Australian Journal of Basic and Applied Sciences, 9(1) January 2015, Pages: 34-40

AENSI Journals
Australian Journal of Basic and Applied Sciences
ISSN:1991-8178

Journal home page: www.ajbasweb.com

Holistic Features for Efficient Age Estimation

[1,2]Ghalib A. Salman Salih and [1]Ghazali Sulong

[1]*Universiti Teknologi Malaysia,Media and Game Centre of Excellence (MaGICX),Faculty of Computing, Skudai, Johor, 81310 MALAYSIA*
[2]*Ministry of Science and Technology, Digital Society Center, Directory of Information Technologies, Baghdad-Jadirya, Iraq*

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form
19 November 2014
Accepted 22 December 2014
Available online 2 January 2015

Keywords:
Age estimation, Face Features, Holistic Features and Statistical Features.

ABSTRACT

This paper presents a new set of holistic features that can be extracted from human face after applying image transform. Result coefficients from discrete cosine transform are statistically inspected to evaluate their performance over different stages of age and over different image samples within each age stage. Coefficients efficiency was measured in term of their changes over different ages and changes over image samples within each specific age; the coefficient with high interclass changes over age and low intraclass changes over sample will be chosen to be an age estimation feature. A set of experiments were conducted on standard FG.net dataset of face images, in addition to private own collected dataset. Encouraging results were yielded in age estimation depending on classification accuracy and mean absolute errors.

To Cite This Article: Ghalib A. Salman Salih and Ghazali Sulong, Holistic Features for Efficient Age Estimation. ***Aust. J. Basic & Appl. Sci.**, 9(1): 34-40, 2015*

INTRODUCTION

Age estimation issue attracted more attention from different researchers in last decade; increasing needs appear for this issue. For security applications, the age of a murdered or an offender is needed to be determined. Some under age persons should be controlled to be preventing from accessing some sites on Internet or from buying wines or cigarettes. The most well-known part of human body that can be used to estimate the age is the face; whereas it contains a lot of aging information, and at same time it's the most captured in the personal or official photos (Mathew G, 2009).

This work provides new efficient features that can be used for automatic age estimation; these features are extracted statistically from the result matrix of discrete cosine transform.

Background of problem:

Different researches had been conducted in age estimation area. These researches are classified into two main types, first type processed the whole face image using image transformations from which frequency domain features can be extracted. Second type researches focused on spatial domain features that belong to the face skin and components; these features are directly extracted from spatial domain face image (Sung, *et al* 2011). John Hatzis (2004) used the most perceptible indication of age progression, which are wrinkles and their measurements; his results were concerning about specific age period for senior adult. A combination between texture and geometric features were proposed to represent age progression (Narayanan and Rama 2004), yet their results had some weakness in age estimation. Karl, *et al*, (2010) proposed set of biometric face features for age estimation; their results suffered from illumination effects. In our previous work Ghalib and Ghazali 2014), we provided a new set of features for age estimation that are robust to illumination and rotation; we depended on some statistical measurements to estimate the age, our estimation were general in classifying the age into three main classes child, adult and senior adult.

In spite of interesting results of age estimation spatial domain features, the success of such estimation are restricted to some factors such as the efficiency of face detection, and correctly chosen features that can represent an age period; another factor is features robustness to the scaling, rotation and illumination (Tao, *et al* 2012). Those restrictions increased the importance of frequency domain features; on the other hand, image transformations deal with whole image and produce coefficients matrix with huge number of coefficients; thus, dimensionally reduction methods are needed for choosing discriminative features (Mathew 2009), most crucial factor for best estimation is to choose the most efficient coefficients that represent robust features.

Corresponding Author: Ghalib A. Salman Salih, Universiti Teknologi Malaysia,Media and Game Centre of Excellence (MaGICX),Faculty of Computing, Skudai, Johor, 81310 MALAYSIA

There are two types of image transforms, statistical and deterministic were used to extract the features and avoid redundant coefficients. Statistical approaches were widely used in age estimation issues for their ability of reducing the correlation between data; this leads to enhance the differences over different age stages (inter class). The most well-known statistical approaches are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) (Yun, *et al* 2010).

In spite of the wide usage of deterministic transforms in different classification problems, their appearance in age estimation was significantly lower than statistical approaches. Basis vectors in deterministic transforms are not related to the dataset; thus, their data have undesired correlation, which increases intra class differences making them more suitable for face recognition and other identification issues rather than age estimation problems. Discrete cosine transform is more suitable for age estimation rather than other deterministic transforms like Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) because of its ability to reduce the correlation between data (Jing and Zhang 2004), (Hafed and Levine 2001) and (Chen, *et al* 2005).

Statistical methods were used to reduce redundant data from DCT results. Shuicheng, *et al* (2008) used expectation maximization of Gaussian density distribution and covariance matrix for image representation and dimensional reduction; their results suffered from high Mean Absolute Error (MAE) values in estimated ages; for male faces, MAE was (7.82), and it was (8.53) for female faces. Other approach combines Radon and DCT transformations to extract texture features, which was fused with the appearance features extracted from Active Shape Model (ASM) for age estimation. Although their MAE values (4.18), their work suffer from high complexity computations and bias image sample; they used (710) images for ages between (0 - 19) and (292) image for all other ages.

In previous works, coefficient selection of DCT didn't gather enough attention in spite of its importance, while improper coefficient reduces the efficiency of the feature it represents; most of previous coefficient selection methods provided efficient features for Identification issues like face, palmprint, gender and fingerprint recognition problems (LU, *et al* 2010), (Kekre, *et al* 2010) and(Amornraksa and Tachaphetpiboon 2006). In this paper a new method is proposed for extracting efficient features for age estimation; these features can be represented by selecting proper coefficients that have the highest change over different ages (interclass) and the smallest changes within each age (intraclass).

DCT Coefficients and Features:

DCT is considered as one of the most powerful image transformation; it produces holistic face features that can be extracted from three types of coefficients, low, middle and high frequency bands see Fig1. Using combinations of these bands, many types of features had been proposed and extracted (Mathew 2009) and (Jing and Zhang 2004). As Low frequencies band correlates to the illumination conditions, and noise is represented using high frequencies, extracted features from DCT can be robust to illumination and noise by controlling high and low frequency bands (Chen, *et al* 2005). Most significant information can be found in the middle frequency band; it also constructs the basic useful information to represent the image (Jiang and Feng 2007). Although middle frequencies coefficients are seemed to be the best band that can produce discriminative features, low and high frequency band can also provide significant features if they are manipulated properly.

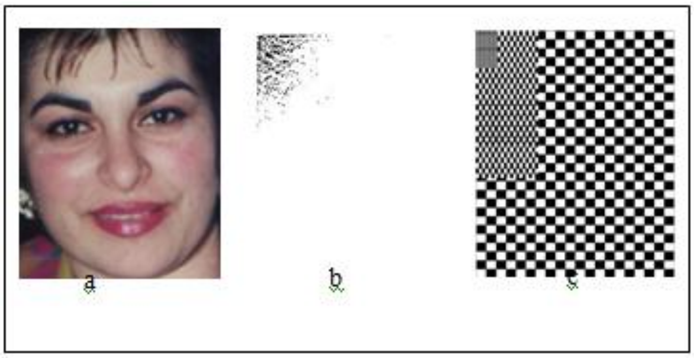


Fig. 1: Applying DCT on a face image where, (a) original image, (b) DCT result and (c) three bands of DCT

Coefficients Selection:

Obviously From previous sections, different coefficients in different arrangements can provide efficient features for different problems; the most crucial factor in providing discriminative features for any classification problem is the selection of suitable coefficients that represent these features.Pan *et al*. (Pan, *et al* 2000) depended on reconstruction error as a base to choose coefficients that minimize the error; this can be more related to face recognition or compression than age estimation. Examination of DCT result as group by group to select discriminant coefficients yielded better results (Jing and Zhang 2004), yet their results still suffer from losing discrimination power because of the effects of lower discriminate coefficients when they are combined in the same group (Dabbaghchian, *et al* 2008). Matthias Steiner [2010] scaled trained and tested image into 64×64 pixels image to avoid extra size of features vector, which actually leads to interferences between extracted features; he combined DCT with ASM to enhance his results. Although he yielded (1.99) MAE in (0 – 9) ages, MAE increased to reach (28.66) at the minimum in (60 – 69) ages; this may indicate that his features are more suitable for childhood than other ages.

***Methodology*:**

This work analyses DCT coefficients to determine the most discriminative ones over age progression. The performance of each coefficient is analyzed using different ages with set of images for each age class. The output of this analysis is a set of positions of the coefficients that should be chosen from each DCT result matrix to be robust features for age estimation.

Efficient coefficient for representing age progression must have two properties, large differences (diff) over age progression (interclass) and at the same time low diff over different images within the same age interval (intraclass); these two types of change can be combined in a ratio of interclass changes to the intraclass changes, which should yield high value for discriminative coefficient; low value for this ratio means that this coefficient has weak discrimination power for age estimation. This ratio can be shown as:

$$Efficiency = \frac{diff^{\,ages}}{diff^{\,samples}} \tag{1}$$

Where:

Eff:is the efficiency of the coefficient.

diffage : is the differences over age.

diffsample: is the differences within each age class and over difference samples.

One of the most well-known measurements that represent the differences between a set of values is the Standard Deviation SD, which will be used to measure diffsample, while an adaptive form of SD will be used to measure diffage. Assuming the input image of N×N size, the result matrix from applying DCT will be:

$$M = \begin{bmatrix} m_{11} m_{12} m_{13} & \ldots\ldots\ldots\ldots\ldots & m_{1N} \\ m_{21} m_{22} m_{23} & \ldots\ldots\ldots\ldots\ldots & m_{2N} \\ \vdots\vdots\vdots\vdots & & \\ m_{N1} m_{N2} m_{N3} & \ldots\ldots\ldots\ldots\ldots & m_{NN} \end{bmatrix}$$

Assuming the available training database is consist of (a) of age classes and (s) of different samples for each age class, each coefficient of DCT result matrix for each age class and different sample m_{ij}will be taken to construct the testing matrix T_{ij}, each element of this matrix t_{hk} (h: 1, 2, 3… s, k: 1, 2, 3… a) corresponds the element m_{ij} taken from sample h of the age class k.

$$T_{ij} = \begin{bmatrix} m_{ij}(11)\ m_{ij}(12) & \ldots\ldots\ldots\ldots\ldots & m_{ij}(1a) \\ m_{ij}(21)\ m_{ij}(22) & \ldots\ldots\ldots\ldots\ldots & m_{ij}(2a) \\ & \vdots\vdots\vdots\vdots & \\ m_{ij}(2s)\ m_{ij}(2s) & \ldots\ldots\ldots\ldots\ldots & m_{ij}(sa) \end{bmatrix}$$

SD is used to measure differences of different image samples within each age class using the well-known form:

$$SD = \sqrt{\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}} \tag{2}$$

In this from values deviation is measured about a center, which is μ; it is computed as:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

This form can be affected by extreme values (Nagabushanam, *et al* 2005); this is expected in age estimation problems where some of available ages in datasets have available ages in some intervals more detailed than other intervals; this may affect the results and yield unreal deviation from the center point.

Another measurement can be used instead of μ; Median (ɱ) is more suitable for sorted values such as ages and not affected by extreme values. It is simply the middle value of odd number of sorted values and the average of the two middle values for even number of sorted values (Nagabushanam, *et al* 2005). From the values of μ1 that represent the average after ignoring the extreme values, obviously, they are nearer to the values of ɱ, which is more stable than ɱ. An adaptive form of SD that measures values deviation from ɱ will be:

$$\Omega = \sqrt{\sum_{i=1}^{n} \frac{(x_i-\text{ɱ})^2}{n}} \tag{4}$$

For each column of the testing matrix T_{ij}, μ and then SD will be computed; SD for each column represents the differences over different samples within each age class; for general view of the differences over samples for each coefficient, the average of SD values will be computed as:

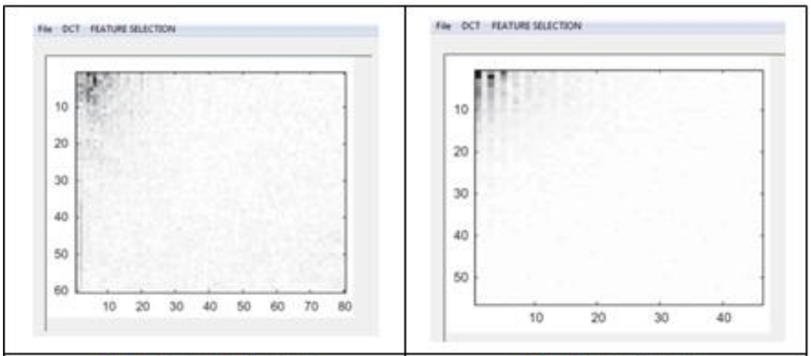
$$\delta_{ij}^{sample} = \frac{\sum_{k=1}^{a} SD_k}{a} \tag{5}$$

This measurement can represent the intraclass differences of each coefficient over image sample. For each raw of T_{ij}, ɱ and then Ω will be computed; the average of Ω values will be computed as:

$$\delta_{ij}^{age} = \frac{\sum_{h=1}^{s} \Omega_h}{s} \tag{6}$$

This measurement summarizes the interclass differences over age progression for each coefficient in T_{ij}. We can rewrite Eq. 1 as:

$$m_{ij} = \frac{\delta_{ij}^{age}}{\delta_{ij}^{sample}} \tag{7}$$



(a): Standard Dataset	(b) Private Dataset

Fig 2: Illustrate m_{ij} efficiencies in standard and private datasets.

The results of this equation construct the matrix of efficiencies *M;* in which, best coefficient should yield highest differences over different ages and smallest differences over different images within each age class, therefor, best coefficients should yield highest values for m_{ij}. The positions of best coefficients will be stored; from each DCT result matrix, coefficients corresponding to these positions will be extracted as age estimation features. Middle frequency coefficients of DCT result matrix yielded the highest efficiency values in both standard (FG.net) and private (own collected) datasets, while high frequency coefficients yielded lowest efficiency values in both two datasets; see Fig 2, which illustrates efficiency values for the standard and private datasets.

Low frequency coefficients yielded higher efficiency in private dataset than the standard one; there are many reasons for this variation between these datasets, most significant of them is the illumination effects that affect low frequency coefficients. In addition, private dataset has some better properties than standard one, which affect the results generally such as:

a. Higher quality
b. Higher number of images for each age class
c. More detailed age classes
d. Stable time period between age classes

Therefore, these properties should be considered in order to utilize low frequency coefficients for standard datasets; proper preprocessing should be applied to low quality images.

Where coefficients with high values are proposed, a threshold between high and low values should be determined. For high quality images or properly manipulated low quality images, low frequency coefficients can be used, and then a threshold need to be determined to exclude high efficiency coefficients; without suitable preprocessing, low frequency coefficients should be excluded for low quality images. High threshold value ensures that best coefficients are to be chosen, but some important coefficients may be lost. On the other hand, low threshold value ensures all significant coefficients are chosen, yet insignificant coefficients may be chosen also; this may lead to the threat of illumination effects.

RESULT AND DISCUSSIONS

This work was conducted using two types of dataset, FG-net standard dataset which is consist of 1002 images for 81 males and females and own collected dataset that consist of 5800 images for 28 males and 31 females. Most of the images in our own collected dataset are of high quality and they were taken in stable time periods such as, daily, monthly and yearly images.

Standard dataset in this work was divided into (600) image for training, (368) images for testing and (34) extremely distorted images were excluded. Private dataset was divided into (3000) images for training and

(2800) images for testing. Using Weka 3-6-9 software, Support Vector Machine (SVM) was used for classification; zigzag, zonal masking and genetic search algorithms were used to search for alternative features as benchmarking. All features were normalized into (0- 1) range.

Controlling threshold value adjusts the number of selected coefficients; with high threshold value and little number of features, Classification Accuracy (CA) was low because of losing significant coefficients. Increasing threshold value enhanced CA level, yet at the same time, increasing chosen coefficient should not exceed to inefficient features to avoid decreasing features efficiency, see Fig3.

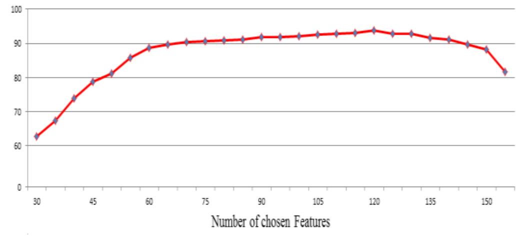


Fig. 3: Classification accuracy with different number of chosen features.

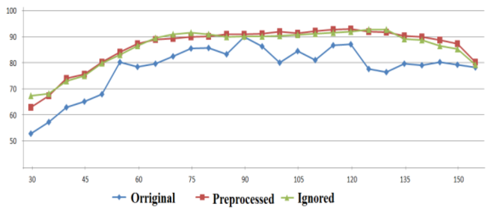


Fig. 4: Effects of low frequency coefficients.

Low coefficients has significant effects with high quality images, while in variant image qualities, they decrease the robustness of the features. In uncontrolled image quality datasets, low frequency coefficients should ignored or proper preprocessing should be applied. See Fig 4 which illustrate that CA for features with low frequency lower than for preprocessed or without low frequency; in most of cases, preprocessed was better than the others.

CA for features chosen used median and adapted form of SD for age changes was in general better and more stablethan using mean and SD; which because Mean is more affected by extreme values than Median, see Fig 5.

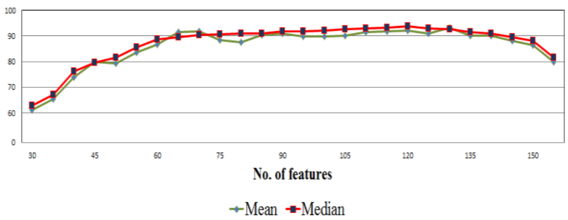


Fig. 5: Effect of using Mean and Median.

In addition to CA Mean Absolute Error (MAE) was used to measure the deviation of misclassified images.Table 2 illustrates that proposed features yielded less errors than other DCT selected features.

Results showed that proposed features are consistent for all age classes, and then it can be used for age estimation in all age periods. Best MAE of previous works in age estimation (4.97) was yielded by Yuyu, *et al* [27], which higher than our work. Results are summarized in table 3. It shows that using proper number of features our proposed features yielded best results than other algorithms that searched for proper features. Although our approach wasn't the best with low number of selected features, but, it was better than the others most of times and also in the average results

***Conclusion and Future Works*:**

According to the previous results, proposed features have perceptible efficiency and made significant improvement in the results of other techniques. Proposed features have stable results over different age intervals, yet they need suitable number of chosen coefficients to give robust features to illumination. In addition to

Australian Journal of Basic and Applied Sciences, 9(1) January 2015, Pages: 34-40

significant classification accuracy, proposed features yielded low and consistence error rates. For future works, low frequency coefficients may be fused in suitable manner in order to decrease illumination effects. This approach can be applied to other image transformation such as Discrete Wavelet Transform DWT,Karhunen–Loeve transform KLT or Discrete Fourier Transform DFT.

Table 2: Classificationaccuracy and MAE: proposed Methods verses (Matthias 2010) and (Shuicheng, *et al* 2008).

Age	Our CA	MAE				
		Our	DCT with Spatial		DCT with AAM	
			Female	Male	App1	APP2
0- 10	92.89%	5.15	7.2	4.51	1.99	3.09
11- 20	93.85%	4.23	4.55	5.43	4.04	3.69
21- 30	93.87%	3.12	4.79	6.53	7.12	3.8
31- 40	94.73%	3.78	8.43	8.27	9.37	8.89
41- 50	93.52%	3.22	9.49	11.62	13.62	16.76
50- 59	93.93%	3.41	7.42	10.44	21.79	24.89
60- 69	93.76%	3.92	12.12	7.5	28.38	35.38
All	93.79%	3.83	7.71	7.76	12.33	13.79

REFERENCES

Amornraksa, T. and S. Tachaphetpiboon, 2006. “Fingerprint recognition using DCT features”, : Electronics Letters, 42(9-27): 522–523, EIT Digital Library.

Chen, W., M.J. Er, S. Wu, 2005. “PCA and LDA in DCT Domain”, Pattern Recognition Letters, 26: 2474–2482.

Dabbaghchian, S., A. Aghagolzadeh, M.S. Moin, 2008. “Reducing the Effects of Small Sample Size in DCT Domain for Face Recognition”, International Symposium on Telecommunication (IST).

Dr. Kekre, H.B., D. Sudeep Thepade and Tejas Chopra, 2010. “Face and Gender Recognition Using Principal Component Analysis”, International Journal on Computer Science and Engineering, 02(04): 959-964.

Freund, J.E., “Modern elementary statistics”, Prentice-Hall.

Ghalib Salman Salih, A. and Ghazali Sulong, 2014. “Localized Statistical Features Representing Face Vitality for Age Estimation. Aust. J. Basic & Appl. Sci., 8(13): 698-703.

Hafed, Z.M., M.D. Levine, 2001. “Face Recognition Using the Discrete Cosine Transform”, International Journal of Computer Vision, 43(3): 167–188.

Jiang, J. and G. Feng, 2007. “Robustness analysis on facial image description in DCT domain”, Electronics Letters, 43(24): 2007.

Jing, X.Y., D. Zhang, 2004. “A face and palmprint recognition approach based on discriminant DCT feature extraction, IEEE Transactions on Systems, Man and Cybernetics, 34(6): 2405–2415.

Jing, X.Y., D. Zhang, 2004. “A Face and Palmprint Recognition Approach Based on Discriminant DCT Feature Extraction”, IEEE Transactions on Systems, Man and Cybernetics, 34(6): 2405–2415.

John Hatzis, 2004. “The wrinkle and its measurement: A skin surface Profilometric method”, Micron, 35: 201–219, ELSEVIER.

Karl Ricanek Jr., MariosSavvides, Damon L. and Gerry Dozier, 2010. “Unconstrained Biometric Identification: Emerging Technologies”, Biometrics Compendium, IEEE.

Lu Leng, Jiashu Zhang, Muhammad Khurram Khan, Xi Chen and Khaled Alghathbar, 2010. “Dynamic weighted discrimination power analysis: A novel approach for face and palmprint recognition in DCT domain”, International Journal of the Physical Sciences, 5(17): 2543-2554.

Matthew, G. Rhodes, 2009. “Age Estimation of Faces: A Review”, Wiley InterScience& Sons, Ltd. Applied Cognitive Psychology, Appl. Cognit. Psychol, 23: 1–12.

Matthias Steiner, 2010. “Study Thesis in Facial Image-based Age Estimation”, faculty of Computer Science, Institute for Anthropomatics Facial Image Processing and Analysis.

Nagabushanam, K., R. Ravanan, V. Varalakshmi and N. Suseela, 2005. "STATISTICS", Tamilnadu University.

Narayanan Ramanathan and Rama Chellappa, 2008. “Modeling Shape and Textural Variations in Aging Faces”, Automatic Face & Gesture Recognition, 2008.FG '08. 8th IEEE International Conference on DOI: 10.1109/AFGR.2008.4813337, IEEE.

Pan, Z., A.G. Rust and H. Bolouri, 2000, “Image Redundancy Reduction For Neural Network Classification Using Discrete Cosine Transforms”, Proceeding of International Joint Conference on Neural Networks (Como, Italy), 3: 149–154.

Shuicheng, Yan, Ming Liu and Thomas S. Huang, 2008. “Extracting Age Information From Local Spatially Flexible Patches”, Acoustics, Speech and Signal Processing, IEEE International Conference on DOI: 10.1109/ICASSP.2008.4517715

Australian Journal of Basic and Applied Sciences, 9(1) January 2015, Pages: 34-40

Sung Eun Choi, YounJoo Lee, Sung Joo Lee, Kang Ryoung Park and Jaihie Kim, 2011. “Age estimation using a hierarchical classifier based on global and local facial features”, Pattern Recognition, pp: 1262-1281.

Tao Wu, PavanTuraga and Rama Chellappa, 2012. “Age Estimation and Face Verification Across Aging Using Landmarks”, IEEE Transactions on Information Forensics and Security, 7(6), IEEE.

Yun, Fu, Guodong Guo and S. Thomas Huang, 2010. “Age Synthesis and Estimation via Faces: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(11).

Yuyu Liang, Xianmei Wang, Li Zhang and Zhiliang Wang, 2014. “A Hierarchical Framework for Facial Age Estimation”, Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2014, Article ID 242846