Research Journal of Computer and Information Technology Sciences ______________________ ISSN 2320 – 6527
Vol. **5(6),** 1-7, August **(2017)** *Res. J. Computer and IT Sci.*

Shot boundary detection using second order statistics of gray level co-occurrence matrix

B. Reddy Mounika* and Ashish Khare
Department of Electronics and Communication, University of Allahabad
reddymounika.b@gmail.com

Available online at: www.isca.in
Received 8th April 2017, revised 10th August 2017, accepted 20th August 2017

Abstract

The readily and easily available nature of capturing devices made enormous amounts of video available in day-to-day life. Processing of such a lengthy video is a time consuming process, therefore researchers have introduced key frames. Key frame in short can be visualized as a frame that represents the information present in entire video shot. Detecting shot boundaries plays a vital role in extracting key fames. The results of shot boundary detection shows effect on performance of further stages of processing, therefore a reliable shot boundary detection task forms corner stone in several applications such as video analysis and summarization, video abstraction and higher contextual segmentation etc. In this article a novel image gray level co-occurrence matrix based technique for shot boundary detection by calculating statistics of the current frame such as homogeneity, energy, correlation, contrast and comparing the same with the next frame. The proposed algorithm successfully detects the shot boundaries by considering the statistics captured by gray level co-occurrence matrix. The method is experimented on animation videos. Performance of the method is evaluated with evaluation parameters boundary recall, accuracy, detection percentage, missing factor. The investigational results demonstrate that the proposed algorithm performs better than state-of-art methods. The results are tabulated, plotted and discussed briefly.

Keywords: Gray level co-occurrence matrix, Contrast, Correlation, Homogeneity, Energy, Shot Boundary Detection.

Introduction

The rapid growth in amount of video data available on the web made researchers to step towards the ways of efficiently managing the available video data. Shot change detection plays a vital role in managing the video content. A shot is basically defined as a continuous recording taken by the camera between start and stop actions. Shot transition detection lies on the principle, visual dissimilarity between two frames is high at a shot change. Shot change is classified as: abrupt change, gradual change. A sudden change of visual information in one frame is known as abrupt shot change. In gradual transition, change in visual information takes place gradually in several frames. Gradual shot change is again classified as dissolve, fadein, fadeout, wipe. The process of gradual replacement of single colored frames by multiple colored frames is known as fade in. In fade out the reverse process takes place, colored frames are gradually replaced by the single color image. In wipe the frames of current shot are replaced by the frames of next shot travelling from the right side of the current frame till the whole frame of current shot is replaced by the next shot. An efficient shot detection algorithm should detect abrupt and all types of gradual transitions. Camera and object motion made shot detection a challenging task.

In the present article shot transitions are detected by calculating the difference between homogeneity, energy, correlation and contrast of the current frame and next frame. The difference then subjected to threshold to detect shot transitions. The proposed algorithm is tested on various kinds of videos and the performance is calculated.

Rest of the paper is organized as follows: Introduction is followed with related works in the field of shot detection. The basics of gray level co-occurrence matrix, the proposed algorithm, experimental results and evaluations were presented and discussed extensively in consecutive sections respectively. Finally, the concluding remarks with significant applications are presented.

Literature survey

A unified model has been formulated by Partha Pratim Mohanta et al. to detect shot transitions[1]. Frame transition and estimation parameters, which are based on scatter matrix of edge strength and motion matrix are utilized by them to classify the frames into no change, abrupt change, gradual change frames using a multilayer perceptron network. Murat Birinci and Serkan Kirnayaz has been introduced a novel and robust shot boundary detection method at significantly low computational costs[2]. They detected visual discontinuities by identifying the objects in the frames using local image features. Mostafa Tavassolipour et al. have used spatiotemporal and compressed domain features to compute visual dissimilarity, which is then applied to support vector machine to detect shot boundaries[3]. ZheMing Lu and Yong Shi have proposed a unified fast SBD scheme using candidate segment selection and singular value decomposition

(SVD)[4]. They performed SVD on candidate segments to form refined feature matrices based on which cut and gradual transitions are detected using pattern matching. G.G. Lakshmi Priya and S. Domnic have been proposed a new automatic shot detection method[5]. They calculated mean, median, maximum intensity value, edge classifier value for each block of the frames to classify edge and non edge pixels based on which shots are clustered. Mendi and Bayrak have employed color histogram difference and self similarity modeling to detect shot boundaries[6]. Shot boundary detection under fire, flicker and explosion (FFE) is one of the under studied areas. Krishna K. Warhade et al. developed an efficient algorithm to detect shot boundaries under FFE using cross correlation coefficient and stationary wavelet transform with the combination of local and adaptive thresholds[7]. Marius Vila et al. have used Tsallis mutual information and Jensen-Tsallis divergence to quantify the similarity between frames based on which the abrupt shot boundaries of a video sequence are detected[8]. Dalton Meitei Thounaojam et al. proposed a shot boundary detection approach using Genetic Algorithm and Fuzzy Logic[9]. They genetic algorithm to calculate the membership functions of the fuzzy system and shot transition classification is done by the fuzzy system. Shuai Wang et al. proposed a shot detection procedure for gastroscopic video summarization[10]. Histogram intersection is used by them to calculate visual dissimilarity based on which shot transitions are identified. Zernike moments have been used by Pablo Toharia et al. to perform shot boundary detection[11]. Krishna K. Warhade et al. developed an algorithm for shot boundary detection in the presence of illumination change, fast object motion, and fast camera motion[12]. They used dual-tree complex wavelet transform to extract structure features and later, spatial domain structure similarity computed between adjacent frames is subjected to threshold for identifying shot boundaries. Feng-feng Duan introduced a shot segmentation method for binocular stereoscopic video based on spatial–temporal feature clustering (STFC)[13]. They used color and brightness of left video frames in temporal domain as well as the depth as features, and differences between them for consecutive frames are calculated and quantified to distinguish shot boundaries. Poonam et al. detected shots using skew, kurtosis, mean obtained from block based image histogram[14]. Thakre et al. detected shots by calculating distance between the wavelet transformed blocks of consecutive frames of the video[15].

In summary, some approaches yielded results with fewer computations and processing time and few approaches offered less computational complexity. However, every method is having its own drawbacks such as reduced accuracy and reliability. In our proposed approach we not only focused on reducing computational complexity and also concentrated on designing an efficient method.

Gray level co-occurrence matrix

Gray-level co-occurrence matrix (GLCM) or the gray-level spatial dependence matrix allows us to examine texture by considering spatial relationship of pixels. GLCM describes about the occurrences of a pixel with a specific value in a specified relationship in the image. Second order statistical parameters such as contrast, correlation, energy, homogeneity etc. can be extracted from GLCM by which information about the texture of an image can be provided.

Let C be GLCM of an image G, and each entry of C is c_{ij}. Then, each C_{ij} denotes the number of times that G(x,y)=I and G(x+1,y+1)=j.

Contrast: Local variations in the gray level co-occurrence matrix are measured by contrast and its mathematical equation is given as:

$$Contrast = \sum_i \sum_j (i-j)^2 C_{ij} \tag{1}$$

Correlation: The joint probability of occurrence of the specified pixel pairs is measured by correlation and its mathematical equation is:

$$Correlation = \sum_i \sum_j \frac{ijC_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{2}$$

Energy: The sum of squared values of all entries in GLCM is given by energy and its mathematical equation is:

$$Energy = \sum_i \sum_j C_{ij}^2 \tag{3}$$

Homogeneity: Measures the closeness of gray levels in the GLCM to the GLCM diagonal and its mathematical equation is:

$$Homogeneity = \sum_i \sum_j \frac{C_{ij}}{1+|i-j|} \tag{4}$$

Proposed Method of Shot Detection

The proposed method of key frame extraction is explored in this section. The basic principle behind shot detection method is visual dissimilarity is more at shot boundary. In the proposed method second order statistics of GLCM: contrast, correlation, energy, homogeneity are used as metrics to determine dissimilarity based on which shot transitions are declared.

Algorithm of Shot detection:
Step 1: Extract frames from video.
Step 2: Convert colored frames of the video into gray color.
Step 3: Obtain gray level co-occurrence matrix of the frame (GLCM).
Step 4: Using GLCM, contrast, correlation, energy, homogeneity are calculated for every frame.
Step 5: Compute the difference between two consecutive frames using values in Step 4 and
MD= CD = Contrast(i) - Contrast(i+1)
CRD = Correlation(i) - Correlation(i+1)

ED= Energy(i) - Energy(i+1)
HD= Homogeneity(i) - Homogeneity(i+1)
Where i and i+1 are two consecutive frames
Step 6: Overall difference is found out by adding all four differences OD.
OD=CD+CRD+ED+HD
Step 7: If OD is positive integer then
(a) Using step 5 and 6 calculate OD for every consecutive frames. Calculate threshold over whole video sequence by taking average of OD and multiplying it with α.
$T=OD_{avg}*\alpha$
α is tuning parameter and can vary from 1 to 10
OD_{avg} = Average of all (OD).
(b) Compare OD with threshold T.
(c) OD (i,i+1) ≥ T then i is the end of previous shot and i+1 is start of next shot.
Step 7: else if OD is negative or not an integer (NaN) then
(a) Replace all NaNs of OD with zeros.
(b) Using step 5 and 6 calculate OD for every consecutive frames. Calculate threshold over whole video sequence by taking average of OD,
$T=OD_{avg}$
OD_{avg} = Average of all absolute values of (OD)
(c) Compare OD with threshold T.
OD (i,i+1) ≥ T or OD (i,i+1)=0 then i is the end of previous shot and i+1 is start of next shot.

Experimental results

The experiment is conducted on 21109 frames taken from animation category videos. The quantitative performance analysis of the proposed method done with the evaluation parameters: boundary recall, accuracy, detection percentage and missing factor. The Table-1 presents total number of frames in the video, stotal number of shots detected by the proposed method and the methods[14,15].

To evaluate the performance of the proposed technique accuracy, boundary recall, detection percentage, figure of merit, missing factor are calculated. Ground truth is built using publicly available tool Virtual Dub[16].

Boundary Recall (BR): According to Debabrata Dutta et al.[17] the boundary precision (BP), boundary recall (BR) is calculated as

$$BR = \frac{TP}{TP + FN} \tag{5}$$

Where true positives (TP) is the number of key frames correctly detected by the proposed method. The false negative (FN) is the number of key frames in ground truth which doesn't present in the segmentation result.

The values of boundary recall lie in the range 0 to 1. Higher values of boundary recall is desired for an efficient method. The results of the proposed method, Poonam et al.[14], Thakre et al.[15] are shown in Table-2. The maximum values of boundary recall are, 0.3185 has been achieved by Poonam et al.[14], 0.2272 have been attained by Thakre et al.[15] where as proposed method shown a maximum of 0.9333 almost an improvement of more than thrice the performance of methods[14,15]. A final conclusion drawn from the observations furnished in Table-2 and the Figure-1 is that proposed method's boundary recall is superior to the methods[14,15].

Table-1: Quantitative presentation of number of shots detected by the proposed method and methods[14,15].

Sr. No.	Total number of frames	Poonam et al.[19]	Thakre et al.[26]	Proposed	Number of Shots in ground
Video 1	2354	476	271	373	192
Video 2	1898	261	121	411	165
Video 3	1621	491	81	142	135
Video 4	2227	128	99	405	195
Video 5	1998	96	171	435	167
Video 6	1588	302	319	388	132
Video 7	3703	466	376	1417	320
Video 8	1155	96	42	246	105
Video 9	1960	164	214	477	181
Video 10	2605	352	82	373	221

Detection Percentage (DP): Detection percentage has been outlined by Poornima and Kanchana[18]. According to them computational equations is given as follows:

$$DP = 100 \times \frac{TP}{TP + FN} \tag{6}$$

High value of detection percentage (DP) is desirable for an efficient method, and it lies in the range 1 to 100. Under animation category the maximum value of DP attained by Poonam et al.[14] is 31.259, Thakre et al.[15] is 26.5625 where as proposed method shown a maximum of 93.33 almost an improvement of more than thrice the performance of methods[14,15]. The values of DP obtained by the state-of-art methods[14,15] and the proposed method are furnished in Table-3 and plots of those values are shown in Figure-2 from which we observe that the value of DP attained by the proposed method is superior to the methods[14,15].

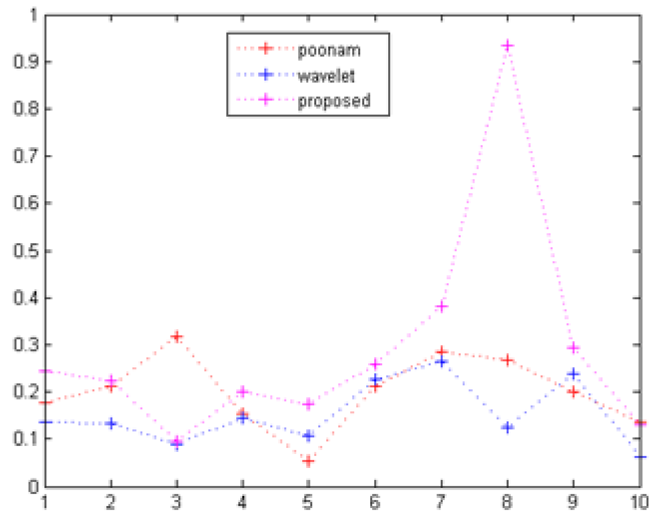

Figure-1: Plot of Boundary recall for Poonam et al.[14], Thakre et al.[15] and proposed method.

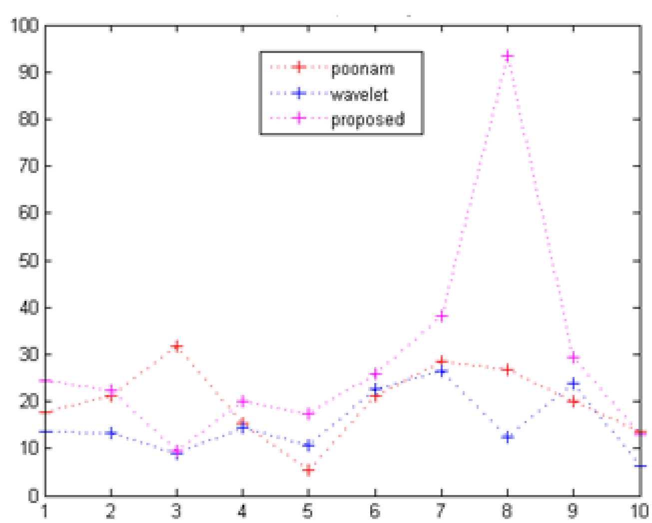

Figure-2: Plot of Detection percentage for Poonam et al.[14] , Thakre et al.[15] methods and proposed method.

Accuracy: Accuracy can also be stated as percentage of correct classification. Good segmentation will have higher values of accuracy. According to Khare et al.[19], accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

The maximum value of accuracy achieved by the proposed method is 0.52234 where as it is 0.4762 for Poonam et al. [14] and it is 0.4492 for Thakre et al.[15] methods. A final conclusion drawn from the observations furnished in Table-3 and the plots shown in Figure-3 is that the proposed method's accuracy is superior to the methods[14, 15].

Missing Factor (mf): According to Ibrahim et al.[20] missing factor (mf) is defined as

$$mf = \frac{FN}{TP} \quad (8)$$

Table-2: Quantitative values of boundary recall, Detection % for Poonam et al.[14], Thakre et al[15] methods and the proposed method.

	Boundary Recall			Detection percentage		
Sr. No.	Poonam et al.[19]	Thakre et. al[26]	Proposed	Poonam et al.[20]	Thakre et. al[26]	Proposed
Video 1	0.1777	0.137	0.24365	17.7655	13.7056	24.3655
Video 2	0.2121	0.1333	0.22424	21.2121	13.3333	22.4242
Video 3	0.3185	0.0888	0.0963	31.2519	8.8889	9.62963
Video 4	0.1538	0.1435	0.2	15.3846	14.359	20
Video 5	0.1538	0.1077	0.17365	5.3892	10.7784	17.3653
Video 6	0.2121	0.2272	0.25758	21.2121	22.7273	25.7576
Video 7	0.2844	0.2656	0.38125	28.4375	26.5625	38.125
Video 8	0.2667	0.1238	0.93333	26.6667	12.381	93.3333
Video 9	0.1989	0.2375	0.29282	19.8895	23.7569	29.2818
Video 10	0.1357	0.0633	0.13122	13.5747	6.3348	13.1222

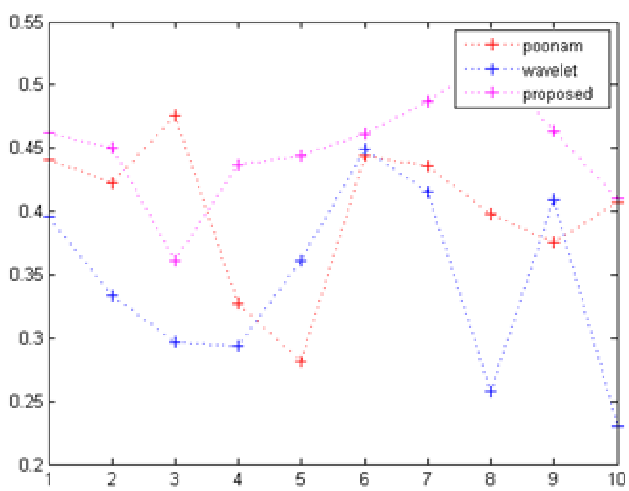

Figure-3: Plot of accuracy for Poonam et al. [14] , Thakre et al.[15] methods and proposed method.

Table-3: Accuracy, Missing factor obtained for Poonam et al.[14], Thakre et al.[15] methods and the proposed method.

	Accuracy			Missing Factor		
Sr. No.	Poonam et al.[14]	Thakre et. al[15]	Proposed	Poonam et al.[14]	Thakre et. al[15]	Proposed
Video 1	0.4411	0.3956	0.462	4.6286	6.2963	3.10417
Video 2	0.423	0.3333	0.45016	3.7143	6.5	3.45946
Video 3	0.4762	0.2967	0.36132	2.1395	10.25	9.38462
Video 4	0.3274	0.2937	0.43689	5.5	5.9643	4
Video 5	0.2815	0.3615	0.44433	17.5556	8.2778	4.75862
Video 6	0.4441	0.4492	0.4619	3.7143	3.4	2.88235
Video 7	0.4355	0.4151	0.48694	2.5165	2.7647	1.62295
Video 8	0.3983	0.2576	0.52234	2.75	7.0769	0.07143
Video 9	0.3753	0.4091	0.46356	4.0278	3.2093	2.41509
Video 10	0.4069	0.2296	0.41034	6.3667	14.7857	6.62069

The lower the value of mf the better the performance of the method. The maximum value of missing factor of the proposed method is 9.38462, Poonam et al.[19] method is 17.5556 and Thakre et al.[15] method is 14.7857. The discussion and the plot of values of missing factor shown in Figure-4 concludes that the values of missing factor of the proposed method are lower compared to the methods[14,15].

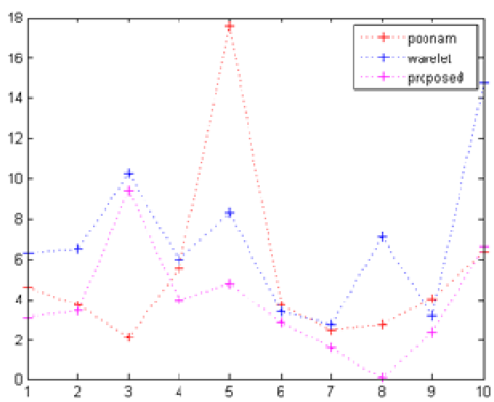

Figure-4: Plot of Missing factor for Poonam et al.[14], Thakre et al.[15] methods and proposed method.

Conclusion

A novel shot detection method based on second order statistics of gray level co-occurrence matrix has been proposed in this article. In the proposed method shots were detected by comparing the differences of contrast, correlation, energy, homogeneity between two consecutive frames with a threshold. The whole dataset containing 10 videos belonging to different categories of animations were collected from the internet and the proposed method was tested on. Performance of the proposed method and the methods of literature were evaluated with the evaluation parameters boundary recall, detection percentage, accuracy, missing factor. The boundary recall of the proposed method shown superiority of 0.6, detection percentage is enhanced by 60%, an improvement of 0.05 is shown in accuracy and the least missing factor of 0.07143 is achieved by the proposed method compared to state-of-art methods. The experimental results demonstrated that the proposed algorithm yielded good results.

References

1. Mohanta Partha Pratim, Saha Kumar Sanjoy and Chanda Bhabatosh (2012). A Model-Based Shot Boundary Detection Using Frame Transition Parameters. *IEEE Transactions on Multimedia.*, 14(1), 223-233.
2. Birinci Murat and Kiranyaz Serkan (2014). A perceptual scheme for fully automatic video shot boundary detection. *Signal Processing: Image Communication.*, 29(3), 410-423.
3. Tavassolipour Mostafa, Karimian Mahmood and Kasaei Shohreh (2014). Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(2), 291-304.

4. Lu ZheMing and Shi Yong (2013). Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. *IEEE Transactions on Image processing,* 22(12), 5136-5145.

5. Lakshmi Priya. G.G. and Domnic S. (2014). Shot based keyframe extraction for ecological video indexing and retrieval. *Ecological Informatics*, 23, 107-117.

6. Mendi E. and Bayrak C. (2013). Shot boundary detection and keyframe extraction from neurosurgical video sequences. *The imaging Science Journal,* 60(2), 90-96.

7. Warhade Krishna K., Merchant S.N. and Desai U.B. (2011). Shot boundary detection in the presence of fire flicker and explosion using stationary wavelet transform. *Signal, Image and Video Processing,* 5(4), 507-515.

8. Vila Marius, Bardera Anton, Xu Qing, Feixas Miquel and Sbert Mateu (2013). Tsallis entropy based information measures for shot boundary detection and keyframe selection. *Signal, Image and Video processing*, 7(3), 507-520.

9. Thounaojam Dalton Meitei, Khelchandra Thongam, Singh Kh. Manglem and Roy Sudipta (2016). A Genetic Algorithm and Fuzzy Logic Approach for Video Shot Boundary Detection. *Computational Intelligence and Neuroscience*, 1-12. Article ID 8469428.

10. Wang Shuai, Cong Yang, Cao Jun, Yang Yunsheng, Tang Yandong, Zhao Huaici and Yu Haibin (2016). Scalable gastroscopic video summarization via similar-inhibition dictionary selection. *Artificial Intelligence in Medicine,* 66, 1-13.

11. Toharia Pablo, Robles Oscar D., Suarez Ricardo, Bosque Jose Luis and Pastor Luis (2012). Shot boundary detection using Zernike moments in multi-GPU multi-CPU Architectures. *Journal of Parallel Distribution Computinng,* 72(9), 1127-1133.

12. Warhade Krishna K., Merchant S.N. and Desai U.B. (2013). Shot boundary detection in the presence of illumination and motion. *Signal Image and Video processing,* 7(3), 581-592.

13. Duan Feng-feng (2016). Shot Segmentation for Binocular Stereoscopic Video Based on Spatial–Temporal Feature Clustering. *3D Research*, 7, 29.

14. Jadhava Poonam S. and Jadhav Dipti S. (2015). Video Summarization using Higher Order Color Moments. *Procedia Computer Science*, 45, 275-281.

15. Thakre K.S., Rajurkar A.M. and Manthalkar R.R. (2016). Video Partitioning and Secured Keyframe Extraction of MPEG Video. *Procedia Computer Science*, 78, 790-798.

16. VirtualDub (2017). Proof that I had too much free time in college http://www.virtualdub.org/index.html

17. Dutta Debabrata, Kumar Saha Sanjoy and Chanda Bhabatosh (2016). A shot detection technique using linear regression of shot transition patterns. *Multimedia Tools and Applications*, 75(1), 93-113.

18. Poornima K. and Kanchana R. (2012). A Method to Align Images Using Image Segmentation. *International Journal of Soft Computing and Engineering,* 2(1), 294-298.

19. Khare Manish, Srivastasava Rajneesh Kumar and Khare Ashish (2015). Moving Object Segmentation in Daubechies Complex Wavelet domain. *Journal of Signal, Image and Video Processing*, 9(3), 635-650.

20. Shaker Ibrahim F., Abd-Elrahman Amr, Abdel-Gawad Ahmed K. and Sherief Mohamed A. (2011). Building Extraction from High Resolution Space Images in High Density Residential Areas in the great Cairo Region. *Remote Sensing,* 3(4), 781-791.