(\mathrm{x} - \mathrm{d})^{\mathrm{T}}(\mathrm{x} - \mathrm{d}) + {B_1}^{\mathrm{T}}(\mathrm{x} - \mathrm{d}) + C_1 \qquad (5)$$

By operating the above equation, we get the equation 6:

$$\mathrm{f_2(x)} = A_1(\mathrm{x})^{\mathrm{T}}(\mathrm{x}) + (B_1 - 2A_1\mathrm{d})^{\mathrm{T}}(\mathrm{x}) + A_1(\mathrm{d})^{\mathrm{T}}\mathrm{d} - {B_1}^{\mathrm{T}}\mathrm{d} + C_1 \qquad (6)$$

The coefficients in both functions follow the assumption that the intensity at the two-pixel values of a displaced object (dx, dy) over an interval dt does not change between two successive frames, so the relation coefficients of intensity betweet frame at t and dt are:

$$A_2 = A_1 \qquad (7)$$

$$B_2 = (B_1 - 2A_1 d) \qquad (8)$$

$$C_2 = (\mathrm{d})^{\mathrm{T}} A_1 d - (B_1)^{\mathrm{T}} d + C_1 \qquad (9)$$

From equation eight then we get the intensity displacement equation (dx, dy) can be written with equation 10 and 11 as follows:

$$dA_1 = -\tfrac{1}{2}(B2 - B1) = \Delta B \qquad (10)$$

$$d = \Delta B\ {A_1}^{-1} \qquad (11)$$

Thus the displacement vector in each pixel can be known by substituting the coefficient of the polynomial equation between frames. Assuming that there is an overlapping region between frames showing the movement of an object. It is based on the initial assumption that the neighbouring pixel value of an image has a similar pixel motion. For each kernel contained in the image, the vector displacement equation at an intensity (dx, day) show as the equation 12:

$$\sum_{\Delta x \epsilon 1} w(\Delta x)||\ A(x + \Delta x)d(x) - \ \Delta B(x + \Delta x)||^2 \qquad (12)$$

Where $w\ (\Delta x)$ is the weight function at each point in the neighborhood pixel. The coefficients can be estimated by weighted least square estimate of pixel values about neighborhood. As with all

TABLE 1
VEHICLE DETECTION RESULT

Vehicle Detection	***Day Time***				***Night Time***		***Average (%)***
	Noon	***Afternoon***	***Sundown***	***Rainy***	***Night***	***Rainy***	
Actual Number of Vehicles	228	413	261	128	277	247	
Percentage of correctly detected vehicles (%)							
Rashid et al. [12]	86.8	48.4	82.8	50.0	30.7	43.3	57.0
Shiva and Reza [2]	**99.1**	**99.0**	**98.4**	94.5	**99.3**	**98.8**	**98.2**
OFD – HKF	98.7	92.7	87.0	**99.2**	92.1	91.9	93.6
Percentage of falsely detected vehicles (%)							
Rashid et al. [12]	182.9	146.5	84.7	84.4	141.2	72.1	118.6
Shiva and Reza [2]	2.2	1.9	2.3	4.7	2.9	3.6	2.9
OFD – HKF	**1.7**	**1.7**	**0.4**	**1.6**	**0.7**	**0.8**	**1.2**

TABLE 2
VEHICLE TRACKING RESULT

Vehicle Tracking	***Day Time***				***Night Time***		***Average (%)***
	Noon	***Afternoon***	***Sundown***	***Rainy***	***Night***	***Rainy***	
Actual Number of Vehicles	*228*	*413*	*261*	*128*	*277*	*247*	
Percentage of correctly tracked vehicles (%)							
Rashid et al. [12]	86.8	48.4	**82.8**	50.0	30.7	43.3	57.0
Shiva and Reza [2]	90.5	48.7	68.2	51.6	66.8	46.6	62.1
OFD – HKF	**97.8**	**82.6**	**82.8**	**99.2**	**82.7**	**83.8**	**88.2**
Percentage of miss tracked vehicles (%)							
Rashid et al. [12]	13.2	51.6	**17.2**	50.0	69.3	56.7	43.0
Shiva and Reza [2]	9.5	51.3	31.8	45.4	32.2	53.4	38.9
OFD- HKF	**2.2**	**17.4**	**17.2**	**0.8**	**17.3**	**16.2**	**11.8**

optical flow algorithm, the brightness constancy assumption is made. The brightness of a path of image in adjacent frames is constant. Then the vector displacement equation at intensity (dx, dy) is shown in equation 13:

$$d\,(x) = (\sum wA^T A\,)^{-1} \sum wA^T\,\Delta B \qquad (13)$$

The intensity displacement vector is a two-channel array representing the optical flow vector of the variables u and v. The value and direction of the optical flow vector are represented in hue and value with the HSV colour space. Post-processing uses morphological operations. The purpose of morphological operations is to remove noise, connect disconnected elements, or cover holes in the blob.

Vehicle Tracking

Vehicles tracking is the process of monitoring the position of the same object over several specified frames. Vehicle tracking serves to distinguish between one objects with another. [20] use Hungarian Kalman filter algorithm for vehicles tracking. Hungarian Kalman Filter is an algorithm that combines two existing algorithms, the Kalman filter and the Hungarian algorithm [18][20][19]. Kalman filter is used as an estimator [24][21], while Hungarian algorithm is used as optimal assignment based on minimum cost function [18][20][19].

Kalman filters work using the least square estimation principle of the linear displacement of an object [22]. Kalman filters are efficient to implement, as its method only store previous state information into memory. Kalman filter consists of two estimation processes. The first estimation is called a prediction whereas the second estimation is called a correction or update [21]. The correction or update step were evaluated from the result of prediction step.

Mathematically both estimates can be written in equation 14-19.

$$u'_{\{k|k-1\}} = Fu'_{\{k-1|k-1\}} \qquad (14)$$

$$P_{\{k|k-1\}} = FP_{\{k-1|k-1\}}\,F.T + Q \qquad (15)$$

Correction or update step:

$$u'_{\{k|k\}} = u'_{\{k|k-1\}} + K_{\{k\}}(b_{\{k\}} - Au'_{\{k|k-1\}}) \qquad (16)$$

$$P_{\{k|k\}} = P_{\{k|k-1\}} - K_{\{k\}}(CK.T) \quad (17)$$

$$K_{\{k\}} = P_{\{k|k-1\}} A.T(C.Inv) \quad (18)$$

$$C = AP_{\{k|k-1\}} A.T + R \quad (19)$$

Where u is the mean matrix, F is the displacement matrix, P is the covariance matrix, T is time, Q is the process noise, K is the Kalman Gain, R is the observation noise, b is the observation, C is the pre-fit residual covariance, and A is observation matrix. Kalman Gain serves to correct the estimation result from mean and covariant error [22]. Kalman Gain works using the principle of MMSE (Minimum Mean Square Error). The observation noise value indicates the sensitivity to the update of the tracking result of the filter used. The higher the observed noise matrix value, the tracker will be more sensitive to the observation result, whereas if the matrix observation noise value is lower, the tracker will be very responsive to the result of the estimation of the filter track [24].

Vehicle Counting

Vehicle counting consists of two approaches, id based counting and single line counting [2,11]. The idea of id based counting is each object has a different id with another object, so every object counts as some vehicles [1]. While the idea of single line counting is every vehicle has passed the single line in a particular frame will be counted as the number of counting vehicles [2]. A good accuracy performance for id based counting is each object must always be detected at any time in every frame. If a miss detection occurs, then the object id will be different from the previous id. This case will lead to miss counting depending on the number of ids formed. This will significantly affect the accuracy of performance in vehicle counting.

2. Methods

Dataset

We use open vehicle traffic dataset namely Amirkabir from Computer Vision Laboratory[1]. The dataset consists of 6 traffic videos such as noon, afternoon, sundown, rainy day, night, and rainy night.

Research methodology framework

The proposed research method is divided into three main processes, including vehicle detection, vehicle tracking, and vehicle counting. The proposed research methodology framework is shown in Figure 1. Dataset video input is processed into a set of image frames with size (480 x 640). Preprocessing in this paper uses ROI polygon with sizes at points [220.0], [50,280], [600,280] and [430,0].

The vehicle detection process uses optical flow density which has been developed by OpenCV library using Farneback Algorithm. The Farneback method uses the base quadratic polynomial assumption. OpenCV gets a 2-channel array with optical flow vectors, (u, v). Next step is to find the magnitude and direction of flow vectors. Direction corresponds to Hue value of the image. Magnitude corresponds to Value plane. Image binarization is used to get the blob of vehicle object. The object of vehicles then optimized using dilation operation to connect the parts of an unbounded object or to close the hole in the object. Detection of vehicles objects in this paper using the segmentation of Contours from blob detection in the ROI region. Each vehicle object detected then find a centroid value by utilizing the moment feature. The centroid is used as position input for vehicle tracking using Hungarian Kalman filter.

The representation of objects on the Kalman filter using multivariate Gaussian assumptions. Mean (centroid) represents the existence of each object. While covariance represents uncertainty. Process noise and observation noise represent noise in the system.

$$u = \begin{bmatrix} Cx \\ Cy \end{bmatrix} \quad (20)$$

The transition matrix represents the movement of an object by using the concept of linear displacement. The displacement is linearly represented by the movement of objects with fixed speed and minimal acceleration changes.

$$F = \begin{bmatrix} 1 & dt \\ dt & 1 \end{bmatrix} \quad (21)$$

Where dt is the time difference during the process of object displacement occurs. So based on the kalman filter equation, the state prediction from the previous input state is present in equation 22:

$$u'_{\{k|k-1\}} = Fu'_{\{k-1|k-1\}} \quad (22)$$

$$\begin{bmatrix} Cx' \\ Cy' \end{bmatrix} = \begin{bmatrix} 1 & dt \\ dt & 1 \end{bmatrix} \begin{bmatrix} Cx \\ Cy \end{bmatrix} \quad (23)$$

Uncertainty predictions on the filter calm can be calculated using the equation 24 and 25:

$$P_{\{k|k-1\}} = FP_{\{k-1|k-1\}} F.T + Q \quad (24)$$

[1] http://cvlab.aut.ac.ir/old/node/28.html

TABLE 3
VEHICLE COUNTING RESULT

Vehicle Counting	*Day Time*				*Night Time*		*Average (%)*
	Noon	*Afternoon*	*Sundown*	*Rainy*	*Night*	*Rainy*	
Actual Number of Vehicles	*228*	*413*	*261*	*128*	*277*	*247*	
Percentage of correctly counted vehicles (%)							
Rashid et al. [12]	86.8	48.4	**82.8**	50.0	30.7	43.3	57.0
Shiva and Reza [2]	90.5	48.7	68.2	51.6	66.8	46.6	62.1
OFD – HKF	**97.8**	**82.6**	**82.8**	**99.2**	**82.7**	**83.8**	**88.2**
Percentage of falsely counted vehicles (%)							
Rashid et al. [12]	182.9	146.5	84.7	84.4	141.2	72.1	118.6
Shiva and Reza [2]	7.9	**2.9**	**4.2**	4.7	21.7	**1.6**	**7.2**
OFD – HKF	**3.9**	13	6.1	**3.1**	**15.5**	18.2	10

$$\begin{bmatrix} Px' \\ Py' \end{bmatrix} = \begin{bmatrix} 1 & dt \\ dt & 1 \end{bmatrix} \begin{bmatrix} Px \\ Py \end{bmatrix} \begin{bmatrix} 1 & dt \\ dt & 1 \end{bmatrix} dt + \begin{bmatrix} Qx \\ Qy \end{bmatrix} \quad (25)$$

Q represents the noise process caused by the system's ability to detect blobs and camera quality capabilities. The prediction result of mean and covariance matrix is then corrected to update the position of the object by using observation result and Kalman gain. Kalman gain can be written mathematically with the equation 26

$$K_{\{k\}} = P_{\{k|k-1\}} A.T(C.Inv) \quad (26)$$

Where A is a matrix that links the observation with the predicted result and C represents pre-fit residual covariance that can provide uncertainty information on observation noise.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (27)$$

$$C = AP_{\{k|k-1\}} A.T + R \quad (28)$$

The equation $b_{\{k\}} - Au'_{\{k|k-1\}} = B$ is a residual pre-fit measurement that can estimate and calculate the difference values between the actual measurements with the best predicted results available under the previous system model and measurement. The final result of updating the object position and uncertainty on the system can then be calculated using the equation 29 and 30.

$$u'_{\{k|k\}} = u'_{\{k|k-1\}} + K_{\{k\}} B \quad (29)$$

$$P_{\{k|k\}} = P_{\{k|k-1\}} - K_{\{k\}}(CK.T) \quad (30)$$

The Hungarian algorithm is an algorithm used for assignment of two pairs of matrix inputs. Matrix input in this research is the distance between the object positions obtained from detection and prediction. The Hungarian Algorithm aims to create an optimal mapping for each component of the observed object [20]. The distance measurement parameter of the Hungarian algorithm used in this paper is a Euclidian distance such as equation 31:

$$\mathrm{d} = \sqrt{(\mathrm{C_{xd}} - \mathrm{C_{xkf}})^2 + (\mathrm{C_{yd}} - \mathrm{C_{ykf}})^2} \quad (31)$$

Where d is the distance, $\mathrm{C_{xd}}$ and $\mathrm{C_{yd}}$ are centroid values on the x and y-axes of detection using OFD, $\mathrm{C_{xkf}}$ and $\mathrm{C_{ykf}}$ are centroid values on the x and y-axes of detection using KF.

Vehicles counting becomes one of the critical processes. Because the detection based counting has some weaknesses due to double and miscounting. This research uses single line counting which is placed at one specified frame location, 5/12 from frame width. In [2], counting line located according to the best results through trial and error. Changing the location of the counting line will affect the configuration of the development ITS system. In the video dataset, the occurrence of miss detection on a particular frame can cause a miscounting.

The measurement performances

To Evaluate our methods, we use six evaluation measurement performances. These are Percentage of correctly detected vehicles, falsely detected vehicles, correctly tracked vehicles, misses tracked vehicles, correctly counted vehicles, and falsely counted vehicles. Percentage of correctly detected vehicles is the ratio of the number of vehicles

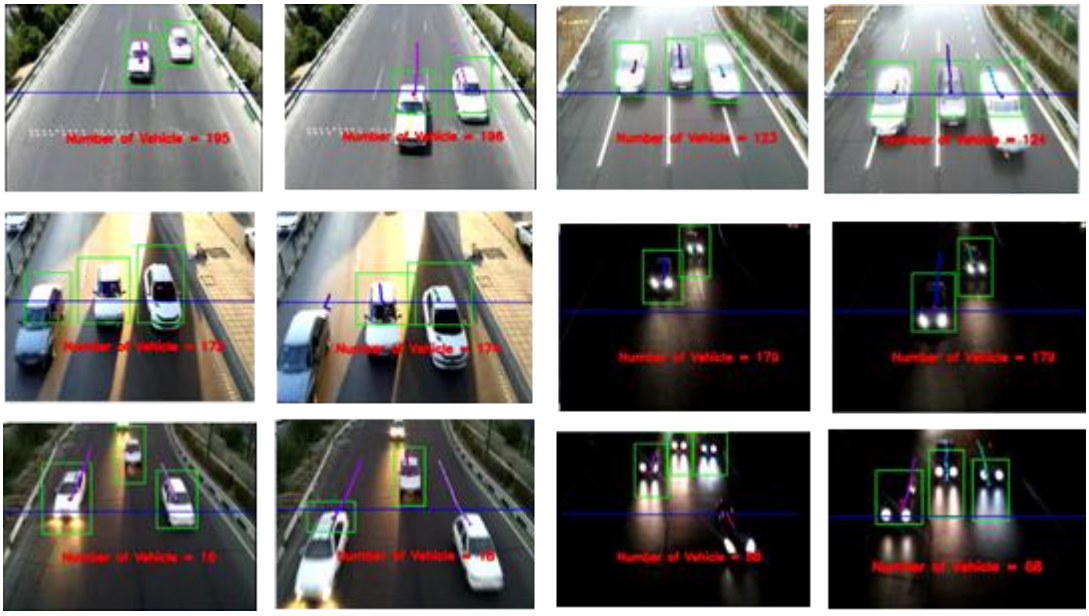


Figure 2. Visualization Result of Vehicle Detection, Tracking, and Counting for Noon, Afternoon, Sundown, Rainy Day, Night, and Rainy Night

detected and the number of vehicles matching ground truth multiplied by 100%; Percentage of falsely detected vehicles is the ratio of the number non-vehicle detected and the number of vehicles matching ground truth multiplied by 100%; Percentage of correctly tracked vehicles is the ratio of the number of successfully tracked vehicles and the number of vehicles matching ground truth multiplied by 100%; Percentage of misses tracked vehicles is the ratio of the number of untracked vehicles and the number of vehicles matching ground truth multiplied by 100%; Percentage of correctly counted vehicles is the ratio of the number of counted vehicles and the number of vehicles matching ground truth multiplied by 100%; and Percentage of falsely counted vehicles is the ratio of non-vehicle counted and the number of vehicles matching ground truth multiplied by 100%.

This research was tested on the Amirkabir dataset and compared with previous related studies [12] and [2]. The implementation algorithm conducted on laptop with specification Intel Core I5-6200U CPU @ 2.3 GHz 2.4 GHz with total RAM of 4 GB

3. Results and Analysis

Vehicle detection is the first step in research related to the development of intelligent traffic system (ITS). Further vehicle detection results can be utilized for tracking and counting the number of vehicles. Vehicle detection is the primary foundation for the development of high-quality ITS. Detection of vehicles with good results will make ease to the next process. Therefore, vehicle detection becomes very important for developing the ITS. The dataset used consists of six data sets that have different characteristics. Each character has its challenges that must be addressed to avoid errors during object detection. Possible errors in object detection are miss detection and false detection. Miss detection object detection can occur due to differences in colour intensity between objects with the background while false detection can occur due to the movement of other objects such as shadows and emission of light from moving vehicles. Tables 1 shows the results of vehicle detection using Optical flow density method.

Table 1 shows the results of vehicle detection performance using Optical Flow Density (OFD) method compared with two previous research [12] and [2]. From table 1, it is known that the vehicle detection using optical flow density methods, implemented by the authors have an average accuracy 93.6%. Vehicle detection using optical flow density has a better result than [12]. [12] using TSI-VDL to detect the vehicles. Besides that, the proposed method shows the approach research to Shiva and Reza [2]. [2] use the ABM-SC method to detect the vehicles. The result of percentage false rate value shows the better result than [12] and [2]. The average value of false percentage rate is 1.2%. Vehicle verification use to minimize the false rate from this research. The parameters using the minimum contour area, length, and width of each bounding box size.

OFD has been implemented utilizing the features of gradient image. Vehicles detection using OFD in this paper was conducted to improve the performance of previous research, i.e. TSI-VDL [12] and ABM [2]. The implemented OFD method does not take much time to recognize the vehicles on the video dataset, compared with the TSI-VDL and ABM-SC methods. The TSI-VDL is only capable of detecting vehicles with an average accuracy performance only 57.0%.

Especially for night and afternoon datasets, the TSI-VDL [12] does not even achieve 50%. This is due to the movement of vehicle lights and shadow of vehicles. Also, the error rates using TSI-VDL method is also too high up to 118.6% on average. In table 1 the average accuracy performance of [2] reaching 98.2%. The vehicles detection result was obtained by taking five video frames from each dataset. Vehicles will be counted into detected vehicles if at least one frame detects the vehicle. This approach is not capable of detecting the vehicle at any frame.

The consistency of vehicle detection for each frame is one of the essential factors to be considered. This will be seen in the tracking and counting result. Regarding this issue, we use motion based appearance to detect the vehicle. So that the moving vehicle can be detected for every frame. Although the result is not as a good as ABM-SC. Vehicles are tracking in this paper using Hungarian Kalman filter algorithm. Hungarian Kalman filter is an algorithm used to track the object with the input of the previous object position.

Table 2 shows the results of vehicle tracking performance. The results of vehicle tracking performance were compared using two previous studies [12] and [2]. The Hungarian Kalman filter implemented in this paper performs better result than the previous methods. The average accuracy is 88.2%.

The Hungarian Kalman filter works well in the case of multi-object tracking. It can predict the position of objects in the future and associated data on each vehicle object. Tracking results using distance similarity measurement conducted by [2] have an average accuracy of 61.2%. The difference between vehicle detection and vehicle tracking performance showed the significant decrease from 98.2% to 61.2%.

In [12], the number of tracked vehicles is the number of detected vehicles. In this research, vehicle tracking and counting are determined by the particular id during vehicle detection. The performance of vehicles tracking is highly dependent on vehicle detection performance results. The accuracy performance from vehicles tracking only 57% on average. The average performance of miss tracked using Hungarian Kalman filter method is 11.8%. This result is the best result compared to [2] and [12]. While the average performance of miss tracked by using the id detection based [12] and distance similarity measurement [2] reached 43% and 38% in average respectively. Miss tracked due to the occlusion problem so that adjacent vehicles are detected and tracked as one vehicle only. Moreover, the configuration of determining time and covariance matrix variables on the Kalman filter can also affect the results of object tracking. Time variable in this paper set as 0,025.

Vehicles are counting in this paper using single line counting (SLC) algorithm. SCL used to count the number of vehicles which have a small error rate [2]. Table 3 shows the results of vehicle counting. Based on table 3, the proposed framework is the best method for vehicles counting. The developed framework has several advantages over the comparator. In [12], The accuracy performance obtained from this method only reached 57%. Some vehicles are mostly undetected. Same as vehicle tracking, due to double detection problems, so the error rate for vehicles is counting up to 118.6%. In [2], the accuracy performance below 70% in average, except for the Noon dataset. Although the location of the counting line was arranged based on the best trial and error. Visualization results for multi-object detection and tracking using OFD - HKF for vehicle counting shown in figure 2. Object centroid, counting line, and detected vehicles are visualized by using a red dot, blue line, and green bounding box, respectively. Trajectory tracking and the number of counted vehicles visualized using RGB and text “Number of Vehicles = n” in red colour randomly. Where *“n”* represents the number of counted vehicles.

4. Conclusion

In this paper, the proposed framework has been able to detect, track, and count the number of vehicles on the video dataset. Vehicles detection, tracking, and counting using optical flow density, Hungarian Kalman filter, and Single line counting. Those methods combine to perform the fundamental process of building the integrated traffic systems. The proposed framework has been successfully tested despite the appearance of a shadow, illumination change, challenging weather, motion blur and dynamic background in the video traffic dataset. The average accuracy of each process detection, tracking, and vehicle counting was 93.6%, 88.2%, and 88.2% respectively. While

the average value of a false rate for each process detection, tracking, and vehicle counting is 1.2%, 11.8% and 10.0% respectively.

References

[1] N. Kosaka and G. Ohashi, "Vision-Based Nighttime Vehicle Detection Using CenSurE and SVM," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 5, pp. 2599–2608, 2015.

[2] S. Kamkar and R. Safabakhsh, "Vehicle detection, counting and classification in various conditions," IET Intell. Transp. Syst., vol. 10, no. 6, pp. 406–413, 2016.

[3] T. Chen and S. Lu, "Robust Vehicle Detection and Viewpoint Estimation with Soft Discriminative Mixture Model," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 2, pp. 394–403, 2017.

[4] M. Rezaei, M. Terauchi, and R. Klette, "Under Challenging Lighting Conditions," vol. 16, no. 5, pp. 2723–2743, 2015.

[5] Q. Zou, H. Ling, S. Luo, Y. Huang, and M. Tian, "Robust Nighttime Vehicle Detection by Tracking," vol. 16, no. 5, pp. 2838–2849, 2015.

[6] M. Nieto, G. Vélez, O. Otaegui, S. Gaines, and G. Van Cutsem, "Optimising computer vision based ADAS: a vehicle detection case study," IET Intell. Transp. Syst., vol. 10, no. 3, pp. 157–164, 2016.

[7] X. Wang, L. Xu, H. Sun, J. Xin, and N. Zheng, "On-Road Vehicle Detection and Tracking Using MMW Radar and Monovision Fusion," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 7, 2016.

[8] J. M. Guo, C. H. Hsia, K. S. Wong, J. Y. Wu, Y. T. Wu, and N. J. Wang, "Nighttime Vehicle Lamp Detection and Tracking with Adaptive Mask Training," IEEE Trans. Veh. Technol., vol. 65, no. 6, 2016.

[9] H. T. Chen, Y. C. Wu, and C. C. Hsu, "Daytime Preceding Vehicle Brake Light Detection Using Monocular Vision," IEEE Sens. J., vol. 16, no. 1, pp. 120–131, 2016.

[10] S. Noh, D. Shim, and M. Jeon, "Adaptive Sliding-Window Strategy for Vehicle Detection in Highway Environments," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 2, pp. 323–335, 2016.

[11] Yi Wu and Jongwoo Lim and Ming-Hsuan Yang, "Online Object Tracking: A Benchmark" IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013

[12] N. U. Rashid, N. C. Mithun, B. R. Joy, and S. M. M. Rahman, "Detection and classification of vehicles from a video using a time-spatial image," ICECE 2010 - 6th Int. Conf. Electr. Comput. Eng., no. December, pp. 502–505, 2010.

[13] Indrabayu, R. Y. Bakti, I. S. Areni, and A. A. Prayogi, "Vehicle detection and tracking using Gaussian Mixture Model and Kalman Filter," Proc. - Cybern. 2016 Int. Conf. Comput. Intell. Cybern., 2017.

[14] A. Agarwal, S. Gupta, and D. K. Singh, "Review of optical flow technique for moving object detection," Proc. 2016 2nd Int. Conf. Contemp. Comput. Informatics, IC3I 2016, pp. 409–413, 2016.

[15] S. Aslani and H. Mahdavi-nasab, "Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance," vol. 7, no. 9, pp. 761–765, 2013.

[16] D. X. Zhou and H. Zhang, "Modified GMM Background Modeling and Optical Flow for Detection of Moving Objects," 2005 IEEE Int. Conf. Syst. Man Cybern., vol. 3, no. December, pp. 2224–2229, 2005.

[17] K. Kale, S. Pawar, and P. Dhulekar, "Moving Object Tracking using Optical Flow and Motion Vector Estimation," pp. 2–7, 2015.

[18] E. Engineering, I. I. T. B. Mumbai, and I. I. T. B. Mumbai, "Dynamic Arbitrary Shape Boundary Detection and Tracking Using Hungarian Kalman Filter," no. Boundary Detection, pp. 243–249, 2016.

[19] Z. Meng, Z. U. O. Yan, F. Y. Yu, and L. I. M. Di, "Sensor Assignment Method Based on Time-Varying Measurement Variance for Tracking Multi-targets," no. 1, pp. 3368–3372, 2016.

[20] B. Sahbani and W. Adiprawita, "Kalman Filter and Iterative-Hungarian Algorithm Implementation for Low Complexity Point Tracking as Part of Fast Multiple Object Tracking System," 6th Int. Conf. Syst. Eng. Technol., pp. 109–115, 2016.

[21] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in the low-exposure colour video for night conditions," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 453–462, 2010.

[22] R. O'Malley, M. Glavin, and E. Jones, "Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions," IET Intell. Transp. Syst., vol. 5, no. 1, pp. 1–10, 2011.

[23] H. Kaur, "Vehicle Tracking using Fractional order Kalman Filter for Non-Linear System," Int. Conf. Comput. Commun. Autom. Veh., pp. 474–479, 2015.

[24] K. M. Jeong and B. C. Song, “Night time vehicle detection using rear-lamp intensity,” 2016 IEEE Int. Conf. Consum. Electron. ICCE-Asia 2016, pp. 9–11, 2017

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 27-33
DOI: http://dx.doi.org/10.21609/jiki.v11i1.549

EEG CLASSIFICATION FOR EPILEPSY BASED ON WAVELET PACKET DECOMPOSITION AND RANDOM FOREST

Yuna Sugianela, Qonita Luthfia Sutino, and Darlis Herumurti

Department of Informatics, Faculty of Information and Communication Technology, Sepuluh Nopember Institute of Technology, Surabaya, 60111, Indonesia

nelaneliyuna@gmail.com, qonitaluthfia@gmail.com, darlis@if.its.ac.id

Abstract

EEG (electroencephalogram) can detect epileptic seizures by neurophysiologists in clinical practice with visually scan long recordings. Epilepsy seizure is a condition of brain disorder with chronic noncommunicable that affects people of all ages. The challenge of study is how to develop a method for signal processing that extract the subtle information of EEG and use it for automating the detection of epileptic with high accuration, so we can use it for monitoring and treatment the epileptic patient. In this study we developed a method to classify the EEG signal based on Wavelet Packet Decomposition that decompose the EEG signal and Random Forest for seizure detection. The result of study shows that Random Forest classification has the best performance than KNN, ANN, and SVM. The best combination of statisctical features is standard deviation, maximum and minimum value, and bandpower. WPD is has best decomposition in 5th level.

Keywords: *EEG, epilepsy, seizure, wavelet, random forest*

Abstrak

EEG (electroencephalogram) dapat mendeteksi serangan epilepsi oleh ahli neurofisiologi dalam praktik klinis dengan memindai rekaman secara visual. Kejang epilepsi adalah kondisi gangguan otak kronis yang tidak dapat berkomunikasi, menyerang orang dari segala umur. Tantangan penelitian yang dapat dilakukan adalah bagaimana mengembangkan metode untuk pemrosesan sinyal yang mengekstrak informasi EEG halus dan menggunakannya untuk mengotomatisasi deteksi epilepsi dengan akurasi tinggi, sehingga kita dapat menggunakannya untuk memantau dan mengobati pasien epilepsi. Dalam penelitian ini kami mengembangkan sebuah metode untuk mengklasifikasikan sinyal EEG berdasarkan Dekomposisi Packet Wavelet yang menguraikan sinyal EEG dan Random Forest untuk deteksi kejang. Hasil penelitian menunjukkan bahwa performa Random Forest lebih baik dibandingakna dengan metode klasifikasi. Kombinasi fitur statistika terbaik adalah standar deviasi, nilai maksimum dan minimum, serta bandpower. WPD memiliki dekomposisi terbaik pada level 5.

Kata Kunci: *EEG, epilepsi, kejang, wavelet, random forest*

1. Introduction

Epilepsy seizure is a condition of brain disorder with chronic noncommunicable that affects people of all ages. The intricate chemical changes in brain nerve cells lead to sudden activity of electric current and magnetic fields during seizures [1]. EEG (electroencephalogram) can detect epileptic seizures by neurophysiologists in clinical practice with visually scan long recordings [2].

There are many various of EEG analysis and classification methods use the fact that the processing of information in the EEG signal as dynamical changes of the electrical activity [3]. The challenge of study is how to develop a method for signal processing that extract the subtle information of EEG and use it for automating the detection of epileptic with high accuration, so we can use it for monitoring and treatment the epileptic patient.

In this study we developed a method to classify the EEG signal based on Wavelet Packet Decomposition that decompose the EEG signal and Random Forest for seizure detection. There are many research about EEG analysis and classification, they will be presented in section 2. In section 3 is the explanation of proposed method based on Wavelet Packet Decomposition, how to select the features and Random Forest for classification data. The results and discussion of research are presented in section 4, and finally the section 5 is conclussion of our study.

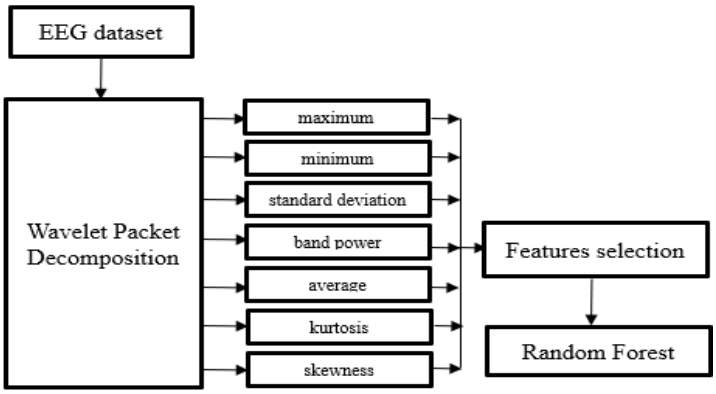


Figure 1. Block diagram of proposed method

Related work

Some research have proposed epilepsy diagnosis by using wavelet transformation (WT) for processing signal and feature extraction. Compared to Fourier Transform, WT has smoother representation, because it captures changes of EEG signals in details [3].

Discrete Wavelet Transformation (DWT) is one of wavelet analysis used in common. DWT processes the EEG signal into approximations (low frequency) in every levels and detail coefficients (high frequency) only in first level. The output of this analysis is wavelet coefficients. Xie and Krishnan [11] proposed research to detect seizure and diagnose epilepsy using DWT for EEG signal feature extraction with Haar wavelet.

The performance in detecting epilepsy has been evaluated using Wavelet Packet Decomposition (WPD), DWT, and Empirical Mode Decomposition (EMD) [4]. DWT and WPD are wavelet based methods which the difference with EMD is EMD has no decrease in the number of features. Because WPD has developed DWT capabilities which WPD decomposes signal into both approximations and detail coefficients, it results better frequency resolution. It has been proved that overall accuracy in three classes case using Random Forest classifier and WPD to process signal results 99.66% where using DWT and EMD to process signal results 98.4% and 90.4% [4]. For other dataset classification that has more than 3 classes case, WPD has lower accuracy than DWT [5].

2. Methods

The proposed method in this study is the classification of EEG signal based on Random Forest method, and the decomposition of signal EEG is based on Wavelet Packet Decomposition.

We select the statistical features from the coefficent's result in WPD. Figure 1 shows the block diagram of proposed method.

Dataset

The dataset of EEG was downloaded from The Epileptologie, Univertat Bonn. The dataset consists of 5 EEG records set, those are A, B, C, D, and E. Each EEG dataset contains 100 single-channel brain, they were taken from different people recording process of dataset A and B were taken from healthy volunteers but with different conditions. Dataset A was recorded with open eyes but B with close eyes. Dataset C, D, and E were taken from epileptic patients. Same with A and B, the recording process of dataset C, D, and E were taken from different conditions. Dataset C was recorded from epileptic patients with no seizure and open eyes, but D with close eyes. Dataset E were taken from seizure epileptic patients.

As mentioned before, we can conclude that there are 200 healthy people's brain signal data, 200 epileptics with no seizure patient's brain signal data, and 100 seizure epileptic patient's brain signal data [6].

Mother Wavelet

To get information from EEG signals, there are tools called wavelets. Part of wavelets that is called Mother Wavelet has been used to extract frequency and time information smoother [12]. Daubechies is one of orthogonal wavelet families that is able to get optimal set of EEG signal [13]. Some researches have been found using DWT based on Daubechies 4 (db4) as mother wavelet. They show that db4 is commonly suitable for detecting epilepsy case [3].

TABLE 1
STATISTIC FEATURES COMBINATION

No	Statistical Features			
	1	2	3	4
1	Mean	Max	Min	Bp
2	Mean	Max	Min	Krt
3	Max	Min	Bp	Krt
4	Skw	Max	Min	Bp
5	Std	Max	Min	Bp
6	Mean	Bp	Krt	Skw
7	Max	Bp	Krt	Skw
8	Min	Bp	Krt	Skw
9	Std	Bp	Krt	Skw
10	Mean	Krt	Skw	Std
11	Max	Krt	Skw	Std
12	Min	Krt	Skw	Std
13	Min	Max	Mean	Std
14	Max	Min	Bp	Krt

TABLE 2
COMPARISON DWT AND WPD WITH DB4

Classi fi-cation Meth od	Decomposition Level					
	3		4		5	
	DWT	WPD	DWT	WPD	DWT	WPD
KNN	**88.91 %**	86.37 %	**86.46 %**	85.61 %	**86.87 %**	85.39 %
ANN	**91.58 %**	91.50 %	**91.46 %**	91.37 %	**92.97 %**	91.96 %
SVM	73.31 %	**79.13 %**	77.73 %	**85.39 %**	81.47 %	**90.33 %**
RF	96.83 %	**97.23 %**	96.49 %	**97.50 %**	97.34 %	**98.11 %**

Wavelet packet decompostion

The wavelet packet decomposition (WPD) extends the capabilities of the WD (wavelet decomposition) and DWT (discrete wavelet transform). Whereas DWT decomposes the approximations records only, WPD does the decomposition of both approximation and detail records into sublevels. WD only partitions the frequency axis finely toward low frequency, and WPD is a generalized version, which also decomposes the high frequency bands that are kept intact in wavelet decomposition [4].

WPD applies a complete wavelet package tree, it delivers better frequency resolution for the signal being decomposed. Another benefit of the WPD is that it represents the reconstruction of the original signal by combining various decomposition level [4]. In this study, daubechies4 (db4) mother wavelet function is used as previous research with DWT for detecting epilepsy .

Feature selection

Wavelet packet decomposition can be used for denoising and feature extraction [3]. In this study, we select seven different statistical features for EEG classification, aiming at decreasing the dimensionality of dataset. The signal statistics are used in order to capture important information while keeping the low data dimensions. The seven statistical features are mean, standard deviation (Std), minimum value (Min), maximum value (Max), bandpower (Bp), skewness (Skw), and kurtosis (Krt).

In this experiment, we combine statistic features that have been mentioned. The number of features each combination used is 4 statistical features. There are 14 experimental combinations of statistical features performed.

The purpose of combining statistic features is to find the best accuracy between combinations made. Beside statistic features, we also evaluate the level of decomposition based on WPD. The number of sub-bands every level is 2^k, where k is

TABLE 3
EVALUATION OF WPD'S DECOMPOSITION LEVEL

Statistical Features	RF Decomposition Level			Best Acc
	3	4	5	
1	96.60%	97.40%	98%	5
2	95%	95.80%	97.20%	5
3	97.20%	97.20%	97.80%	5
4	97.40%	97%	97.80%	5
5	97.80%	98.20%	98.40%	5
6	97.20%	97.20%	98.20%	5
7	97.60%	97.40%	98.20%	5
8	96.60%	97.40%	98%	5
9	98.60%	98%	98.60%	3 and 5
10	97.80%	97.80%	98.40%	5
11	97.80%	98%	98.20%	5
12	97%	98.20%	98.20%	4 and 5
13	97.20%	97.80%	98.20%	5
14	97.40%	97.20%	98%	5

TABLE 4
EVALUATION OF STATISTICAL FEATURES

Statistical Features	Decomposition Level		
	3	4	5
1	87.55%	90.00%	93.65%
2	89.45%	90.65%	92.65%
3	91.20%	92.25%	93.75%
4	88.55%	90.55%	89.27%
5	**93.25%**	**94.15%**	**96.40%**
6	79.65%	81.10%	81.90%
7	87.45%	88.25%	89.30%
8	86.75%	88.25%	88.80%
9	89.90%	91.15%	90.75%
10	87.85%	89.80%	90.85%
11	87.85%	89.80%	90.85%
12	90.25%	91.35%	92.55%
13	90.55%	92.25%	94.60%
14	86.45%	88.10%	91.25%

number of level. In level 3 decomposition, there are $2^1 + 2^2 + 2^3 = 14$ sub-bands. So, there are 15×4 statistic features $= 60$ features in level 3. In the level 4 are $(14 + 2^4) \times 4$ statistic features $=$ 120 features and in level 5 there are $(30 + 2^5) \times 4$ statistic features $= 248$ features.

EEG classification based on random forest (RF)

Random forest (RF) are an ensemble-based learning technique for classification [7], consists of many individual classification trees, where each tree is a classifier by itself that is given a certain weight for its classification output. The classification outputs from all trees is used to determine the overall classification output is done by choosing the mode of all trees classification [8].

In the case of classification, a large number of classification trees is generated, whereby each tree is asigned an input vector sampled independently from the same distribution using bootstrap samples. In the present case, one third of the observations (out-of-bag (OOB) set) were used for cross-validation. The OOB data are also used to obtain estimates of feature importance. Specifically, during the forest building process, an internal unbiased estimate of the generalization error (OOB error) is generated and used to identify the most important features. The final OOB prediction is the average score achieved from the

TABLE 5
COMPARISON WITH OTHER CLASSIFICATION METHODS

Classification Method	Decomposition Level					
	3		4		5	
	Acc. Avg (%)	Run. Time Avg (s)	Acc. Avg (%)	Run. Time Avg (s)	Acc. Avg (%)	Run. Time Avg (s)
KNN	86.37	6.21	85.61	35.92	85.39	110.31
ANN	91.50	0.00	91.37	0.00	91.96	0.00
SVM	79.13	0.12	85.39	0.13	90.33	0.16
RF	**97.23**	0.02	**97.50**	0.03	**98.11**	0.07

majority vote within the forest, excluding trees that included this observation during their training phase. The features that yield large score values are ranked as more important. The subset of the most important features was selected by sequential RF classification, i.e. by using labelled data to remove the least important feature in each run. By doing this, the minimum number of features required to achieve good classification accuracy was eventually selected [9].

Performance evaluation

In order to evaluate the proposed method, we use 10-fold cross-validation. This k-fold (in this study k=10) technique is implemented to create the training set and testing set for evaluation. With this technique, the feature vector set is divided into 10 subsets of equal size [10]. Of the 10 subsets, a single subset is retained as the validation data for testing the model and the remaining (k-1 or 9 in this study) subsets are used as training data. Then, the cross-validation process is repeated 10 times (the folds), with each of the 10 subsets used exactly once as the validation data. The average accuracy across all 10 trials is computed for consideration. The accuracy (Acc) are defined as:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (1)$$

TP (true positive) was determined when the epilepsy signals classfication as epilepsy signal. FP (false positive) was determined when the normal signals are classification as epileptic signals. TN (true negative) was determined when the normal signals are classified as an epilepsy signals. FN (false negative) was determined when the epilepsy signals are classified as normal signal as ilustrated in the confusion matrix [11].

3. Results and Analysis

Results

In this study we use the dataset for input in WPD process. We decomposed the signal of dataset and selected statistical features from each coefisiens. We compared the number of level in signal decomposition. In this experiment, we applied different combination of statistical features.

There were 7 statistical features that selected in this study. From 7 features, we combined them into 4 for each experiment. There were 14 experiments of statistical features. In this study we got the best performance of statistical features combination.

The statistical features were used for classification process based on Random Forest method. We also compare the performance of Random Forest with other classification methods, i.e. ANN (Artificial Neural Network), KNN (K-nearest Neighbor), and SVM (Support Vector Machine).

Comparison DWT and WPD decomposition with db4 mother wavelet

Daubechies4 (db4) mother wavelet function is used as previous research with DWT for detecting epilepsy. We also used db4 mother wavelet with WPD in this study. To compare the performa of DWT and WPD with db4, we try these decomposition methods to extract signal data and use them to classify data. We use statistical features from extracting data with the combination in Table 1.

The evaluation of classification use accuracy. In this comparison, we use the average value of accuracy from 14 statistical features combination. Table 2 shows that in KNN and ANN classification, DWT has better performance than WPD, but in SVM and Random Forest classification, WPD has better performance than DWT.

Evaluation of WPD's decomposition level

The number of decomposition level effects the number of features that are used for classification process. In this study we try 14 tests with different statistical features. From the experiment of decomposition level's number, we get the result that is shown in Table 3. The table presents that the

best performance of decomposition is in level 5. But in some test cases, we get the fact that there are 4 test cases that the accuration of 4th level is less that the 5th. Those test cases are combination of statistical features in number 4, 7, 9, and 14 in Table 1.

Evaluation of statistical features

In this study we combine 7 statistical features to get the best performance. Table 4 shows the average of 4 test cases based on 4 classification methods (ANN, KNN, SVM, and RF). The result shows that the best combination of statistical features for 3 level are same, number 5 (standard deviation, maximum value, minimum value, and bandpower).

Comparison with other classification methods

To eval the performance of classification proposed method, we compare with ANN, KNN, and SVM. From the average of 14 statiscical features combination shown in Table 5, we get the high accuracy average of level 3, 4, 5 are in RF classification methods. Furthermore, based on running time average, level 5 shows more significant increase than level 4 in KNN, SVM, and RF classification methods. It indicates level 3 and 4 using RF classification method result the better performance.

4. Conclusion

The proposed method to classify EEG signal for epilepsy disease is based on Wavelet Packet Decomposition (WPD) and Random Forest (RF) classifier. We used single channel EEG signal to analysis. To evaluate the performance of proposed method, we use 10-fold cross validation and compute the accuration of classification's result.

In this study we analyze the decomposition level of WPD and the result is 5th level has the highest average of accuration which is 91.44%. On the other hand, according to the average of running time, result of 3rd level has more efficient running time average which is 1.60 s. It shows that the less decomposition level of WPD results the the less time of running and the more decomposition level of WPD results the higher accuracy in several classification methods.

We extracted 7 statistical features from decomposition process based on WPD. The best combination of statistical features for classification process is standard deviation, maximum value, minimum value, and bandpower. The combination has 93.25% accuration in 3th level, 94.15% in 4th, and 96.40% in 5th level. Classification that is based on RF has the better performance than ANN, KNN, and SVM. The accuration of RF is 97.23% in 3th level, 97.50% in 4th level, and 98.11% in 5th level.

References

[1] A. Zahra, N. Kanwal, S. Ehsan, and K. D. Mcdonald-maier, "Seizure detection from EEG signals using Multivariate Empirical Mode Decomposition," *Comput. Biol. Med.*, vol. 88, no. June, pp. 132–141, 2017.

[2] J. Malmivuo and R. Plonsey, "Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields."

[3] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure Eur. J. Epilepsy*, vol. 26, pp. 56–64, 2015.

[4] E. Alickovic, J. Kevric, and A. Subasi, "Biomedical Signal Processing and Control Performance evaluation of empirical mode decomposition , discrete wavelet transform , and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomed. Signal Process. Control*, vol. 39, pp. 94–102, 2018.

[5] S. Sunny, D. P. S, and K. P. Jacob, "A Comparative Study of Wavelet Based Feature Extraction Techniques in Recognizing Isolated Spoken Words."

[6] V. Bajaj and R. B. Pachori, "EEG Signal Classification using Empirical Mode Decomposition and Support Vector Machine," pp. 581–592, 2012.

[7] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.

[8] L. Fraiwan, K. Lweesy, N. Khasawneh, and H. Wenz, "Automated sleep stage identification system based on time – frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2011.

[9] M. N. Anastasiadou, M. Christodoulakis, E. S. Papathanasiou, S. S. Papacostas, and G. D. Mitsis, "Clinical Neurophysiology Unsupervised detection and removal of muscle artifacts from scalp EEG recordings using canonical correlation analysis , wavelets and random forests," *Clin. Neurophysiol.*, vol. 128, no. 9, pp. 1755–1769, 2017.

[10] Y. Zhang, Y. Chen, and N. V Chawla, "Neurocomputing Automated epileptic

seizure detection using improved correlation-based feature selection with random forest classifier," vol. 241, pp. 204–214, 2017.

[11] D. Rahmawati, U. C. N. R, and R. Sarno, "Classify Epilepsy and Normal Electroencephalogram (EEG) Signal Using Wavelet Transform and K-Nearest ... Classify Epilepsy and Normal Electroencephalogram (EEG) Signal Using Wavelet Transform and K- Nearest Neighbor," no. November, 2017.

[12] H. A. Akkar and F. A. Jasim, "Optimal Mother Wavelet Function for EEG Signal Analyze Based on Packet Wavelet Transform," vol. 8, pp. 1222–1227, 2017.

[13] J. Rafiee, M. Rafiee, N. Prause, M. Schoen. (2011), "Wavelet basis functions in biomedical signal Processing," Expert Syst. Appl.vol. (38), pp. 6190–6201.

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 34-41
DOI: http://dx.doi.org/10.21609/jiki.v11i1.478

DETECTING CONTROVERSIAL ARTICLES ON CITIZEN JOURNALISM

Alfan Farizki Wicaksono, Sharon Raissa Herdiyana, and Mirna Adriani

Information Retrieval Lab., Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Email: alfan@cs.ui.ac.id

Abstract

Someone's understanding and stance on a particular controversial topic can be influenced by daily news or articles he consume everyday. Unfortunately, readers usually do not realize that they are reading controversial articles. In this paper, we address the problem of automatically detecting controversial article from citizen journalism media. To solve the problem, we employ a supervised machine learning approach with several hand-crafted features that exploits linguistic information, meta-data of an article, structural information in the commentary section, and sentiment expressed inside the body of an article. The experimental results shows that our proposed method manages to perform the addressed task effectively. The best performance so far is achieved when we use all proposed feature with Logistic Regression as our model (82.89% in terms of accuracy). Moreover, we found that information from commentary section (structural features) contributes most to the classification task.

Keywords: *controversy detection, text classification, supervised learning*

Abstrak

Pendirian dan pemahaman seseorang terhadap suatu topik kontroversial dipengaruhi oleh sumber berita yang dikonsumsinya. Namun, pembaca seringkali tidak menyadari bahwa ia sedang membaca sebuah artikel yang kontroversial. Padahal, dengan mengetahui bahwa sebuah artikel bersifat kontroversial, pembaca dapat lebih kritis dalam menerima informasi yang disampaikan di artikel tersebut. Penelitian ini bertujuan untuk mengembangkan sebuah model yang dapat secara otomatis mengklasifikasikan sebuah artikel jurnalisme warga berbahasa Indonesia sebagai kontroversial atau non-kontroversial. Digunakan metode berbasis *supervised learning* dengan dua model klasifikasi, yaitu *Logistic Regression* dan *Support Vector Machine*. Model dibangun dengan menggunakan empat kategori fitur, yaitu fitur metadata yang terdapat pada artikel, fitu struktural yang ada pada bagian komentar artikel, fitur linguistik, dan fitur yang mengeksploitasi informasi sentimen pada artikel. Hasil eksperimen menunjukkan bahwa model yang diusulkan oleh penelitian ini berhasil melakukan pendeteksian kontroversi dengan cukup efektif. Didapatkan akurasi terbaik sebesar 82,89% dengan menggunakan kombinasi semua fitur dan model *Logistic Regression*. Hasil eksperimen juga menunjukkan bahwa fitur struktural adalah fitur yang paling kontributif. Didapatkannya kombinasi semua fitur sebagai konfigurasi terbaik menandakan bahwa masalah pendeteksian kontroversi perlu didekati dari berbagai aspek.

Kata Kunci: *deteksi kontroversi, klasifikasi teks, pembelajaran mesin*

1. Introduction

The growing reach of Internet technology and the increasing of its usability has been able to bring the world mutually in a small country, where everyone is strongly connected to each other, just like one community. One can certainly mention incredible contributions of internet technology in many domains, such as education, research, public health, economics, entertainment, communication, journalism, etc. It is really clear how this technology has become one of the most important needs in our daily life.

In the area of journalism, Internet has provided many platforms that enables everyone to product and distribute reports on the interaction of events, facts, and ideas. We refer to this as citizen journalistic media, which is basically one of the forms of collaborative and social media. In the early stage of Internet, the one-way communication style of media has hindered citizen participation in terms of online journalistic activities. But, nowadays, the presence of social media, such as Weblogs, Microblogs, and Internet Forums has formed a new concept of communication, so called participatory journalism. Based on Bowman and Willis, [1], participatory journalism is defined as "the act of a citizen, or group of citizens, playing an active role in the process of collecting, reporting, analyzing and disseminating news and information". In addition to that, they also mentioned that "The intent of this participation is to provide indepen-

dent, reliable, accurate, wide-ranging and relevant information that a democracy requires." Differ from common professional journalism, participatory journalism requires no editorial oversight or formal journalistic workflow. Instead, it is the result of conversations in the online social media [1].

The opportunity to actively contribute in the participatory journalism has become great attractions for (non-professional) people. As a result, many online news websites have shifted towards facilitating a two-way communication platforms that enables citizen to share their journalistic writings within their websites. For example, one of the news agencies in the USA, CNN, has launched iReport 1. In Indonesia, there are several similar websites, such as Kompasiana 2 and Citizen 6 3.

Participatory journalistic media has several advantages as compared to common professional journalistic media, in the sense that the content of participatory journalistic media is actual and has more variation than common journalistic media. Unfortunately, it also has downsides since the creation process does not involve thorough editorial process. As a result, the accuracy of the content cannot be guaranteed. Moreover, the content has tendency to be biased, controversial, or provocative. This kind of readings can mislead many readers.

In our work, we focus on proposing a computational models to detect controversial articles due to its usefulness. Based on Merriam-Webster 4 dictionary, controversy is defined as "argument that involves many people who strongly disagree about something" or "strong disagreement about something among a large group of people". Controversial topics often involve many pros and cons around the topics. Wiley [2] mentioned that someone's understanding and stance on a particular controversial topic can be influenced by daily news or articles he consume everyday. Therefore, controversial topics should be carefully written. Unfortunately, readers usually do not realize that they are reading controversial articles. In addition to that, automatically recognizing controversial articles is not trivial. Knowing that a particular article is being controversial can help readers becoming more critical on information conveyed from the article.

One way to detect controversial articles is by looking at authorship debate occurred in the commentary section of the articles. However, when the total number of comments from an article reaches hundreds or thousands, it will be very difficult and tedious to manually read all comments and examine the quality of being controversial from the article. This problem motivates us to develop a computational model that can "automatically read" the article and its comments and determine its controversialness.

This paper is organized as follows. Section 2 describes related work on controversial article detection from the perspective of supervised approach. Then, section 3 presents the information of our annotated dataset for this task as well as our proposed approach. Section 4 discusses our experiment results. Finally, section 5 concludes our work and findings during the experiment.

Related Work

There have been several attempts in developing model for detecting controversial issues from several media, such as Wikipedia [3][4], News articles [5]–[7], and social media [8][9]. Popescu and Pennacchiotti [9] proposed a model for detecting controversial events from microblogs, like Twitter, related to some public figures in a fixed time period. They introduce the notion of twitter snapshot, i.e. a triple consisting of three concepts: target entity (e.g., Donald Trump), time period (e.g., one day), and a set of tweets talking about the target entity during the target time period. Their task is to assign a controversy score to each snapshot and rank the snapshots according to the controversy score. They argued that snapshot of controversial events provoke a public discussion, in which opposing opinions, surprise, or disbelief are easily found in the snapshot. Moreover, they coped with the problem using supervised machine learning models. Hence, several features were proposed to represent a snapshot, such as linguistic features, structural features, sentiment features, and "news buzz" features. They found that linguistic, structural, and sentiment features are highly ranked in terms of discriminative power.

Chimmalgi [6] focused on detecting controversial topics from social media such as comments and blogs, taking into account sentiment expressed in the comments, burstiness of comments, and controversy score. An annotated corpus consisting of 728 news articles was developed for training and evaluation purpose. Besides sentiment and structural features, Chimmalgi [6] also developed controversy-bearing term list from Wikipedia. Moreover, features derived from sentiment orientation score and controversy term list give the best discriminative power.

Allen et al., [10] conducted studies to detect disagree- ment in casual online forums such as Slashdot5. They presented a crowd-sourced annotated corpus for topic level disagreement detection. To develop the corpus, annotators were shown se-

1. http://ireport.cnn.com
2. http://www.kompasiana.com
3. http://citizen6.liputan6.com
4. http://www.meriam-webster.com/dictionary/controversy
5. https://slashdot.org/

veral topics and label them as containing disagreement or not. Furthermore, they found that disagreement detection is a subjective and difficult task since there are 22 topics (of 95 topics) has confidence scores below 55%. They formalized the problem as supervised learning using several hand-crafted features, such as rhetorical relations, sen-timent features, n-gram features, slashdot metafeatures, lexicon features, and structural features. Among those proposed features, the most discriminative features are those that include rhetorical information. Surprisingly, n-gram features harmed the model's performance.

Mejova et al. [7] studied the use of sentiment orientation information and biased words in 15 news portals. They showed empirical proof that controversial articles tend to reveal negative sentiment orientation and contain biased-words. Finally, Dori-Hacohen and Allan [11] proposed an automated approach to detect arbitrary webpages discussing controversial topics. Interestingly, they leveraged Wikipedia articles which bridge the gap between arbitrary webpages and rich metadata available in Wikipedia. They developed nearest neighbor classifier that maps webpages to the Wikipedia articles that discuss the same topic. Thus, the decision was solely based on those Wikipedia articles; if the Wikipedia articles are controversial, the corresponding webpages are also assumed to be controversial.

2. Methods

Data Collection

To build our dataset, we collected several articles from one of the famous participatory journalistic media in Indonesia, i.e. Kompasiana, during 9th - 10th April 2015. For our purpose, we selected those articles that have a considerable number of comments since analyzing comments is our main resource for feature extraction. In detail, first, we run a crawler to discover several popular topics. Moreover, we assumed that popular topics have been discussed in more than 100 articles. Second, for each popular topic, we run the second crawler to retrieve all related articles that have more than 60 comments. Next, after we collected a number of articles, we randomly selected around 500 articles from the collection for gold standard development. Finally, we manually annotated each of those articles as being controversy or non-controversy.

We obtained 304 articles as being controversy and 205 articles as non-controversy. Our controversial topics are mainly about political and law issue, such as Indonesian's presidential election in 2014 and Indonesian Corruption Eradication Commission. Table 1 shows the detail of our annotated corpus. We do not translate some terms since those terms are really specific to our domain.

TABLE 1
THE DETAIL OF OUR CONTROVERSY ARTICLE DATASET

Topic	Total	***Non-C**	***C**
pilpres 2014 (Presidential Election 2014)	203	96	107
pilkada jakarta (Jakarta Elecion)	47	31	16
KPK (Indonesian Corruption Eradication Commision)	37	22	15
pasangan capres (President Candidates)	29	12	17
jokowi capres 2014	22	8	14
jokowi nyapres	19	6	13
kawal pemilu	18	8	10
budi gunawan	15	9	6
debat capres (Election Debate) 2014	15	7	8
kompasiana baru	13	11	2

To create high quality corpus, we need to define annotation's guidelines, in which the definition of "being controversy" must be clear. Based on Indonesian's dictionary, controversy means something that sparks debate, while based on Merriam-Webster dictionary, controversy is defined as "argument that involves many people who strongly disagree about something". Hence, we formulated three questions that can help annotator to decide the label of an article: 1) Does the article form strong opinions toward a given issue or contain topics which are widely known as controversial (several topics are widely known as controversial, such as "gay", "atheism", "middle- east war", etc.), 2) Does the comment section of the article contain strong arguments or even aspersions, 3) Is the ratio between the number of pro's and con's comments of the article considerably balance.

Pre-Processing

Before we extracted the features of the articles, we followed several pre-processing steps: 1) we converted all the characters in the articles into their lowercased-version. 2) Explicit links were then removed from the article. 3) Non-canonical words were normalized into their canonical forms. For example, in Indonesian social media, we often see words, like "gak" and "nggak". Basically, they are non-canonical forms of the word "tidak", which means "not" in English. (4) Duplicate comments were then removed and considered as one distinct comment.

Proposed Methods and Features

We formulate our controversial article detection problem as a supervised classification problem us-

ing machine learning approach. Formally, given a set of label $L = \{\text{Controversy}, \text{NonControversy}\}$ and a set of articles $A = \{a_1, a_2, \ldots, a_n\}$, we seek a classifier function $F: A \rightarrow L$. Therefore, devising hand-crafted features is one of the most important steps in our task. In brief, we proposed 4 categories of features: META, STRUCTURAL, LINGUISTIC, and SENTIMENT features.

Meta Features (META)

This type of features leverage meta-data information found in the participatory journalistic media. There are two features belong to this category: the number of article views and the reader's evaluation.

1. The number of article views (META-1)
   Based on our observation, controversial article usually has a controversial title as well, such that many readers are interested in opening the article. As a result, the number of views upon a controversial article might have discriminative power in our classification task.
2. The reader's evaluation (META-2)
   In the participatory journalistic media, every reader can evaluate an article as being "inspirative", "interesting", "beneficial", or "actual". Therefore, an article is associated with 4 values denoting the number of votes of being "inspirative", "interesting", "beneficial", and "actual", respectively. We treat each value as a separate binary feature value.

Structural Features (STRUCTURAL)

This type of features captures the information regarding discussion activities that happened in the comment section of an article. Actually, the goal is that we need to know whether debate has occured in the comment section since debate is a good indicator for controversy.

1. Reply comments (STRUCTURAL-1)
   The first structural feature is the ratio between the number of reply comments and main comments. There are two types of comments, namely main comments and reply comments. Main comments directly response to the main article, while reply comments response to a particular main comment. We argue that when the number of reply comments is big enough, debates most likely occur in the comment section.
2. Distinct commentators (STRUCTURAL-2)
   We count the number of distinct commentators as our second structural feature. The rationale is that the more distinct commentators that a particular article has, the more intense the discussion among commentators.
3. Average cumber of comments (STRUCTURAL-3)
   This feature measures the activeness of a commentator. Someone who participate in a debate will most likely have more comments than those who don't.
4. Average length of comments (STRUCTURAL-4)
   Based on the observation conducted by Mishne and Glance [8], a comment that has good argument quality is quite lengthy. Therefore, we use the average length of comments as our structural feature. The length of a comment can be defined as the number of non-distinct words in the comment.
5. Maximum thread length (STRUCTURAL-5)
   Thread is identified as a single comment with its reply comments. Hence, thread length is defined as the number of reply comments in the thread itself. Based on our observation, comments that contain good arguments are usually replies to the a previous comment. In our work, we use the maximum thread length as our one of the structural features.

Linguistic Features (LINGUISTIC)

Based on our observation, we found that the controversialness of an article follows several linguistic patterns. Linguistic features mainly focuses on harnessing punctuations and lexical resource of bias words.

1. Question mark (LINGUISTIC-1)
   Question mark had been previously leveraged for the problem of controversial article detection [9][10]. Notice that in writing a sentence, question mark is frequently used to express curiosity or doubt about something. For example, in the following sentence, question mark can trigger debate on the commentary section.

   "oya, soal adipura, saya jamin 1000% hatta tahu. lho anak murid saya kelas 6 SD tahu semua. massa hatta enggak? jd nggak logis kalo hatta tdk tahu bedanya. massa anak SD lebih cerdas dari Hatta? nggak logis kan? hayolah akui itu." (Regarding adipura, i guarantee 1000% that Hatta knows this issue. All my elementary students know this issue very well, why not with Hatta? Are my elementary students smarter than Hatta? it really doesn't make sense, does it?)

   Specifically, we use the number of comments containing at least one question mark and the number of question marks in the main article body as our features.

2. Exclamation mark (LINGUISTIC-2)
   Just like question mark, exclamation mark was also used by Allen et al. [10] as one of their features for the same task. Exclamation mark usually indicates spirit and anger in the journalistic articles, in which this usually happens in a debate situation. We use the number of comments containing at least one exclamation mark and the number of exclamation marks in the main article body as our features.
3. Capital letters (LINGUISTIC-3)
   In a debate situation, writers usually express their arguments using capital letters to emphasize a certain issue or topic. We use the number of comments containing at least three words with capital letters. In addition, we use th-ree as our threshold number to compensate abbreviations, in which all letters are capitalized.
4. Bias lexicon (LINGUISTIC-4)
   Bias words tend to support a certain opinion, which means that they are not neutral. Furthermore, Mejova et al. (2014) found that controversial articles contain many bias words. To create Indonesian bias lexicon, first, we automatically translated 654 English bias words developed by Recansens et al. [4]. After that, we checked the translated bias word manually and discovered 602 bias words ready to use. The following list shows several words from our collection of 602 bias words. For main article body and commentary section, we use the proportion of bias words in the document as our feature.
5. Example of bias words
   aborsi (abortion), *fanatisme* (fanatism), *genosida* (genocide), *bom* (bomb), *homoseksual* (homosexual), *minoritas* (minority), *terorisme* (terorism), *skandal* (scandal), *sosialis* (socialism), *komunis* (communism), *muslim* (moslem), *katolik* (catholic), *rasis* (racism), *revolusi* (revolution), *yahudi* (jew), *zionis* (zionism), *propaganda* (propa-ganda)

Sentiment Features (SENTIMENT)

This type of features leverages subjectivity in the body and commentary section of an article. In our case, we hypothesize that a controversial article tends to reveal pro and contra about a particular topic. In addition, pro and contra toward a particular topic are usually expressed using opinionated words. As a result, we can employ several techniques for detecting sentiment orientation inside the articles and their comments. Furthermore, our techniques harness Indonesian sentiment lexicon developed by Vania [12], which contains 416 positive words and 581 negative words.

1. Sentiment score (SENTIMENT-1)
   Suppose Pos denotes the number of positive words, Neg denotes the number of negative words, and Neu denotes the number of neutral words in a document (body of articles or commentary section), we compute the sentiment score of the document as follow.

$$\text{sentiment} = \frac{\text{Pos} - \text{Neg}}{\text{Pos} + \text{Neg} + \text{Neu}} \quad (1)$$

   Equation(1) yields a value in range $[-1, +1]$, where value less than zero means that the overall document reveals negative sentiment, and vice versa. From this score, we devise three feature values: (1) sentiment score of the article body, (2) average sentiment scores of all comments (commentary section can have more than one comment), the difference between article body's sentiment score and average sentiment scores of all comments. The last feature is proposed since, based on our observation, controversialness of an article is often triggered by the opinion clashes or differences that happen between the content of article body and commentary section.

2. Standard deviation of sentiment score (SENTIMENT-2)
   This feature is based on rationale that the more controversial an article is, the more various the sentiment expressed inside the commentary section. The variance of all sentiment scores inside the comments can be captured using the following standard deviation formula.

$$\text{std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - \mu_s)^2} \quad (2)$$

   where N is the number of comments, S_i is the sentiment score of i-th comment, and μ_s is the average of all comments' sentiment scores.

3. Mixed sentiment score (SENTIMENT-3)
   We also use mixed sentiment score proposed by Popescu and Pennacchiotti [9]. In their original paper, they use this scoring mechanism to detect controversialness on microblogs, such as tweets. In our case, instead of tweets, we apply this scoring formula to all comments.

$$\text{mixSentiment} = \frac{\min(\text{Pos}, \text{Neg})}{\max(\text{Pos}, \text{Neg})} \frac{\text{Pos} + \text{Neg}}{\text{Pos} + \text{Neg} + \text{Neu}} \quad (3)$$

4. Ratio of positive and negative comment (SENTIMENT-4)
   For the last feature, we use the ratio value of positive and negative comments. The ratio of positive and negative comments is determined as Pos/N and Neg/N, respectively, where N is the total number of comments, and Pos and Neg are the number positive and negative comments, respectively.

3. Results and Analysis

Metrics and Experiment Settings

To evaluate the performance of our proposed model, we use precision, recall, and F1-score as our evaluation metrics. We also use 10-fold cross validation and employ Logistic Regression and Support Vector Machine as our classifiers. Furthermore, before we did experiments, we made our dataset balance by employing oversampling technique, namely SMOTE [13]. Table 2 shows the comparison of our dataset before and after oversampling.

After that, we performed several experiments to see the contribution of each feature type, as well as find the best combination of features for our classification model. First, we tried each feature group separately. Second, we performed feature ablation study to see the contribution of each feature group relative to the others. Finally, we compared the effect of two feature sources, i.e., the body of article and the commentary section, upon classification performance to see which source of information contribute most to the task.

Result

First, we run Chi-Square test to see the contribution of every single feature for the classification task. The result can be seen in Table 3. Moreover, we only show top-10 most discriminative features. It is interesting to see that all feature groups have their representatives in top-10, except for META features. Next, to see the effect of each feature group as a whole, we conducted experiment involving only one feature group. As can be seen in Table 4, META features gives the worst performance compared to the other feature groups (around 61.5% and 59.2% for LogReg and SVM, respectively). This result is actually inline with the information described in Table 3. STRUCTURAL and LINGUISTIC feature groups are considerably important our detecting controversial contents. When we only used STRUCTURAL feature group, the accuracy achieved 79.4% with Logistic Regression model.

Next, we performed feature ablation study (Table 5), i.e., empirical analysis task that explores the contribution of each feature group by omitting each group while keeping the other feature groups. As can be seen in Table 5, the worst accuracy was yielded when we omit STRUCTURAL feature group, which means that STRUCTURAL feature group is the most discriminative feature group for this task. Until now, it seems like Logistic Regression model outperforms Support Vector Machine. Finally, we performed an experiment to see which source of features (either from article body or comment section) has the most discrim- inative power. Table 6 shows that information (features) from commentary section has notable contribution for our classification task. When we used features extracted from commentary section, the performance reached more than 81%, while the features extracted from article body resulted in much lower classification performance (below 63% in terms of accuracy). The best performance in our experiment was achieved when we used all features and Logistic Regression as our classifier, i.e., 82.8% and 83.1% in terms of accuracy and F1-score, respectively. Moreover, the value of precision and recall for ”Controversy” label tend to be similar in many scenarios.

It is also worth to know several reasons for False Positive in our classification task. False Positive means that we mis-classify non-controversy article as controversy article. Based on our observation, this is mostly due to lengthy SPAM comments which can contain around 1180 words. This harms the performance of our classifie by reducing the discriminative power of some features, including STRUCTURAL-4 feature. The other case is due to lack of opinion lexical resources (for In-

TABLE 2
OUR DATASET (BEFORE AND AFTER OVERSAMPLING)

Condition	Non-Controversy	Controversy	Total
Original	304	205	509
Oversampled (SMOTE)	304	304	608

TABLE 3
CHI-SQUARE VALUE OF EVERY SINGLE PROPOSED FEATURE

Rank	Feature Name	Category Name
1	Average Length of Comments	STRUCTURAL-4
2	Capital Letters	LINGUISTIC-3
3	The Ratio of Negative Comments	SENTIMENT-4
4	Mixed Sentiment Score	SENTIMENT-3
5	Maximum Thread Length	STRUCTURAL-5
6	Bias Lexicon	LINGUISTIC-4
7	Question Mark	LINGUISTIC-1
8	Average Sentiment Scores in Comments	SENTIMENT-1
9	Average Number of Comments	STRUCTURAL-3
10	Reply Comments	STRUCTURAL-1

TABLE 4
THE PERFORMANCE OF DETECTION - USING ONLY ONE FEATURE GROUP

Feature Group	Class	Logistic Reg				SVM			
		Prec	Rec	F1	Acc	Prec	Rec	F1	Acc
META	Non-Controversy	0.61	0.60	0.61	0.61	0.65	0.38	0.48	0.59
	Controversy	0.61	0.62	0.61		0.56	0.80	0.66	
STRUCTURAL	Non-Controversy	0.77	0.83	0.80	**0.79**	0.67	0.89	0.76	0.73
	Controversy	**0.82**	**0.75**	**0.78**		0.84	0.56	0.67	
LINGUISTIC	Non-Controversy	0.76	0.77	0.76	0.76	0.74	0.79	0.76	0.76
	Controversy	0.76	0.76	0.76		0.78	0.72	0.75	
SENTIMENT	Non-Controversy	0.72	0.71	0.71	0.72	0.75	0.62	0.68	0.71
	Controversy	0.71	0.73	0.72		0.68	0.79	0.73	

TABLE 5
THE PERFORMANCE OF DETECTION – FEATURE ABLATION STUDY

Feature Group (without)	Class	Logistic Reg				SVM			
		Prec	Rec	F1	Acc	Prec	Rec	F1	Acc
META	Non-Controversy	0.80	0.80	0.80	0.80	0.80	0.75	0.78	0.78
	Controversy	0.80	0.80	0.80		0.77	0.81	0.79	
STRUCTURAL	Non-Controversy	0.78	0.78	0.78	0.78	0.80	0.72	0.76	0.77
	Controversy	0.78	0.78	0.78		0.74	0.82	0.78	
LINGUISTIC	Non-Controversy	0.78	0.82	0.80	0.80	0.76	0.74	0.75	0.75
	Controversy	0.81	0.78	0.79		0.75	0.77	0.76	
SENTIMENT	Non-Controversy	0.80	0.83	0.82	**0.81**	0.72	0.88	0.79	0.77
	Controversy	**0.82**	**0.80**	**0.81**		0.84	0.66	0.74	

TABLE 6
THE PERFORMANCE OF DETECTION – THE CONTRIBUTION OF EACH SOURCE OF FEATURES

Feature (Classified based on Source)	Class	Logistic Reg			
		Prec	Rec	F1	Acc
META	Non-Controversy	0.80	0.80	0.80	0.80
	Controversy	0.80	0.80	0.80	
STRUCTURAL	Non-Controversy	0.78	0.78	0.78	0.78
	Controversy	0.78	0.78	0.78	
LINGUISTIC	Non-Controversy	0.78	0.82	0.80	0.80
	Controversy	0.81	0.78	0.79	
SENTIMENT	Non-Controversy	0.80	0.83	0.82	**0.81**
	Controversy	**0.82**	**0.80**	**0.81**	

donesian Language) such that some opinionated words are misclassified in terms of its polarity. Moreover, this problem also raises because our sentiment analysis model does not have capability in detecting target of the sentiment evaluation. For example, the following two sentences have different polarity in the context of Indonesian Presidential Election in 2014, yet they actually support each other towards one candidate.

- *kita yg cerdas sudah pasti pilih no. 2* (we, smart people, will absolutely choose number 2)
- *sudah jelas gak bisa memenuhi janji pertama lalu membuat janji yg lebih tidak masuk akal lagi. cepek deh payah banget.* (It is clear that he was not able to fulfill his first promise, yet he made another new promise which doesn't make sense. how loser you are)

There were two candidates at that time. The first sentence evaluates the first candidate as positive, while the second one evaluates the second candidate as negative, which means that these two sentences actually support the first candidate. As a result, incapability to detect opinion target would certainly drop the performance. We also argue that this phenomenon is the reason why our proposed SENTIMENT feature group is not really superior in our case. In addition, we are not really interested in False Negative since we focus on precision, instead of recall, for this controversial detection task.

4. Conclusion

In this paper, we have shown our approach to automatically detect controversial articles due to several motivations. We proposed a supervised machine learning approach harnessing several handcrafted features. Furthermore, our work mostly focus on feature engineering, in which there are four main feature groups: META, STRUCTURAL, LINGUISTIC, and SENTIMENT feature groups. To see the contribution of every single feature and each feature group, we performed several experiments, including feature ablation study and feature ranking (based on discriminative power). We have found that STRUCTURAL and LINGUISTIC feature groups contribute the most to the classification task, while META feature group seems not to be really important for the task. For SENTIMENT

feature group, we argue that its contribution is considerably low due to the fact that opinion expressed in the sentence is quite complex. As a result, more advanced aspect-based sentiment analysis task is required to improve the performance of the task.

The best performance so far is achieved when we use all proposed feature with Logistic Regression as our model (82.89% in terms of accuracy). Finally, we also conducted experiment to see which source of features (either comments or article body) that contributes the most to the classifier's performance. The result show that features extracted from article body seem not to be really discriminative, compared to features extracted from commentary section. When we used all features extracted from article body, the performance achieved 62.9% in terms of accuracy. On the other hand, features from commentary section can result in 81.7% in terms of accuracy. The fact that the best accuracy is yielded by using all features indicates that the task of controversy detection has to consider many aspects of the article.

In the future, we plan to collect more dataset covering more controversial topics. When our corpus is large enough, we can apply state-of-the-art deep learning approach for text classification, in which feature engineering is no longer needed. In fact, our main reason why we employ feature engineering approach is due to small dataset size. However, our hand-crafted features are really important in under- standing hidden information gleaned inside a controversial articles. We can somehow combine the information from our proposed features and automatic learned-features inferred by deep learning model to achieve better performance.

References

[1] S. Bowman and C. Willis. (2003) We media. the media center.

[2] J. Wiley, "A fair and balanced look at the news: What affects memory for controversial arguments?" *Journal of Memory and Language*, vol. 53, no. 1, pp. 95–109, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0749596X05000215

[3] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi, "He says, she says: conflict and coordination in wikipedia," *In Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 453– 462, 2007.

[4] M. Recasens, C. Danescu-Niculescu-Mizil, and D. Jurafsky, "Linguistic models for analyzing and detecting biased language," in *ACL*, 2013.

[5] Y. Choi, Y. Jung, and S.-H. Myaeng, "Identifying controversial issues and their subtopics in news articles," in *Proceedings of the 2010 Pacific Asia Conference on Intelligence and Security Informatics*, ser. PAISI'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 140–153. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13601-6 16

[6] R. V. Chimmalgi. (2013) Controversy detection in social media. masters thesis. louisiana state university.

[7] Y. Mejova, A. X. Zhang, N. Diakopoulos, and C. Castillo, "Con- troversy and sentiment in online news," *CoRR*, vol. abs/1409.8152, 2014. [Online]. Available: http://arxiv.org/abs/1409.8152

[8] G. Mishne and N. Glance, "Leave a reply: An analysis of weblog comments," in *In Third annual workshop on the Weblogging ecosystem*, 2006.

[9] A.-M. Popescu and M. Pennacchiotti, "Detecting controversial events from twitter," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 1873–1876. [Online]. Available: http://doi.acm.org/10.1145/1871437.1871751

[10] K. Allen, G. Carenini, and R. Ng, "Detecting disagreement in conversations using pseudo-monologic rhetorical structure," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1169–1180. [Online]. Available: http://www.aclweb.org/anthology/D14-1124

[11] S. Dori-Hacohen and J. Allan, *Automated Controversy Detection on the Web*. Cham: Springer International Publishing, 2015, pp. 423–434. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-16354-3 46

[12] C. Vania, "Perolehan opini pada dokumen blog berbahasa indonesia (opinion retrieval on indonesian weblogs)," *Bachelors thesis, Universitas Indonesia*, 2009.

[13] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *CoRR*, vol. abs/1106.1813, 2011. [Online]. Available: http://arxiv.org/abs/1106. 1813

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 42-51
DOI: http://dx.doi.org/10.21609/jiki.v11i1.471

A FLEXIBLE SUB-BLOCK IN REGION BASED IMAGE RETRIEVAL BASED ON TRANSITION REGION

Ahmad Wahyu Rosyadi, Renest Danardono, Siprianus Septian Manek, and Agus Zainal Arifin

Department of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember, Jl. Arief Rahman Hakim, Surabaya 60111, East Java, Indonesia

E-mail: wahyu16@mhs.if.its.ac.id, renest16@mhs.if.its.ac.id, septian16@mhs.if.its.ac.id

Abstract

One of the techniques in region based image retrieval (RBIR) is comparing the global feature of an entire image and the local feature of image's sub-block in query and database image. The determined sub-block must be able to detect an object with varying sizes and locations. So the sub-block with flexible size and location is needed. We propose a new method for local feature extraction by determining the flexible size and location of sub-block based on the transition region in region based image retrieval. Global features of both query and database image are extracted using invariant moment. Local features in database and query image are extracted using hue, saturation, and value (HSV) histogram and local binary patterns (LBP). There are several steps to extract the local feature of sub-block in the query image. The first, preprocessing is conducted to get the transition region, then the flexible sub-block is determined based on the transition region. Afterward, the local feature of sub-block is extracted. The result of this application is the retrieved images ordered by the most similar to the query image. The local feature extraction with the proposed method is effective for image retrieval with precision and recall value are 57%.

Keywords: *RBIR, Transition Region, Invariant moment, HSV, LBP*

Abstrak

Salah satu teknik dalam *region based image retrieval* (RBIR) adalah dengan membandingkan fitur global keseluruhan citra dan fitur lokal dari *sub-block* pada citra *database* dan *query*. *Sub-block* yang ditentukan harus dapat mendeteksi objek yang memiliki beragam ukuran dan lokasi. Jadi *sub-block* yang fleksibel pada ukuran dan lokasi sangatlah dibutuhkan. Kami mengusulkan sebuah metode baru untuk ekstraksi fitur lokal dengan menentukan ukuran dan lokasi *sub-block* yang fleksibel berdasarkan *transition region* dalam *region based image retrieval*. Fitur global pada citra *database* dan *query* diekstraksi menggunakan *invariant moment*. Fitur lokal pada citra *database* dan *query* diekstraksi menggunakan histogram *hue, saturation, and value* (HSV) dan *local binary patterns* (LBP). Terdapat beberapa langkah untuk mengekstraksi fitur lokal *sub-block* di citra *query.* Pertama, *preprocessing* dilakukan untuk mendapat *transition region*, kemudian *sub-block* yang fleksibel ditentukan berdasarkan *transition region* tersebut. Setelah itu, fitur lokal *sub-block* diekstraksi. Hasil dari aplikasi ini adalah citra-citra yang ditemu kembali dan diurutkan berdasarkan yang paling mirip dengan citra *query*. Ekstraksi fitur lokal dengan metode ini efektif untuk temu kembali citra dengan *precision* dan *recall* bernilai 57%.

Kata Kunci: *RBIR, Transition Region, Invariant moment, HSV, LBP*

1. Introduction

Content based image retrieval (CBIR) has been a popular research during recent decades. CBIR takes pictures from a large image database based on their visual resemblance [1-2]. Not all images have annotations and not every annotation can resemble images well, looking for images using their content has many advantages than using their annotation text. Furthermore, the weakness of text-based method can be overcome by using CBIR [1] [3].

One of the techniques in the CBIR system that gives a sample image as a query is query by example (QBE). This technique uses the extracted query feature and compared with the database image feature. This feature includes global and local features. Global features use the entire image to extract their features without considering the user interest[1][4][5], while local features only use some relevant regions of the images with respect to user interest known as region based image retrieval (RBIR). The region of interest (ROI) in the query image must be defined by the user to select the relevant region. Chosen ROI by the user is more relevant to be used in RBIR. However, it becomes troublesome if there are a lot of query images that

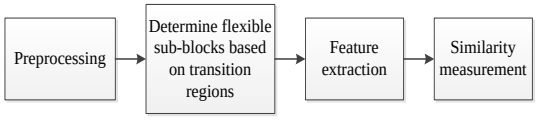


Figure 1. Design System

have to be dealt by the user. Another method is to divide an image into several sub-blocks in a certain size to obtain the ROI [6]. The ROI is determined by selecting the sub-block which is overlapped with the object. Another approach is using Region Important Index (RII) and Saliency Region Overlapping Block (SROB) to select the salient region as a sub-block automatically [7]. Another method is using the ratio of proportional overlapping object (RPOO) which uses a threshold value to determine the degree of overlapping between object and sub-block [1]. In addition, there is a method that make the region of image flexible. This method uses semantic meaningful region (SMR) that is RII with size equal or greater than threshold [8].

Fix location and size sub-block is commonly used in Region Based Image Retrieval. It divides an image into several regions or sub-blocks $b \times b$ in a certain size. Fix location and size sub-block selects the sub-block based on the region of image. If there are several detected objects in all regions of image, sub-blocks will be made in all regions. These sub-blocks may be irrelevant because all regions of image will be used. Fix location and size sub-block also has a weakness in finding the relevant sub-block in image containing small object. It occurs because the size of an object does not meet the minimum threshold to become the relevant sub-block. Another reason is because the use of fix location sub-block. Fix location sub-block cannot adapt the location of an object which means that it can slice and divide one object into several objects if the location or size of the object exceeds the region. If the size of a divided object does not meet the minimum threshold, its region will not be selected as a sub-block even if the actual size of an object meets the minimum threshold. In addition, although method [8] uses the flexible region, that region has to be equal or greather than the threshold to be the ROI. It makes this method not efficient if there is no ROI in the image because the method will use the entire image to campare. So, the flexible sub-block which is able to adapt the size and the location based on the detected object is needed.

In this paper, we propose a new method for local feature extraction by determining the flexible size and location of sub-block based on the transition region in region based image retrieval. Extracted local feature from the flexible sub-block can satisfy the user because it can handle any size of the detected object in the query image and improve the precision and recall value of the retrieving results.

(a) (b)

Figure 2. (a) Original image (b) Transition region image

The rest of this paper is divided into section 2-5. In section 2, proposed method is described. Section 3 reports the experimental results. Evaluations are described in Section 4, and conclusions are described in Section 5.

2. Methods

Query image is an example image that is compared with the database images to get the similar images to it. The database images are images in the database that will be retrieved based on their similarity to the query image. In this study, the query image is compared with the database image by measuring the distance of local and global features. The distance of global and local feature of query and database image is calculated by using Euclidean distance. The global feature is extracted from the entire image. However, the local feature is extracted from sub-block of image. The sub-block has to be the relevant sub-block that has object inside. The proposed method has to select the region or sub-block in flexible size and location based on the transition region.

There are four main steps as we can see in Figure 1 that are conducted in retrieving the images. The first, preprocessing is conducted to get the transition region. After that, the transition region is used to determine the flexible sub-block. Then, local feature is extracted from the determined sub-block. After the features of query image are obtained, the similarity measurement is conducted to determine the similar images.

Preprocessing

Preprocessing is conducted to obtain the transition region image that will be used in the following process. The noise in the query image has to be reduced by using the Gaussian filter. Query image is also required to have a one-color channel in order to speed up the computing process in the next process. We use the grayscale function to change the image with three-color channel into the one-color channel.

To detect the object, some research use image segmentation [1][6]. Image segmentation is a tech-

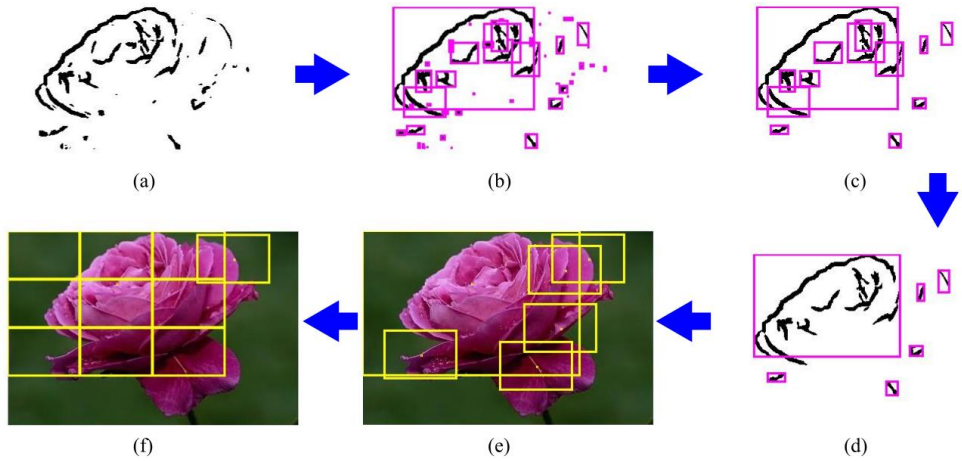


Figure 1. (a) Transition region image (b) Transition region image with its area or block (c) Normalized transition region with its area or block (d) Image after merging the transition region (e) Original image with adjusted block size

nique for separating an object from the background [8–13]. In this study, transition region is used as the salient region which is able to find the transition zone near the contour of the true object [11]. The transition region extraction process is performed after filtering and grayscale converting process to obtain the transition region as we can see in Figure 2(b).

A transition region is a structure of image that is similar with the edge. The transition region has three characteristics [15]. The first is the transition region usually has a width of several pixels near an edge. The second is that it is around the object and should be located between the object and the background. The last characteristic is that there is a dramatically and frequently change in gray levels within pixels in the transition region. Thus, the information for extracting transition regions can be obtained by transitional pixels which have larger magnitude and frequency of gray level changes than non-transitional pixels.

There are many descriptors [15–17] that have been developed for extracting transition regions. In this paper we use local variance to extract the transition region. Local variance can distinguish the area contains edge or not [12]. Edge generally exists in the area that has high value of variance. Hence, local variance is used because it is more important for finding transition region than local complexity [18].

For a center pixel $p(i,j)$ of $m \times m$ local neighborhood, the local variance can be calculated using equation(1).

$$LV(i,j) = \frac{1}{m^2-1}\sum_{x=1}^{m}\sum_{y=1}^{m}(f(x,y) - \bar{f})^2 \quad (1)$$

where $f(x,y)$ defines gray level value of a local coordinate in a neighborhood and the mean of gray level of that neighborhood is denoted as $\bar{f}$.

By sliding the window from left to right and top to bottom, the local variance measurement is conducted throughout the image to achieve the matrix of variance as shown in equation(2).

After local variance matrix is obtained, it is converted into a vector and sorted all values in descending order. Transitional pixels are generated by choosing the first αn pixels from sorted vector, where n is the total pixel number in the image and the value of α is 0.05 based on [11]. A label matrix TP is defined by using equation(3) to denote transitional pixels.

$$LV = \begin{bmatrix} LV(1,1) & LV(1,2) & \dots & LV(1,N) \\ LV(2,1) & LV(2,2) & \dots & LV(2,N) \\ \dots & \dots & \dots & \dots \\ LV(M,1) & LV(M,1) & \dots & LV(M,N) \end{bmatrix} \quad (2)$$

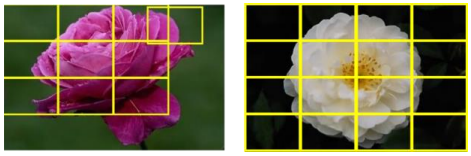
where, M denotes the height of the image and N denotes the width of the image.

$$TP(i,j) = \begin{cases} 0, \text{if } p(i,j) \text{is a transitional pixel} \\ 1, \qquad \text{otherwise} \end{cases} \quad (3)$$

The transition region is composed by grouping connected transitional pixels and making every different group into the different transition region TR. Then all transition regions are labeled with different number k.

Determine flexible sub-blocks based on transition regions

Flexible sub-block is used for the query image. However, fix sub-block is used for the database image. Database image is divided into several sub-blocks in a certain size $b \times b$. Certain size $b \times b$ is also used for the query image to obtain the standard size of sub-block. The standard size of sub-block



(a) (b)

Figure 2. (a) Sub-block in query image (b) Sub-block in database image

can be obtained by dividing the query image into several sub-blocks in a certain size $b \times b$, then the width and the height of a sub-block are used as a standard size of sub-block.

The goal of determining flexible sub-blocks is to make a sub-block adjust to the size and the place of the transition region. Flexible sub-block can focus on the transition region and minimize the selected background in the sub-block. Determining flexible sub-block can be conducted via following steps:

1. Measure the standard size of sub-block's area LSB by multiplying the standard size of sub-block's width WSB and the standard size of sub-block's height HSB using equation(4) and (5) respectively.

$$WSB = N/b \tag{4}$$

$$HSB = M/b \tag{5}$$

2. Make an area for every transition region as shown in Figure 3(b). The area is based on the top-left and the bottom-right of the transition region. The top-left consist of two values: minimum index of row from transition region k bt_k and minimum index of column from transition region bl_k. And the bottom-right consist of maximum index of row from transition region k bb_k and maximum index of column from transition region br_k.
3. Measure the wide of every transition region LTR_k by multiplying its width WTR_k and its height HTR_k. It can be calculated as:

$$WTR_k = br_k - bl_k \tag{6}$$

$$HTR_k = bb_k - bt_k \tag{7}$$

$$LTR_k = HTR_k * WTR_k \tag{8}$$

4. Delete a transition region that has an area less than 3% of a standard sub-block's area LSB, the result is shown in Figure 3(c).
5. Merge a transition region with another 7transition region if its center is in area of another transition region as illustrated in Figure 3(d).
6. Make a block BTR_k for every transition region that its width WTR_k and its height HTR_k adjust their size according to the standard size of sub-block's width WSB and height HSB as shown in Figure 3(e). It can be defined as:

$$WTR_k = \left\{ \begin{array}{c} WSB * z(v)| \\ (WSB * (z(v) - 1)) \\ < WTR_k < (WSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{9}$$

$$HTR_k = \left\{ \begin{array}{c} HSB * z(v)| \\ (HSB * (z(v) - 1)) \\ < HTR_k < (HSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{10}$$

where z is an array having values ranging from 1 to b and defined as follows

$$z = 1, 2, 3, \dots, b \tag{11}$$

The width WTR_k and the height HTR_k of block are updated but the center is not. It can adjust the top-left, top-right, bottom-left, and bottom-right automatically. Record the probability number of sub-block that can be made vertically $SHTR_k$ and horizontally $SWTR_k$ from HTR_k and WTR_k. It defined as follows.

$$SWTR_k = \left\{ \begin{array}{c} z(v)| \\ (WSB * (z(v) - 1)) \\ < WTR_k < (WSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{12}$$

$$SHTR_k = \left\{ \begin{array}{c} z(v)| \\ (HSB * (z(v) - 1)) \\ < HTR_k < (HSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{13}$$

7. Delete the block that its center is in area of another not smaller block.
8. Record the index of row iy_k and column ix_k where the block BTR_k exists. The index is based on matrix b x b of image. It can be defined as

$$iy_k = \left\{ \begin{array}{c} z(v)| \\ (HSB * (z(v) - 1)) \\ < bt_k < (HSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{14}$$

$$ix_k = \left\{ \begin{array}{c} z(v)| \\ (WSB * (z(v) - 1)) \\ < bl_k < (WSB * z(v))| \\ \forall z(v) \in z \end{array} \right\} \tag{15}$$

TABLE 1
PRECISION VALUE OF THRESHOLD 1%, 2%, 3%, 4%, AND 5%

Threshold	Precision
1%	0.55
2%	0.56
3%	**0.57**
4%	0.56
5%	0.56

TABLE 2
PRECISION OF RETRIEVING 100 IMAGES WITH DIFFERENT B

Label query	Precision	
	b=4	b=5
Africa	0.55	**0.57**
Beaches	**0.42**	0.34
Buildings	**0.51**	0.34
Bus	**0.84**	**0.84**
Dinosaur	**0.90**	0.74
Elephant	**0.25**	0.23
Flowers	0.43	**0.46**
Horse	**0.74**	0.70
Mountains	0.26	**0.34**
Food	**0.75**	0.58
Average	**0.57**	0.51

9. Divide block BTR_k into sub-block $STR_{k,it}$. BTR_k is divided into a tb_knumber of sub-block. The value of tb_k is obtained by multiplying $SWTR_k$ and $SHTR_k$ that is shown in equation(16). Sub-block $STR_{k,it}$ has width equal to WSB and height equal to HSB as we can see in Figure 3(f).

$$tb_k = SWTR_k * SHTR_k \quad (16)$$

$$it = 1,2,\dots,tb_k \quad (17)$$

$$STR_{k,it} \subseteq BTR_k \quad (18)$$

10. Record the index of row $iys_{k,it}$ and column $ixs_{k,it}$ where the sub-block exists. The index is based on the BTR_k's index, $SWTR_k$, and $SHTR_k$. It can be defined as

$$ixs_{k,it} = ix_k + ic(v) - 1 \quad (19)$$

$$iys_{k,it} = iy_k + ir(v) - 1 \quad (20)$$

where

$$ic = 1,2,\dots,SWTR_k \quad (21)$$

$$ir = 1,2,\dots,SHTR_k \quad (22)$$

11. Use the block as the sub-block SQ_k in image query, and the sub-block of a block as sub of sub-block $SSQ_{k,it}$

Feature Extraction

Local and global feature are used in the similarity seasurement process. Local feature of database image is extracted from all sub-blocks in it. On the other hand, local feature of query image is extracted from the selected sub-block only. Color and texture are used as local feature, while shape of object is used as descriptor for the global feature.

HSV is used as the color feature in this research. HSV color space is developed to adapt to the visual characteristic of human, considering hue, saturation, and value [19]. In this research, color is generated from every determined sub-block in query and database image.

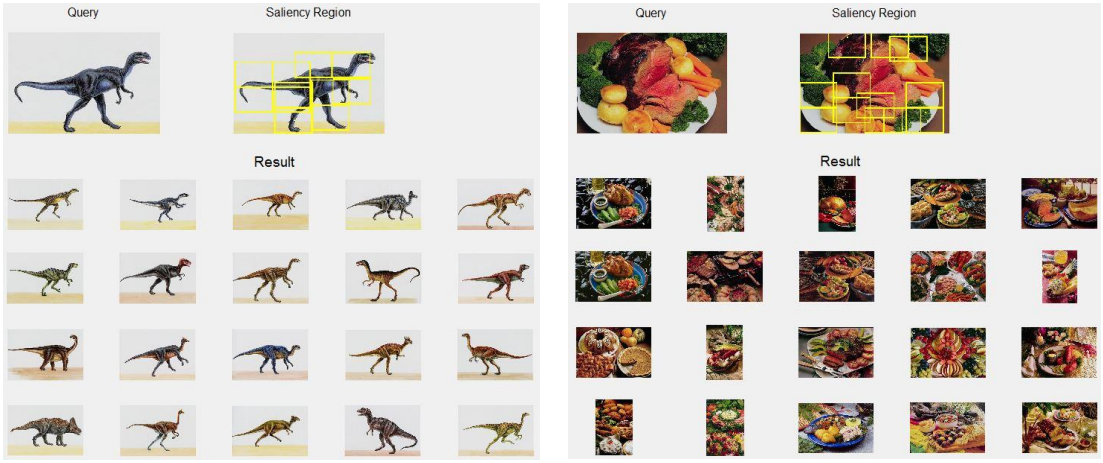
Local binary pattern (LBP) is used as texture feature because it is efficient and has a good classification ability [20]. LBP has the ability to define the surface of an object and its relationship with the surrounding area [7]. In this research, LBP is calculated for every pixel in the determined sub-block. To represent the sub-block's texture, then the histogram of LBP is created.

Shape is used as the global feature in this research. Shape descriptor can be divided into two types that is region-based shape descriptor and contour-based shape descriptor. One of the important shape descriptors is moment invariants. Hu invariant moments is used as the global feature because it is invariant under rotation, translation, and also changes in scale [1].

Similarity Measurement

Similarity between query image and database image is computed by using Euclidean distance for every feature. The similarity distance of local feature is obtained by calculating the distance for each selected sub-blocks of query and database image as seen in Figure 4. Sub-blocks of query image are convoluted in database image in its own area to measure the minimum distance of all iterations. The distance of sub-block in query image that has some subs of sub-block is done by convoluting the sub-block of query image in database image, but the distance value of every iteration is determined by taking the average of all distance values of sub-block's subs. The distance between qu-ery image sub-block and database image sub-block can be calculated using equation(24), where $FSQ_{k,it}$ defined the local feature of sub-block's sub $SSQ_{k,it}$, $FDB_{q+isy,z+isx}$ defined the local feature of sub-block in database image, $q = \{-1,0,1\}$, $z = \{-1,0,1\}$, isy defined $iys_{k,it}$ and isx defined $ixs_{k,it}$.

The final distance of a local feature is described as average distance for every sub-block shown as



(a) (b)

Figure 3. Example query and results for retrieval 20 images from category (a) Dinosaur (b) Food

$$d(SQ, SDB) = \frac{\sum_{k=1}^{n} d(SQ_k - SDB_{q,z})}{n} \quad (23)$$

$$d\left(SQ_k, SDB_{q,z}\right) = \min_{q,z}\left(\frac{\sum_{it=1}^{tb_k}\sqrt{(FSQ_{k,it} - FDB_{q+isy,z+isx})^2}}{tb_k}\right) \quad (24)$$

Euclidean distance is applied directly to measure the similarity of the global feature of query and database image. A weight is assigned to every feature before calculating the total distance. Based on [1], the optimal weight is 0.1, 0.4, 0.5 for the weight of color feature w_c, the weight of texture feature w_t, and the weight of shape feature w_s, respectively. The distances of two local features and one global feature are combined to obtain the total distance as shown in equation(25).

$$d(q, db) = w_c . d_c + w_t . d_t + w_s . d_s \quad (25)$$

The sorting by ascending order is performed after obtaining the distances to all images in the database. The smaller the distance, indicate that two images have higher similarity.

3. Results and Analysis

This research used Wang's image datasets that is usually used in image retrieval research. This dataset has 10 categories. Each category consists of 100 images. So there are 1000 images in this dataset.

Parameter to Delete Transition Region

In this study, the threshold used for determining the transition region is $0.05 \times n$ pixels as explained in [11]. The determined transition region is not always big, there will be some small transition regions. The small transition region may represent part of object or even part of background. We assume that the very small transition region represents the part of background. So it is important to delete the very small transition region, by deleting the very small transition region, we can reduce the number of transition region and reduce the computing time for retrieving the images. In this study we use a threshold value to delete the transition region. The threshold value is 3% of a sub-block's area based on our experiment and shown in Table 1.

Performance Measure Using Precision and Recall

Precision is used to measure the number of selected images are relevant, while recall is used to measure the number of the relevant images are selected. Precision and recall can be calculated by equation(26) and (27) respectively.

$$\text{precision} = \frac{TP}{TP+FP} \quad (26)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (27)$$

The number of relevant images that can be retrieved by the system is defined as *TP* (True Positive), the number of relevant images that cannot be retrieved by the system is defined as *FN* (False Negative), the number of irrelevant images but retrieved by the system is defined as *FP* (False Positive),and the number of irrelevant images that is not retrieved by the system is defined as *TN* (True Negative). The results of a query image are shown based on their ranking. The distance between query and image database is calculated to obtain the ranking of each image in database.

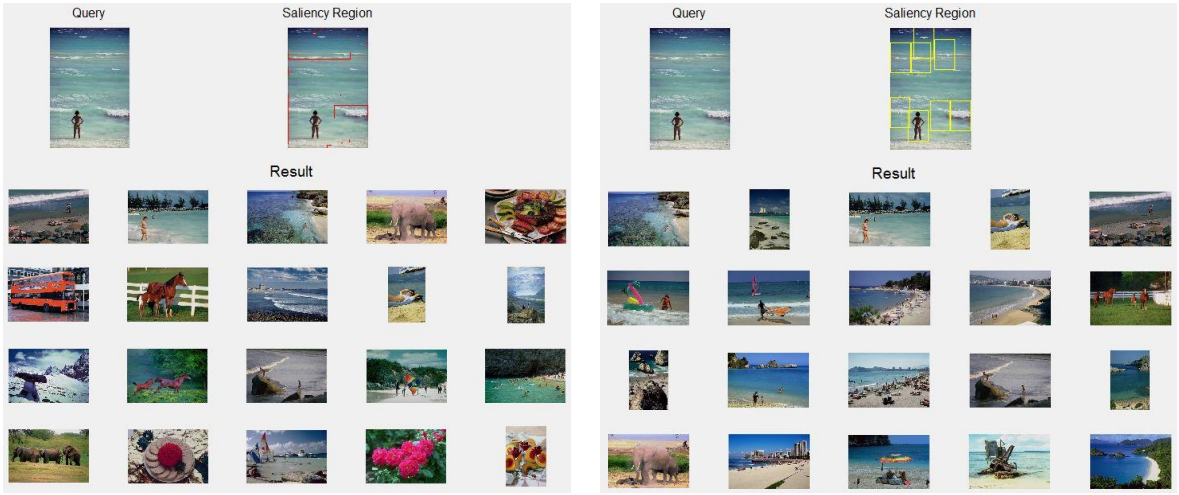


Figure 4. Example query and results for retrieval 20 image on query image with small object, (a) Query and results of RPOO, (b) Query and results of proposed method

The number of sub-block is very important for retrieving images. Parameter *b* with value of 4 has better precision value than *b* with value of 5 as shown in Table 2. So in this study, *b* with value of 4 is used.

It can be seen from Table 3 that the proposed method has better average precision value than Vimina & Jacob [6] and RPOO [1] in retrieving 20 images. The proposed method also has better precision value than RPOO [1] in most of cases except Elephant and Horse in retrieving 100 images as shown in Table 4. In addition, we also conducted image retrieval in image containing small object as shown in Table 5 and Figure 6. From Table 5 we can see that the proposed method has better average precision and recall value than RPOO in image retrieval of small object.

Evaluation

There are several points that could be discussed from the experimental results. As we can see in Table 2 that *b* with value of 4 has better precision value than 5. It is because the value of *b* affects the wide of sub-block. Bigger value of *b* means smaller wide of sub-block. When the wide of sub-block is small, the similarity measurement process is conducted in the small area because the convolution process is done in surrounding area of sub-block instead of in the entire image. Bigger value of *b* can focus the sub-block on small object well. However, it may not be good for similarity measurement process if the similar object is outside the area because the similarity value will be low. Moreover, bigger value of *b* may result bigger number of sub-block that can make the computing time is longer because there will be many sub-blocks to compare by convoluting each sub-block in its surrounding area of sub-block. Based on our experiment, *b* equals 4 has better precision value than 5 because the size of sub-block with *b* equals 4 is fit for most of objects of the dataset image even for object that has small size.

Table 3 showed that Africa has precision value of 100% in retrieving 20 images and 55% in retrieving 100 images because the proposed method can determine sub-block from the parts of object. The determined sub-block has unique color and texture which is similar in its own category in database images.

Bus has precision value of 100% in retrieving 20 images and 84% in retrieving 100 images because Bus has unique color in the body and the glass, so it can determine a good transition region. Moreover, the shape of bus is similar in bus category, so the global feature shape can identify it well.

Dinosaur images also has precision value of 100% in retrieving 20 images and 90% in retrieving 100 images, it is because the selected sub-blocks in the query image are determined by the transition regions surrounding the body of dinosaur that is shown in Figure 5(a). In dinosaur, all the determined transition regions are from the dinosaur body because the background image is relatively homogeneous.

Although the horse category has heterogeneous background, the proposed method can result a good precision value of 100% in retrieving 20 images and 74% in retrieving 100 images because the sub-block is made surrounding the object. It is happened because the transitional pixels which have big value of local variance are determined from pixels surrounding the horse body. Another cate-

TABLE 1
PRECISION OF RETRIEVING 20 IMAGES USING SAME IMAGES IN DIFFERENT METHODS

Label query	Precision of retrieving 20 images		
	Vimina and Jacob	RPOO	Proposed method
Africa	0.71	0.80	**1.00**
Beaches	0.43	0.45	**0.85**
Buildings	0.53	0.70	**0.90**
Bus	0.85	0.95	**1.00**
Dinosaur	0.99	**1.00**	**1.00**
Elephant	**0.59**	0.55	0.20
Flowers	**0.90**	0.75	0.80
Horse	0.92	**1.00**	**1.00**
Mountains	0.38	0.55	**0.65**
Food	0.72	0.65	**1.00**
Average	0.70	0.74	**0.84**

TABLE 3
PRECISION AND RECALL OF RETRIEVING 100 IMAGES USING SAME IMAGES IN DIFFERENT METHODS

Label query	RPOO		Proposed method	
	Precision	Recall	Precision	Recall
Africa	0.42	0.42	**0.55**	**0.55**
Beaches	0.23	0.23	**0.42**	**0.42**
Buildings	0.39	0.39	**0.51**	**0.51**
Bus	0.67	0.67	**0.84**	**0.84**
Dinosaur	0.80	0.80	**0.90**	**0.90**
Elephant	**0.29**	**0.29**	0.25	0.25
Flowers	0.25	0.25	**0.43**	**0.43**
Horse	**0.79**	**0.79**	0.74	0.74
Mountains	0.21	0.21	**0.26**	**0.26**
Food	0.38	0.38	**0.75**	**0.75**
Average	0.44	0.44	**0.57**	**0.57**

TABLE 2
PRECISION AND RECALL OF RETRIEVING 100 IMAGES OF IMAGE CONTAINING SMALL OBJECT USING DIFFERENT METHODS

Label query	RPOO		Proposed method	
	Precision	Recall	Precision	Recall
Beach-49	0.26	0.26	**0.30**	**0.30**
Beach-62	0.21	0.21	**0.26**	**0.26**
Beach-64	0.23	0.23	**0.42**	**0.42**
Beach-74	0.24	0.24	**0.40**	**0.40**
Average	0.24	0.24	**0.35**	**0.35**

gory that has precision value of 100% in retrieving 20 images is food, it is because the determined sub-blocks from query image is not overall of image. Although the object is big, the determined sub-blocks from query image are part of the object, so it just compare the saliency regions.

In elephant category the proposed method has worst precision value, it is caused by the convolution process. Even the determined transition region is good, the convolution process makes it like a separate object and many images in database have similar color of elephant category.

The proposed method successfully retrieved the images with varying sizes of object as shown in the Table 3–5. Table 3 and 4 showed the result of retrieval in all categories with varying sizes of object, while the Table 5 showed the results of image retrieval containing small object.

For general image with varying sizes of object, the proposed method has better average precision and recall value than the result of RPOO as shown in Table 3 and 4. Furthermore, in cases of image retrieval of image contains small object, the proposed method also has the better result than the result of RPOO in all cases. It means that the proposed method has the ability in retrieving the images from image containing varying sizes of object.

The proposed method is capable of retrieving the images of image containing varying sizes of object because it uses the flexible sub-block. It has the ability to create sub-block in any size and any location based on the determined transition region. It can adapt the size of sub-block in the determined transition region by fitting the transition region's size with the standard size of sub-block. It is different from the fix sub-block that cannot adapt the size of sub-block and cannot make a sub-block in the region of a small object. The proposed method also places the sub-block based on the transition region instead of placing sub-block based on the region of an image. Placing sub-block based on the region of an image that is used in the fix sub-block will extract the irrelevant local feature if the object is not in the right place because the sub-block will contain too much background. Consequently, the extracted local feature from the flexible sub-block will be more relevant because the sub-block contains all the transition regions of object and minimizes the selected background.

Extracted local feature from the sub-block that has a lot of object inside is a good local feature. It is explained in [1] and [6] that the better segmentation result from the segmentation process will select sub-block based on the detected object more effectively and obtain a good image retrieval result.

Moreover, convolution process can detect the similar sub-block of database image in surrounding area of query image's sub-block. In this proposed method, sub-blocks of query image maybe a lot, but the selected sub-blocks is not overall image, so it can affect the performance of local feature. However, the convolution process sometimes determines the bad result although the determined sub-block is good, it occurs because many images in database have similar color or texture.

In addition, the generated sub-block may be irrelevant if the object is too small because the sub-block contains too much background. It happens

because the proposed method still needs a standard size to generate the flexible sub-block. The size of its flexibility has to be set by fitting its size with the standard size of sub-block.

Sometimes the deleted sub-block is a useful sub-block. It happens because the proposed method uses the size of sub-block as the only condition to delete sub-block. So the improvement in selecting the sub-block that will be deleted is needed. And there will be improvement in the convolution process, because in case of elephant category, the selected sub-block was good but the convolution process produced the bad precision value.

4. Conclusion

The proposed method successfully retrieved the images with varying sizes of object as shown in the experimental results. In most of categories in dataset, the proposed method has better average precision and recall value than the result of RPOO. Moreover, the proposed method showed superior results than RPOO in retrieving images from an image contains small object by surpassing the result of RPOO in all testing. However, the proposed method is not good enough to retrieve images from an image contains too small object. In addition, the proposed method does not have a good condition to determine which sub-block to be deleted.

References

[1] A. Z. Arifin, R. W. Sholikah, D. F. H. P., and D. A. Navastara, "Region Based Image Retrieval Using Ratio of Proportional Overlapping Object," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 4, p. 1608, 2016.

[2] M. Athoillah, I. Irawan, M., and M. Imah, Elly, "Study Comparison of SVM-, K-NN- and Backpropagation-Based Classifier for Image Retrieval," *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Information)*, vol. 8, no. 1, pp. 11–18, 2015.

[3] H. Maghfirah, F. Arnia, and K. Munadi, "Temu Kembali Citra Busana Muslimah Berdasarkan Bentuk Menggunakan Curvature Scale Space (CSS)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 1, pp. 74–83, 2017.

[4] X.-Y. Wang, Y.-J. Yu, and H.-Y. Yang, "An effective image retrieval scheme using color, texture and shape features," *Comput. Stand. Interfaces*, vol. 33, no. 1, pp. 59–68, 2011.

[5] X. Wang, Z. Chen, and J. Yun, "An effective method for color image retrieval based on texture," *Comput. Stand. Interfaces*, vol. 34, no. 1, pp. 31–35, 2012.

[6] E. R. Vimina and K. P. Jacob, "A Sub-block Based Image Retrieval Using Modified Integrated Region Matching," vol. 10, no. 1, pp. 686–692, 2013.

[7] N. Shrivastava and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," *Inf. Sci. (Ny).*, vol. 259, pp. 212–224, 2014.

[8] X. Yang and L. Cai, "Adaptive region matching for region-based image retrieval by constructing region importance index," *IET Comput. Vis.*, vol. 8, no. 2, pp. 141–151, 2014.

[9] G. P. Cahyono, A. Y. Aprilio, and H. Ramadhan, "Multithresholding in Grayscale Image Using Peak Finding Approach and Hierarchical Cluster Analysis," *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Information)*, vol. 7, no. 2, pp. 83–89, 2014.

[10] M. Hani'ah, C. S. kusuma Aditya, A. Harto, and A. Z. Arifin, "Cortical Bone Segmentation Using Watershed and Region Merging Based on Statistical Features," *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Information)*, vol. 8, no. 2, pp. 76–82, 2015.

[11] Z. Li, G. Liu, D. Zhang, and Y. Xu, "Robust single-object image segmentation based on salient transition region," *Pattern Recognit.*, vol. 52, pp. 317–331, 2016.

[12] P. Parida and N. Bhoi, "Transition region based single and multiple object segmentation of gray scale images," *Eng. Sci. Technol. an Int. J.*, vol. 19, no. 3, pp. 1206–1215, 2016.

[13] N. P. Husain and C. Fatichah, "Segmentasi Citra Sel Tunggal Smear Serviks Menggunakan Radiating Component Normalized Generalized GVFS," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 1, pp. 107–114, 2017.

[14] D. Tuwohingide and C. Fatichah, "Spatial Fuzzy C-means dan Rapid Region Merging untuk Pemisahan Sel Kanker Payudara," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 1, pp. 51–57, 2017.

[15] Y. J. Zhang and J. J. Gerbrands, "Transition region determination based thresholding," Pattern *Recognit. Lett.*, vol. 12, no. 1, pp. 13–23, 1991.

[16] C. Yan, N. Sang, and T. Zhang, "Local entropy-based transition region extraction and thresholding," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 2935–2941, 2003.

[17] Z. Li and C. Liu, "Gray level difference-based transition region extraction and thresholding," *Comput. Electr. Eng.*, vol. 35,

no. 5, pp. 696–704, 2009.

[18] Z. Li, D. Zhang, Y. Xu, and C. Liu, "Modified local entropy-based transition region extraction and thresholding," *Appl. Soft Comput. J.*, vol. 11, no. 8, pp. 5630–5638, 2011.

[19] H. Zhang and X. Jiang, "An Improved Algorithm Based on Texture Feature Extraction for Image Retrieval," *Int. Conf. Intell. Human-Machine Syst. Cybern.*, vol. 8, pp. 281–285, 2016.

[20] G. Cheng and J. Chen, "A local feature descriptor based on Local Binary Patterns," pp. 251–258, 2016.

Jurnal Ilmu Komputer dan Informasi (Journal of a Science and Information). 11/1 (2018), 52-58
DOI: http://dx.doi.org/10.21609/jiki.v11i1.468

AUTOMATIC DETERMINATION OF SEEDS FOR RANDOM WALKER BY SEEDED WATERSHED TRANSFORM FOR TUNA IMAGE SEGMENTATION

Moch Zawaruddin A., Dinial Utami Q.N., Lafnidita Farosanti, and Agus Zainal Arifin

Department of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh November, Kampur ITS Sukolilo, Surabaya, 60111, Indonesia

Email: zawaruddin017@gmail.com, sayadinial@gmail.com, ochadieta@gmail.com, agusza@cs.its.ac.id

Abstract

Tuna fish image classification is an important part to sort out the type and quality of the tuna based on the shape. The image of tuna should have good segmentation results before entering the classification stage. It has uneven lighting and complex texture resulting in inappropriate segmentation. This research proposed method of automatic determination seeded random walker in the watershed region for tuna image segmentation. Random walker is a noise-resistant segmentation method that requires two types of seeds defined by the user, the seed pixels for background and seed pixels for the object. We evaluated the proposed method on 30 images of tuna using relative foreground area error (RAE), misclassification error (ME), and modified Hausdorff distances (MHD) evaluation methods with values of 4.38%, 1.34%, and 1.11%, respectively. This suggests that the seeded random walker method is more ef-fective than existing methods for tuna image segmentation.

Keywords: *image segmentation, watershed, random walker*

Abstrak

Klasifikasi ikan tuna merupakan bagian penting untuk memilah jenis dan kualitas ikan tuna berdasarkan ukurannya. Citra ikan tuna harus memiliki hasil segmentasi yang baik sebelum dilakukan tahap klasifikasi. Citra ikan tuna memiliki pencahayaan yang tidak merata dan teksture yang kompleks sehingga dapat menghasilkan segmentasi yang salah. Penenilitian ini mengusulkan metode penentuan otomatis *seeded random walker* pada *region watershed* untuk segmentasi citra ikan tuna. *Random walker* merupakan metode segmentasi yang tahan terhadap *noise* yang membutuhkan dua tipe *seed* yang ditentukan oleh *user*, yaitu *seed* piksel untuk background dan *seed* piksel untuk objek. Uji coba dilakukan pada 30 citra ikan tuna menggunakan metode evaluasi *relative foreground area error* (RAE), *misclassification error* (ME), dan *modified Hausdorff distances* (MHD) dengan nilai masing-masing 4.38%, 1.34% dan 1.11%. Hal ini menunjukkan bahwa metode *seeded random walker* lebih efektif daripada metode yang telah ada untuk segmentasi ikan tuna.

Kata Kunci: *segmentasi citra, watershed, random walker*

1. Introduction

The process of sorting the tuna is done manually and requires the help of experts, while the number of fish to be sorted is very much. This is will be a consequence of the length of the sorting and the impact on the freshness of the tuna to be processed. To maintain freshness of tuna fish needed tuna classification process quickly and automatically, so the quality of tuna fish production increase [1].

The classification is done after the form object from the image of tuna has been processed. The fish object is derived from a segmentation process that separates the object against the background. Good segmentation will result in an object's shape similar to the original. Random walker is one of the image segmentation methods that are resistant to noise. In general, the random walker is used for image segmentation interactively by user scribble. The user scribble is used manually specify about seeds in which pixels indicate the background and seeds in which pixels indicate the object area. The user's manual determination has the possibility of misleading between the object and the background.

In addition, the use of user scribble on the large fish images will take a long time to process. Massive repetition of the large data is ineffective and impractical in the field. Automatic seeding is required in the segmentation of tuna images.

There are several studies on the determination of the automatic seed for the segmentation process as Fadllullah's proposed method [2]. This method uses gradient barrier watershed based on hierarchal cluster analysis and regional credibility merging

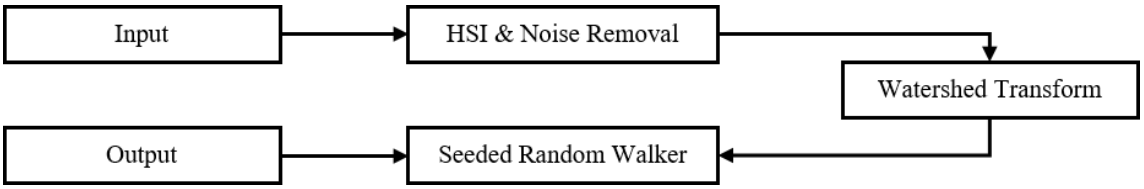


Figure 1. Flowchart of proposed method

(HAC-RCM) for optimization of adaptive threshold determination with the aim of reducing oversegmentation by watershed region. This method combines objects based on two merging criteria. However, in the normalization stage of the image, this method is not optimal in reducing the rough texture of the background area so that the segmentation results are inaccurate. Another method is proposed by Saputra, [3,4]. This method uses the automation of seeded region growing throughout the watershed region. The image of the tuna is transformed within the HSI color space, and only the hue color space is processed to the watershed transform. The determination of the region is obtained based on the highest density region. The threshold parameter is obtained by the difference between the averages of the region's intensity to the average of the region's neighboring intensity. Determination of seed parameters in seeded region growing does not concern to image enhancement and noise removal features so that there is a chance that the selected seed is noise. Seeds located on the noise can produce false segmentation.

Random walker has many advantages including weak boundary detection, noise robustness, and fast computation [5] and has many related and important methods based on it [6][7][8]. In this method, the user should give labels to a small number of pixels about the background and object. Then start a random walker at each unlabeled pixel and calculate the probabilities. These probabilities will determine the probability of each unlabeled pixel belonging to a label. By assigning each pixel to a label with the greatest probability, the interactive image segmentation result can be obtained. This method has shown to perform well on different types of images but is strongly influenced by the placement of the labels within the image. Because of that, we used a random walker for noise removal and weak boundary detection that it can increase the accuracy of tuna image segmentation.

Furthermore, the random walker is a noise-resistant segmentation method that requires two types of seeds defined by the user, the seed pixels for background and seed pixels for the object. In the reality about sorting of tuna fish, there is inefficient for labeling each tuna image by user scribbles while the number of fish to be sorted is very much. We need the determination of the seeds to be automatic so that segmentation with a random walker can be fast and does not require user scribbling. This research proposed method of automatic determination seeded random walker in the watershed region for tuna image segmentation. Seeds on the label for the object and background pixels that generated by watershed can be a substantial resource for a random walker to segmenting image of tuna.

Data in this research is obtained from PT. Aneka Tuna Indonesia with total 30 jpg images. The dimension of each image is 2889×1625 pixels with yellow bucket's background. Groundtruth is obtained by manually process using Adobe Photo-shop's software. Groundtruth consist of two com-ponents that are white as an object and black as the background.

2. Methods

The proposed method is an automatic determination of seed for random walker using region watershed transform for tuna image segmentation. The Figure 1 shown about step by step in our proposed method. The first step, an input image of tuna is transformed into HSI color space and removing noise on hue color space. Then hue image is processed through a watershed transform to generate seeds. The seeds are the pixel position that represents part of an object and part of the background. Seeds that generated by watershed than become an input labels of object and background pixel for a random walker to segmenting image of tuna.

HSI and Noise Removal

There are three processes at this stage, image optimization, HSI transformation of RGB color space, and noise removal. Image optimization is to improve image quality. In this process, we perform the image resizing into 10% of the original image, contrast enhancement, image sharpening, and Gaussian filtering. The RGB image as an optimization result is transformed into an HSI (Hue, Saturation, Intensity) color space and only the hue color spaces are used. The process can be formulated by:

$$H = \begin{cases} 0, & B \leq G \\ 360 - \theta, & B > G' \end{cases} \quad (1)$$

from the Equation(1), H is hue value, and θ formulated by the equation:

$$\theta = \cos^{-1}\left\{\frac{1/2[(R-G)+(R-B)]}{[(R-G)^2(R-B)(G-B)]^{1/2}}\right\}, \quad (2)$$

then noise removing process on the image at hue color space with the following equation:

$$I = \begin{cases} 0, & if\ size\ pixels\ \leq 50 \\ 1, & if\ size\ pixels\ > 50' \end{cases} \quad (3)$$

$$H' = H * I. \quad (4)$$

Noise removal is used for removing pixels in which indicated as noise. It's can improve hue color space shape and increase the segmentation result.

The Figure 2 shown that image of *a1*, *b1*, and *c1* are original images of tuna. Image of *a2*, *b2*, and *c2* are images on hue color space without noise removal. Image of *a3*, *b3*, and *c3* are images on hue color space with noise removal.

Watershed Transform

There are two steps at this stage such as determination of seeds for foreground (object) and background [6,7]. The first step is a determination of seed for foreground on the image using following equation:

$$cost(p,q) = \begin{cases} LS(p).d(p,q), & if\ f(p) > f(q) \\ LS(q).d(p,q), & if\ f(p) < f(q) \\ 0{,}5.\left(LS(p)+LS(q)\right).d(p,q), & if\ f(p) = f(q) \end{cases} \quad (5)$$

$$LS(p) = \underset{q \in N_G}{MAX}\left(\frac{f(p)-f(q)}{d(p,q)}\right), \quad (6)$$

$$LS(q) = \underset{p \in N_G}{MAX}\left(\frac{f(q)-f(p)}{d(p,q)}\right), \quad (7)$$

$$d(p,q) = \sqrt{(p_1-p_2)^2+(q_1-q_2)^2}. \quad (8)$$

where $f(p)$ value on matrix p will compare with $f(q)$ value on matrix q, if $f(p)$ greater than $f(q)$, then calculate gradient value $LS(p)$ at one-pixel point with all the pixel points on the image. $LS(p)$ calculates based on different value $f(p)$ with $f(q)$ and divided by the distance between pixels. Number of pixel steps at matrix $cost(p,q)$ calculated based on $LS(p)$ with $d(p,q)$. The highest $LS(p)$ value will be processed by determining temporary gradient value and otherwise, it will be ignored.

$$T_f^{\pi}(p,q) = \sum_{i=0}^{l-1} d(p_i,p_{i+1})\ cost\ (p_i,p_{i+1}) \quad (9)$$

$$T_f(p,q) = \underset{\pi\ \epsilon\ [p\ \rightsquigarrow\ q]}{MIN} T_f^{\pi}(p,q) \quad (10)$$

Temporary gradient value $T_f^{\pi}(p,q)$ count considered highest $LS(p)$ that is multiplied with $cost(p,q)$ on each pixel. The optimal topographic value $T_f^{\pi}(p,q)$ is taking minimum value $T_f^{\pi}(p,q)$. $T_f(p,q)$ will be processed for determination optimum total number of region that used as seed of object.

Determination optimum number of region considered value of peak point in histogram from hue image *H'*. Then, smoothing process on his-

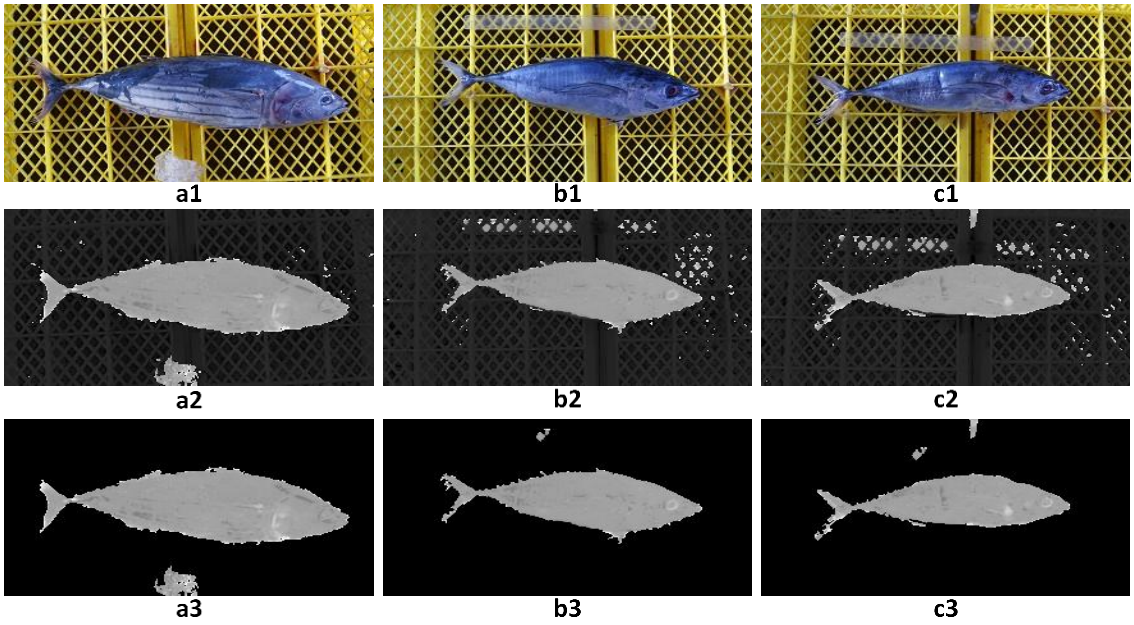


Figure 2. HSI transformation and noise removing

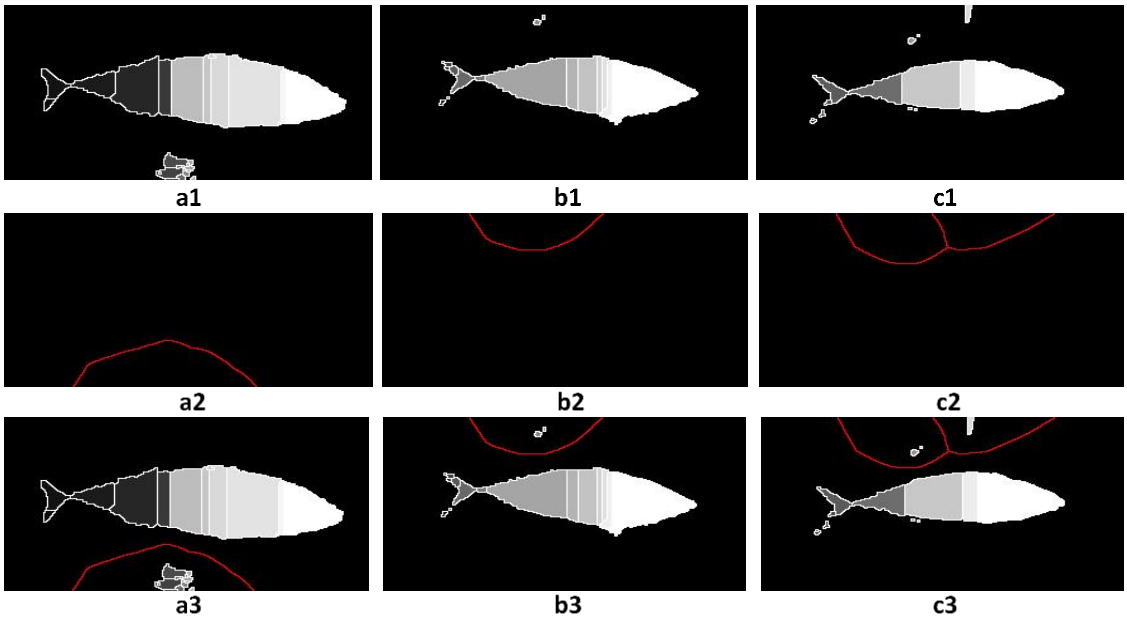


Figure 3. Image result of watershed transform

togram *H'* will be executed to reduce the total number of peak point in a histogram of the following equation:

$$SH(i,1) = \sum_{i=2}^{255} \frac{(i-1)+i+(i+1)}{3}. \quad (11)$$

Next step calculates the total number of pixel frequency on peak point f divided by a total number of peak point n.

$$T_{SH} = \frac{\sum_{i=1}^{n} f}{n}. \quad (12)$$

Value of T_{SH} becomes the optimum region to be taken for the seed of object and then the determinate region will be selected. Afterward, we calculated the size of each region and sorting it descending. Regions selected will be taken the value of its centroid into seeds of foreground *(SF)*.

The second step is determination seeds of background. Value of seeds obtained by calculates the distance of the pixel from H' with Euclidean Distance and watershed transform. The result of its calculation will be taken in pixels with gray value equals to zero and selected as seed for background (*SB*). Determination of seed considered by dividing line between background area and object area to avoid background markers to be too close to the edges of the object.

In Figure 3, images of *a1*, *b1*, and *c1* are the result of watershed transform for the foreground that the centroid on each region will be used as seeds label for an object on random walker process. While, the image on *a2*, *b2*, and *c2* are result from watershed transform for the background that value on each pixel (red marker) will be used as seeds label for the background. The image of *a3*, *b3*, and *c3* are merger image from watershed for foreground and background.

The input seeds for a random walker to segmenting image of tuna can get by define *SB* as seeds for background label and *SF* as seeds for object label.

Random Walker

Dirichlet is used to obtain the probability values of random walks [5,8,9]. The Dirichlet integral $D[q]$ using following equation:

$$[q] = \frac{1}{2}\int_{\Omega} |\nabla q|^2 d\Omega, \quad (13)$$

where q as a field and Ω as region from the image. Harmonic function is a function that fulfill Laplace's formula $\nabla^2 q = 0$. We mean a Component combination matrix of Laplacian L where elemen L_{ij} each point v_i and v_j Can be an equation as

$$L_i = \begin{cases} d_i, & i = j, \\ -w_{ij}, & v_i \text{ and } v_j \text{ are incident node}, \\ 0, & \text{else}. \end{cases} \quad (14)$$

We determine the size of the matrix an $m \times n$ matrix of edge-point incidence A where elements $Ae_{ij}v_k$ for vertex v_k and edge e_{ij} respectively can be an equation as

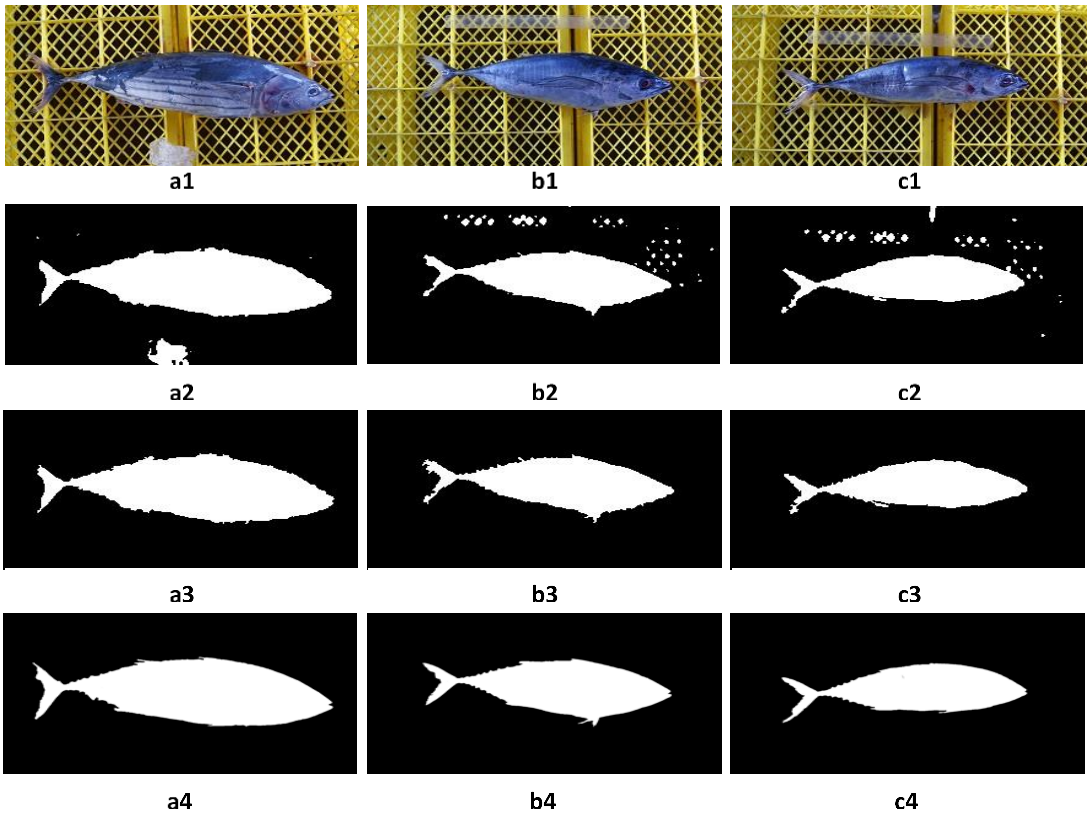


Figure 4. Experiment result using proposed method and Watershed Region Growing

$$Ae_{ij}v_k = \begin{cases} +1, & i = k, \\ -1, & j = k, \\ 0, & \text{else.} \end{cases} \quad (15)$$

Matrix A serve as a composite operator of gradient and matrix A^{T} like composite differences. An isotropic composite Laplacian is obtained from $L = A^{\mathrm{T}}A$. determination of $m \times m$ constitutive matrix C become the diagonal matrix with weights of every edge and every the diagonal. A constitutive matrix should be determined like describe a matrix, in this experience that it describes a parameter weight the inner product from vector space of equation determines each edge point. In this experience, the composite Generalizations of Laplacian to the composite Laplace-Beltrami operator via $L = A^{\mathrm{T}}CA$. The composite formula from formula Integral Dirichlet can be an equation as

$$D[x] = \frac{1}{2}(Ax)^{\mathrm{T}}C(Ax) = \frac{1}{2}x^{\mathrm{T}}Lx = \frac{1}{2}\sum_{e_{ij}} w_{ij}(x_i - x_j)^2 \quad (16)$$

when we divided the point become two sets, V_m describing the prelabeled point seed and V_u describing the unseeded point, then $V_m \cup V_u = V$ and $V_m \cap V_u = \emptyset$. We can be an equation(16) as:

$$D[x_U] = \frac{1}{2}[x_M^T\ x_U^T]\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} x_M \\ x_U \end{bmatrix} = \frac{1}{2}(x_M^T L_M x_M + 2x_U^T B^T x_M + x_U^T L_U x_U), \quad (17)$$

where x_M and x_U the probability is appropriate for the seeded and unseeded pixels. For every unseeded pixel,

$$v_{ui} \in V_u, X = [x_u^j, 0 < j \le K]^T, \quad (18)$$

where the probability of random walker x_u^j, started at v_{ui} will achieve each the j^{th} seeded pixel, $v_{m_j} \in V_m$, and the total number of seed K obtained from system. For every seed pixel

$$v_{m_i} \in V_m, M = [\delta_i^j, 0 < j \le K]^{\mathrm{T}}, \quad (19)$$

distinguish $D[x_U]$ with merge to x_U and finding critical points as a result

$$L_U x_u = -B^{\mathrm{T}}x_M. \quad (20)$$

Equation(20) describes that the system from the linear formula with $|Vu|$ is not defined. The solution is obtained by combination from problem of Dirichlet for the label j then can be found by equ-

TABEL 1
AVERAGE VALUE OF COMPARISON METHOD

	Watershed Region Growing	Proposed Method
ME	1,781	1,407
RAE	6,767	4,075
MHD	0,180	0,100
Time	1,715	0,957

ations. For all seed, where X have K columns obtained by every x_u and M have many columns obtained by each x_m. The probabilities at any point will become to unity, i.e.

$$\sum_j x_u^j = 1 \; \forall \; v_{u_i} \in \; V_u, \tag{21}$$

where systems of sparse linear $K - 1$ must be solved, where the number of seed described as K.

3. Results and Analysis

We do an experiment by comparing between the proposed method and exiting method, seeded region growing in the watershed region (Fadllullah et al., 2015). The experiment does in specification OS windows 10, 8 GB RAM, and Intel processor core i3. Figure 4 shows segmentation process on tuna image where the image of *a1*, *b1*, and *c1* are input images. Image of *a2*, *b2*, and *c2* are images output images from the existing method Watershed Region Growing. Image of *a3*, *b3*, and *c3* are the image result from proposed method. Image of *a4*, *b4*, and *c4* are groundtruth images created manually using Adobe Photoshop software. Figure 4 shows that our proposed method results better than Watershed Region Growing that has noise around the object.

Evaluation of segmentation result using seeded Random Walker is calculated by ME (Misclassification Error), RAE (Relative Area Foreground Error), and MHD (Modified Hausdorff Distance) with the following equation:

$$ME = 1 - \frac{|B_o \cap B_T| + |F_o \cap F_T|}{|B_o| + |F_o|}, \tag{22}$$

where B_0 and F_0 is background and foreground of the groundtruth image. B_T and F_T is background and foreground of image segmentation result.

$$RAE = \begin{cases} \dfrac{A_o - A_T}{A_o} & jika \; A_T < \; A_o, \\ \dfrac{A_T - A_o}{A_T} & jika \; A_T \geq \; A_o, \end{cases} \tag{23}$$

where A_0 is the area from the input image, A_T is an area of segmentation result.

$$MHD(F_0, F_T) = \frac{1}{|F_0|} \sum_{f0 \epsilon F0} \min fT \epsilon FT \; ||f0 - fT|| \tag{24}$$

where F_0 and F_T is pixel area on the input image and a pixel area on image segmentation result.

Based on Table 1, this research concluded that proposed method has accuracy and runtime more effective than the existing method. Random walker method can robust to noise and reduce the time consuming in the segmentation of tuna. Preprocessing on the first stage can produce a better performance of segmentation using contrast enhancement, image sharpening, and Gaussian filtering. In our experiment, a parameter of β in Dirichlet problem formula determinate manually. We need to optimize the β value to getting the optimum of segmentation result by a seeded random walker.

4. Conclusion

Segmentation of tuna image can use seeded random walker method with its parameters determined. Determination of seed automatically on the random walker is obtained in region watershed by establishing total optimum region will be used. Performance of tuna segmentation using seeded random walker is evaluated by Misclassification error (ME), Relative area foreground error (RAE), and Modified Hausdorff Distance (MHD) with average value sequentially 1,407, 4,075, and 0,100. The average value of time execution for each image is 0,957 second. Average value of evaluation method is lower than segmentation using seeded region growing. It shows that the proposed method has better performance and effectiveness than the existing method.

References

[1] I. Widyastuti and S. Putro, "Analisis Mutu Ikan Tuna Selama Lepas Tangkap," *Maspari J.*, vol. 1, pp. 22–29, 2010.

[2] A. Fadllullah, A. Z. Arifin, and D. A. Navastara, "Segmentasi Citra Ikan Tuna Menggunakan Gradient-Barrier Watershed Berbasis Analisis Hierarki Klaster dan Regional Credibility Merging," *J. Buana Inform.*, vol. 7, no. 3, pp. 225–234, Jul. 2016.

[3] W. A. Saputra, "Penentuan Otomatis Seeded Region Growing Pada Region Watershed Untuk Segmentasi Citra Ikan Tuna," 2017.

[4] W. A. Saputra and A. Z. Arifin, "Seeded Region Growing pada Ruang Warna HSI Untuk Segmentasi Citra Ikan Tuna," *J. Infotel*, vol. 9, no. 1, p. 56, Feb. 2017.

[5] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783,

2006.

[6] L. Grady, "Multilabel Random Walker Image Segmentation Using Prior Model," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 763–770, 2005.

[7] S. Ram and J. J. Rodriguez, "Random walker watersheds: A new image segmentation approach," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 1473–1477, 2013.

[8] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov Random Walk for Image Segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.

[9] L. Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, 1991.

[10] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, Jul. 1994.