International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Modified SIFT Algorithm for Image Feature Detection

Syama K Nair[1], Ragimol[2]

[1]Mtech Student, Sree Buddha College of Engineering, Pattoor, Nooranad, Pandalam, Kerala, India

[2]Professor, Sree Buddha College of Engineering, Pattoor, Nooranad, Pandalam, Kerala, India

Abstract: ***Image feature detection is the fundamental task in most of the video analytic and computer vision systems. The feature keypoints from the images taken at different instant is detected in these systems. To most of the embedded systems, the challenge is its large complexity in computation. This computational complexities and time consuming properties are eliminated by proposing a new feature detector based on SIFT algorithm. Efficient image feature detection and matching is a fundamental problem in object recognition, image indexing, visual localization etc. The thesis work proposes a new FPGA-based embedded architecture for image feature detection. In this the modified Scale Invariant Feature Transform (SIFT) feature detector is used to reduce the utilization of FPGA resources. An optimized Gaussian filtering is undergone in this design for reducing the memory utilization. The corresponding key-points obtained can be either used in an embedded application or accessed via Ethernet for remotely computer vision applications.***

Keywords: FPGA, SIFT detector, Optimized Gaussian Filter, Keypoints, Video analytics, embedded architecture.

1. Introduction

Video analytics and computer vision systems are the wide area which focusses on the computerized processing of video sequences. Most of the computer vision applications such as object recognition, robot navigation, image indexing, visual localization etc. are based on the analysis of video streams. These all processes extracts events and data from a pre-recorded video. But in all the applications real time performance is a crucial demand due to its computational complexities. Several feature detection algorithms are already implemented for these applications. The feature detection methods have been discussed in the literature are based on pure software implementations. Their complexities make them far away from a hardware implementation to meet the real time demands. The proposed work focusses on the pure hardware implementation of Scale Invariant Feature transform (SIFT) algorithm for image feature detection or the keypoint extraction. A modified version of SIFT is proposed with an optimized Gaussian filtering operation.

In the literature, many different local feature detection methods are proposed. They are Harris corner edge detector [1], SIFT [2], SURF [3] etc. Among these SIFT is the most efficient feature detection method due to its invariant properties. It has several advantages over the other detection methods based on the invariance in rotation, scaling, illumination changes etc. It is noted that it is very difficult for a software evaluation process for detection using SIFT. However, recently several studies are done for an efficient hardware implementation used in real time analysis [4], [5]. Many works are implemented to accelerate the detection process of SIFT algorithm, such as [6]. The major challenge in integrating the feature detection part into a chip is its operational complexity.

There are several modified versions of this SIFT algorithm proposed in the literature. All these modifications are arises in order to speed up the SIFT algorithm to meet the real time demands. SIFT feature descriptors are also modified by applying dimensionality reduction to original SIFT descriptor, such as principle component analysis (PCA-SIFT). Various quantization techniques are applied to convert the floating point co-ordinates into integer codes. From all these modifications a variance in SIFT descriptor is emerged, such as Binary Robust Independent Elementary Features (BRIEF) descriptor. It yields better recognition accuracy compared to other descriptors and also reduces the memory demanded for feature storage during description method.

In the proposed work the whole detection module is implemented on a single FPGA by considering power, memory and real-time constraints of an embedded system. It combines the stability and repeatability of SIFT detection by optimizing the design to meet the real time requirements. The proposed design is known to be one of the fastest feature detection module with less resource utilization on FPGA.

2. Literature Review

In general, image processing is mainly carried out by two important approaches. They are the feature based approach and the intensity based approach. The first method is based on the local features extracted from the image taken, such as lines, points, edges etc. While the later method is based on the correlation among intensity patterns in images. For most of the real time applications the feature based approach is recommended due to its good result for image detection.

One of the first key points detection algorithms which has been widely used was developed by Harris and Stephens is the Harris corner detector [1]. It achieved great improvement in terms of noise immunity and invariance to intensity change and rotation. Later Lowe proposed the scale-invariant feature transform (SIFT) algorithm [2] which has been considered as one of the most robust approaches.

Zhong *et al.* [4] presented a design for SIFT feature detection and description based on FPGA+DSP architecture for 320 × 256 images. SIFT has shown a great success in various computer vision applications. However, its large computational complexity has been a challenge to most embedded implementations. Bonato *et al.* [5] presented an FPGA-based architecture for SIFT feature detection. In this design a parallel hardware architecture for image feature detection based on the Scale Invariant Feature Transform algorithm was proposed. This work presents an optimized robotic control system on-a-chip for an embedded platform.

Huang *et al.* [6] presented a hardware-based SIFT feature extraction architecture, which can extract SIFT features for 640 × 480 images within 33 ms when the number of feature points to be extracted is fewer than 890. In [7] modified SIFT algorithm for high speed feature detector was presented. This system took 31 ms to detect multiple features from 640×480 images. Schaeferling *et al.* [8] proposed a SURF based object recognition system on a single FPGA. This system contains an array of difference elements (DEs) to overcome the irregular memory access behavior exposed by SURF during the computation of image filter responses.

In [10] proposed a single FPGA-based design to detect and match visual feature in real time for real-life applications. Here SIFT is used for image feature detection by considering the stability and repeatability of the image matching system. In order to achieve real-time performance, replace the SIFT descriptor by the BRIEF descriptor.

3. Proposed Method

The main purpose of this design is to provide an FPGA implementation of SIFT feature detection module for image processing applications. Feature detection is carried out by one of the most efficient feature detection algorithm, i.e., Scale Invariant Feature Transform algorithm. This design is based on two stages of processing. It includes scale space extrema detection and keypoint localization. Here we are considering the reduction in complexity of software implementation by using Xilinx system generator platform with Verilog HDL coding. The input image is driven from a digital camera for real time implementation and the pixel values are recorded for the key point detection purpose. This is an optimized pure software implementation of feature detector using the Xilinx ISIM simulator with Verilog coding. The process diagram of the proposed system is shown in Fig 1. Here the modification is done in the Gaussian filter section by reducing the resource utilization and thus produce an optimized design for feature detection in SIFT algorithm.

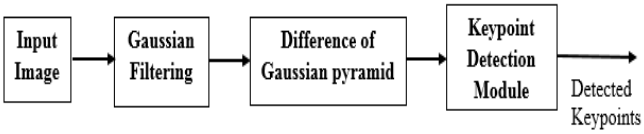

Figure 1: Process Diagram of proposed system

4. Scale Invariant Feature Transform

Features of an image is also known as its keypoints. These features have several peculiarities such as repeatability, distinctiveness and quantity. It means there should be a large number of features (hundreds or thousands) in a single image that are differentiated in a large database with same point of interest should present at various viewpoints. These feature points must be invariant with scale, rotation, illumination, noise occurred by changing of viewpoints.

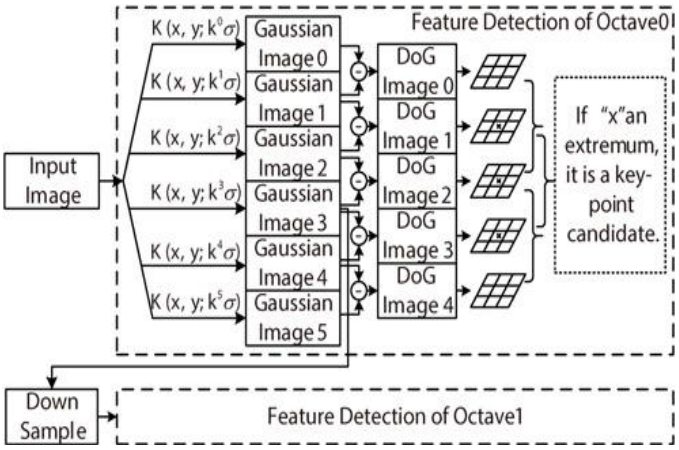

Figure 2: Block Diagram of SIFT feature detection with two octaves

Detection of these features is the fundamental task in video analytical applications. Scale Invariant Feature Transform (SIFT) is an efficient detection algorithm with good results. This process consists of Difference of Gaussian pyramid construction and stable keypoint localization as shown in Fig 2.

4.1 Difference of Gaussian pyramid construction

The main aim of the SIFT feature detector is to determine the variations in the local features of an image with sufficient visual differences. They are mainly focused as the candidate for keypoints. For finding these candidates we have to construct a Difference of Gaussian (DoG) pyramid. In this construction first a Gaussian filtered image is obtained by convolving the input image *I(x, y)* with the Gaussian kernel *K(x, y; σ)* represented in (2) where σ refers to the scale of the Gaussian kernel. Here the 2-D convolution is carried out. The Gaussian filtered image is given as in (1).

$$G(x, y; \sigma) = conv2\,(I(x, y), K(x, y; \sigma)) \quad (1)$$

where *conv2 (,)* refers to the two dimensional convolution and

$$K(x, y; \sigma) = (1/2\pi\sigma^2)\, e^{-(x2+y2)/2\sigma2} \quad (2)$$

The Difference of Gaussian D(x, y; σ) between two Gaussian filtered images are given by

$$D(x, y; \sigma) = G(x, y; k\sigma) - G(x, y; \sigma) \quad (3)$$

where *k* denotes the constant multiplicative factor.

The DoG pyramid construction is based on different octaves. Each octave consists of images with same resolution. There are six scales in each octave. Here we are considering two octaves, by down sampling the input image by 2 and given it

as the source image of the next octave for feature detection. In each octave the image resolution gets halved which makes it possible to select more refined keypoints as possible.

The Gaussian filtering of the source image is done in order to make it smoothened in each of the scales present in each octaves. This smoothening makes the image free from noise and provides accurate results. The source image can be read and written as pixel-by-pixel values in row and column format. After the difference of Gaussian construction image gets blurred and fed to the detection module for searching of the keypoints as features. Gaussian filtered images from successive scales are subtracted to build up the difference of Gaussian pyramid. In this Gaussian filtering section several adders and multipliers are used. Here we have designed an efficient design with reduction in the multiplier and adder sections utilizing less area during FPGA implementations. Instead of using normal multipliers, array multipliers are used.

4.2 Stable Keypoint Detection

After obtaining the DoG pyramid pixel, we have to identify its local maxima and minima by using a stable keypoint detection module. This was detected by comparing the 3×3 pixels with its 26 neighbors and treat this local maxima and minima as candidates for keypoint detection. Keypoints are usually a circular image region with better visual appearance and orientation. Here we are considering these keypoints as points represented in pixel values.

Keypoint detection method includes extreme value detection, edge response rejection and low contrast rejection. Sometimes the identified keypoints may be unstable due to low contrast appearance and occurrence at image edges. So we have to eliminate such keypoints by comparing the DoG output with a low contrast value and an edge threshold value from a Hessian matrix construction.

5. Hardware Architecture

In this section the overall architecture of the proposed system is introduced. The SIFT feature detection in real time is done by 1280 × 720 images. In real time applications the input section of SIFT detector is driven by camera. The entire system performance can be accelerated by using the parallel architecture. The detection module provides a constant throughput defined by the number of octaves and scales in the DoG section. In this design we are using 12 Gaussian filters for the two image octaves. The detailed description of each block is discussed in later sections. The architecture of both DoG module and the stable keypoint detection module is coded in Verilog using Xilinx ISE design suite for simulation.

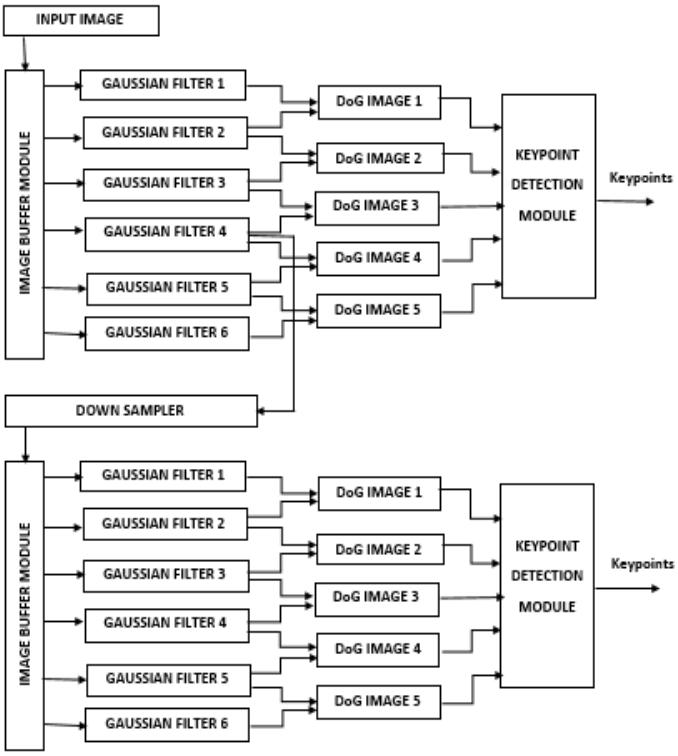

Figure 3: System architecture for SIFT feature detector.

5.1 Implementation of DoG module

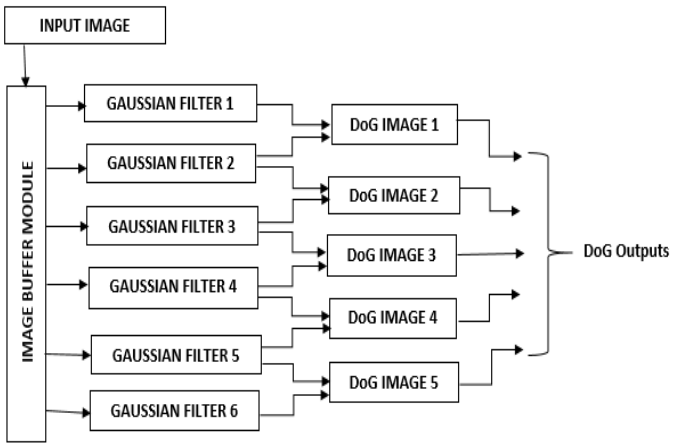

Figure 4: Architecture of DoG module

The difference of gaussian module incorporates the three operations such as image buffering, 2-D Gaussian filtering and image subtraction as in Fig 4. The image buffer module used here is a line buffer section providing a series of delay elements that works in parallel. These line buffers provides sharing in the same octave as in Fig 5. The filtering section is seperable and provides row and column operation for every pixels, which provides better smoothening of the image. Here the conventional multiplication is replaced by array multiplication processes.

The 2-D convolution is more effective by first convolving the gaussian kernel in vertical direction and the second convolution is done in horizontal direction as shown in Fig 6. of [4], which is the traditional filtering approach. In this design the vertical filtering is done first as [10] in contrast to

the traditional approach used in [4]. This will reduce the resource utilization compared to the previous implementations. The SUM used in this filter section is selected as 1024 by considering the characteristics of FPGA, i.e., SUM = $\sum K_i$ = 1024. The gaussian pixel is represented as 8.10 where 8 bits for integer part and 10 bits for fractional part. Instead of division, shift right operation is used since it is more efficient during synthesis.

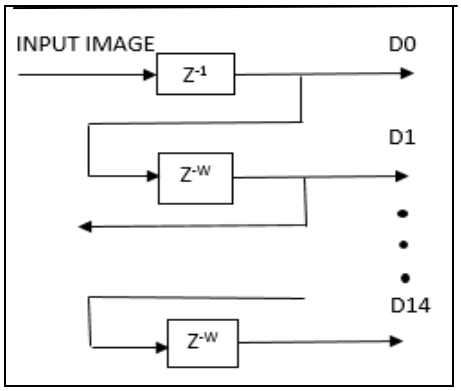

Figure 5: Architecture of the Image Buffer module with delay element where w is the width of gaussian kernel

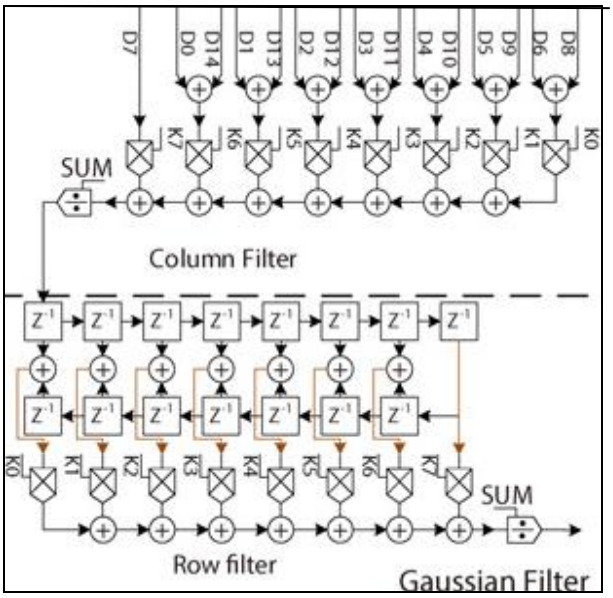

Figure 6: Architecture of 2-D Gaussian filtering with SUM=1024

5.1 Implementation of Key-point Detection module

The feature detection module is a fully parallelized section with four separate modules. They are window generator, extremum detection, edge response rejection and the low contrast rejection modules as explained in [10]. The overall architecture of the stable key-point detection module is shown in Fig 7. Here the window generator module is to generate 3×3 pixel windows from the DoG outputs. The extremum detection module is used for detecting the local maxima by comparing each pixel value with its neighbors. The edge response rejection module is used to eliminate edge features of the image which makes its key-points unstable. This was done by some edge threshold value. The low contrast rejection module is also used to eliminate or reject the less contrast features to make the key-point detection more accurate. Finally all these modules are combined to implement the stable key-point detector section using SIFT. Due to its parallelized design we can detect a large number of features in a defined clock frequency.

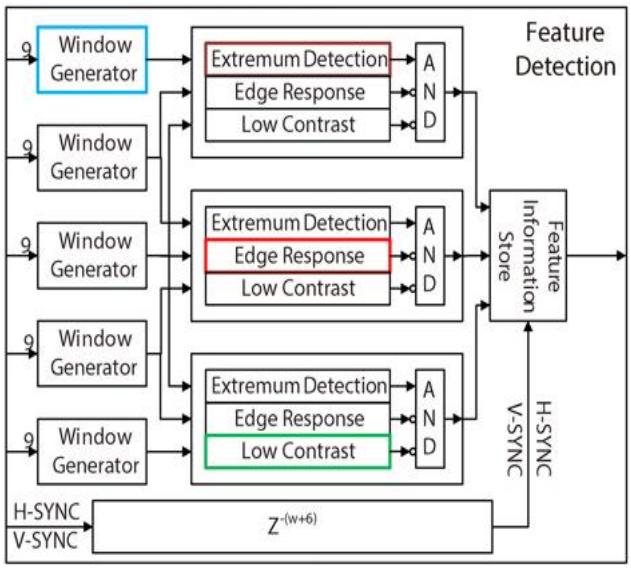

Figure 7: Overall architecture of feature detection module

6. Simulation Results

The design is simulated using Xilinx ISE Design suite 14.2 with Verilog coding language. For the feature detection an image should be given as input. A test input image is converted into corresponding pixel using MATLAB for verification. Fig 8 indicates the simulation output for the line buffer circuit with the image given. It outputs each pixel values one after the other with one clock pulse delay. Fig 9 indicates the difference of Gaussian simulation output with image smoothening by using Gaussian filtering. Fig 10 indicates the key-point detected from the feature detection module which shows that usually the extreme pixel value is considered as the keypoints.

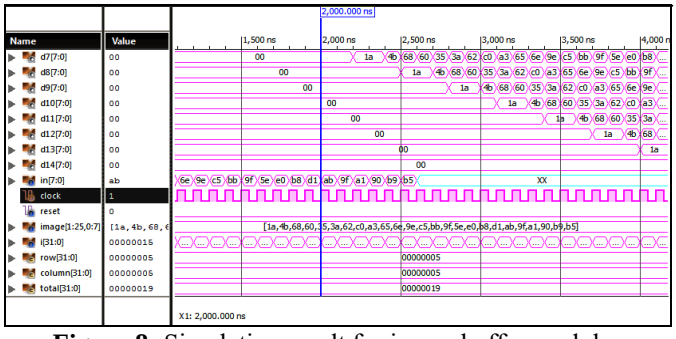

Figure 8: Simulation result for image buffer module

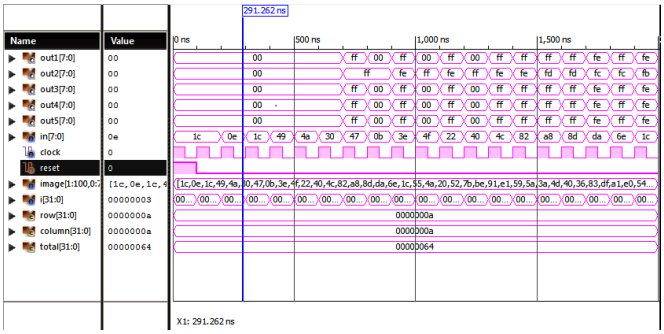

Figure 9: Simulation result for DoG module

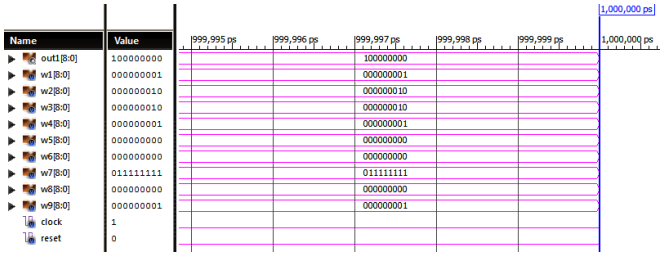

Figure 10: Simulation result for Key-point Detection module

7. Conclusion and Future Work

Most of the video analytical and computer vision applications, images plays a fundamental role. It receives images as inputs and outputs may be either an image or some features of the image indexed as pixel values. A computationally efficient feature detector based on SIFT algorithm is implemented in this design. In this thesis work an FPGA based architecture for image feature detection in real time was designed. By this design a stable and reliable scale space extrema detection is achieved based on SIFT algorithm. This design is focused on the feature detection with invariance to image rotation, scaling, translation and illumination changes.

The future work includes an embedded SoC architecture incorporating the SIFT detector and BRIEF feature descriptor and matching. In this an optimized feature descriptor invariant to scale and orientation is also planned to implement on a single FPGA chip with less resource utilization.

8. Acknowledgment

For the successful completion of my thesis work, I would like to express my deep gratitude towards my guide Prof. Ragimol. She steered me in the right direction with valuable suggestions and continuous assessment in the journey of thesis completion. And I would also like to thank all the reviewers for their esteemed support and encouragement.

References

[1] C. Harris, “A combined corner and edge detector,” in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–152.
[2] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Speeded-up robust features (SURF),” *Comput. Vision Image Understand,* vol. 110, no. 3, pp. 346–359, 2008.
[4] S. Zhong, J. Wang, L. Yan, L. Kang, and Z. Cao, “A real-time embedded architecture for SIFT,” *J. Syst. Arch.*, vol. 59, no. 1, pp. 16–29, Jan. 2013.
[5] V. Bonato, E. Marques, and G. A. Constantinides, “A parallel hardware architecture for scale and rotation invariant feature detection,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 12, pp. 1703–1712, Dec. 2008.
[6] F.-C. Huang, S.-Y. Huang, J.-W. Ker, and Y.-C. Chen, “High performance SIFT hardware accelerator for real-time image feature extraction,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 340–351, Mar. 2012.
[7] L. Yao, H. Feng, Y. Zhu, Z. Jiang, D. Zhao, and W. Feng, “An architecture of optimized SIFT feature detection for an FPGA implementation of an image matcher,” in *Proc. Int. Conf. Field Program. Technol.*, 2009, pp. 30–37.
[8] M. Schaeferling and G. Kiefer, “Object recognition on a chip: A complete SURF-based system on a single FPGA,” in *Proc. Int. Conf. FPGAs ReconFig*, 2011, pp. 49–54.
[9] S. Mitra, *Digital Signal Processing: A Computer Based Approach*, vol. 1, 3rd ed. New York, NY, USA: McGraw-Hill, 2004.
[10] J. Wang, S. Zhong, L. Yan, and Z. Cao *“*An Embedded System- on-Chip Architecture for Real-time Visual Detection and Matching*” IEEE Transactions on circuits and systems for video technology,* vol. 24, no. 3, pp.525-538, March 2014.