Improving distributed video coding side information by intelligently combining macro-blocks from multiple algorithms

Conference or Workshop Item

Improving distributed video coding side information by intelligently combining macro-blocks from multiple algorithms

M.O. Akinola, L.S. Dooley, K.C.P. Wong

Department of Computing and Communications, neXt Generation Multimedia Technology (XGMT) Research Group, The Open University, Milton Keynes, MK7 6AA, United Kingdom

Keywords: Wyner-ziv, side information, distributed video coding, adaptive overlapped block motion compensation, rough set theory.

Abstract

The performance of distributed video coding (DVC) greatly relies on the quality of Side information (SI). This paper investigates a novel way of producing SI by intelligently combining macroblocks (MB) produced by two SI generation algorithms, namely higher-order piecewise temporal trajectory interpolation (HOPTTI) and adaptive overlapped block motion compensation (AOBMC). The two algorithms address the problem differently. HOPTTI attempts to improve the motion estimation using higher order trajectory interpolation while AOMBC addresses the blocking and overlapping problem caused by inaccurate block matching. By judiciously selecting when to incorporate AOBMC with HOPTTI, it would give a peak signal-to-noise ratio (PSNR) improvement in SI quality. Two switching mechanisms, which exploit the spatial-temporal correlation at the macro-block level, have been investigated and the RST-based intelligent mode switching (IMS) algorithm is found to produce enhanced SI quality. Experimental results show that the basic mode switching algorithm gives a PSNR improvement of up to 1.8dB in SI quality compared to using only HOPTTI. The more intelligent RST-based switching provides a further PSNR enhancement of up to 1.1dB for certain test sequences.

1 Introduction

DVC is a practical implementation of the Slepian-Wolf (SW) [1] and Wyner-Ziv (WZ) [2] theorems that reverses the traditional complex encoder - basic decoder paradigm to one where the complexity is incurred at the decoder. This enables relatively simple encoder designs which can be readily deployed in resource-constrained portable devices [3]. In DVC, selected frames (known as key frames) of a sequence are encoded using conventional encoding method such as H.264 at encoder and send to decoder through a communication channel. The frame(s) situated between two key frames, known as Wyner-Ziv (WZ) frames are not directly transmitted to the decoder. Instead, these frame(s) are to be reproduced based on the received key frames. SIs are the coarse representation of these WZ frames, which are not available at the decoder. Hence the quality of SI has a major impact on the resulting DVC output quality [4–7]. SI is commonly generated using linear-motion compensated temporal interpolation (LMCTI) [3-5] and while this generally provides reasonable quality, it does not always afford an adequate formulation as motion in real sequences is not always linear. More accurate SI have been generated in DVC by the modelling of higher-order trajectories [8-11] which includes the HOPTTI algorithm [10]. As the higher order trajectory is determined by using three or more motion vectors (MV) from previous and future frames, HOPTTI is able to model non-linear motion more accurately and deliver improvement in SI quality [10].

However, due to its use of block matching algorithms (BMA), blocking artefacts and overlapping can be observed especially where abrupt changes occur in motion trajectory and multiple objects occupy the same MB. The twin problem of blocking artefacts and overlapping were addressed by incorporating adaptive overlapped block motion compensation (AOBMC) algorithm [11] with HOPTTI. AOBMC allows the use of a raised cosine window in order that MV of neighbouring pixels to the MB under investigation can more accurately modify the predicted MV. AOBMC have been used to tackle the effects of BMA in LMCTI based interpolation and they include motion compensated frame interpolation and adaptive object motion compensation [12] where AOBMC is used along with a MV clustering technique. However, experiments shows that HOPTTI combined with AOBMC gives superior performance [11]. In some complex sequences where the neighbouring pixels to the MB under consideration are not correlated so the modification of the MV using them produces erroneous results, or where the predicted MV is already accurate, the consideration given to neighbouring pixels introduces new errors giving rise to situations where original HOPTTI outperforms HOPTTI-AOBMC output. It is evident that improved SI quality can be achieved if their outputs are combined intelligently on a macro-block basis. Using the spatial-temporal video content characteristics of sum of boundary absolute difference (SBAD) and sum of mean absolute difference (SMAD), a mode switching (MS) mechanism, which applies a matching criterion (MC) introduced by [11] to empirically switch between original HOPTTI and HOPTTI-AOBMC, achieved a further improvement in SI performance. However, it was noted from [11] that the accuracy of MS, which has direct impact on the quality of SI, can be improved further if the switching is conducted more intelligently. This leads to an investigation of incorporating the rough set theory (RST) for governing the video content based IMS. RST has been employed for data mining and analysis and characterization for reasoning about data [13] that provides a formal robust method for manipulating the various features and attributes in data sets and has been successfully employed to increase performance in de-interlacing [13], Experimental results shows that the RST based classifier IMS can produce significant improvement in SI quality.

The remainder of this paper is organized as follows: Section 2 reviews the theories of HOPTTI formulation, the AOBMC algorithm, and empirical mode switching concept and introduces intelligence via the rough set concept. Section 3 presents quantitative and qualitative results and analysis of this SI generation scheme. Section 4 provides some conclusions.

2 Theoretical foundations

2.1. HOPTTI Formulation

In contrast to conventional linear trajectory estimation, HOPTTI estimates the motion trajectory of an object by a set of piecewise cubic (3rd order) polynomials which allow the modeling of the trajectory with variable accelerations. By adding a jolt term (rate of change of acceleration) in the motion trajectory equation, it provides a more accurate motion estimation of objects in the real world. The motion trajectory of an object can be represented by a set of piecewise cubic polynomials:

$$C(t) = \begin{cases} p_1(t) & for \quad t_1 \le t \le t_4 \\ p_2(t) & for \quad t_4 \le t \le t_8 \\ \vdots & \\ p_n(t) & for \quad t_n \le t \le t_{n+3} \end{cases} \tag{1}$$

where each segment of the trajectory $p_i(t)$ is represented by an equation of motion similar to [8] and considering a constant jolt given by:

$$p_i(t) = \frac{1}{6} j_i (t-t_i)^3 + \frac{1}{2} a_i (t-t_i)^2 + v_i (t-t_i) + d_i \tag{2}$$

For $i = 1, 2 \ldots\ldots n$. In (1), n is the number of available key frames, while in (2), j_i is the average *jolt* (the rate of change of acceleration), a_i the average acceleration, v_i the average velocity between t_i and t_{i+1} and d_i the initial displacement at t_i.

The trajectory is built using the MVs of MBs in the previous and future key frames and using motion compensated interpolation to predict intermediate object positions and the SI frames [10]. To calculate the four parameters j_i, a_i, v_i and d_i, a minimum of 4 key frames are required, and if it is assumed the respective displacements of the blocks at these key frames are A_i, B_i, C_i and D_i, then the following holds:

$$d_i = A_i \tag{3}$$

$$v_i = \frac{B_i - A_i}{T} \tag{4}$$

$$a_i = \frac{v_{i+1} - v_i}{2T} = \frac{C_i - 2B_i + A_i}{2T^2} \tag{5}$$

$$j_i = \frac{a_{i+1} - a_i}{3T} = \frac{D_i - 3C_i + 3B_i + A_i}{6T^3} \tag{6}$$

where T is the time between two consecutive key- frames, $A_{i+1} = B_i, B_{i+1} = C_i, C_{i+1} = D_i$.

The forward motion trajectory of an object can be evaluated using (1) – (6), thus enabling the MV of the object at any time between t_1 and t_{n+1} to be accurately interpolated. The backward motion trajectory is evaluated the same way as the forward one using (1) – (6) as described but in reverse direction i.e. D_i, C_i, B_i *and* A_i.

2.2. AOBMC Algorithm

Though HOPTTI improved SI generation [10] due to more accurate motion modeling, the MV estimation using BMA algorithm can sometimes have issues of MB overlapping caused by inaccuracies in forward and backward trajectories or blocky artifacts caused by multiple or deformable objects in a single MB. AOBMC has been employed with HOPTTI [11] (known as HOPTTI-AOBMC) to tackle the above mentioned issues by allowing pixels of the surrounding blocks to moderate the predicted MV, as illustrated in Figure 1, then, using a raised cosine weighting window enclosing the neighboring MBs and the size of the window is determined by the reliability of the MV of neighboring MBs and their distance from the MB under consideration. This is achieved by minimizing SBAD [10], which measures the spatial error of the MB under consideration between the reference and current HOPTTI frames. Since the original current frame is not available at the decoder, the previous key frame is used as the reference frame [7].

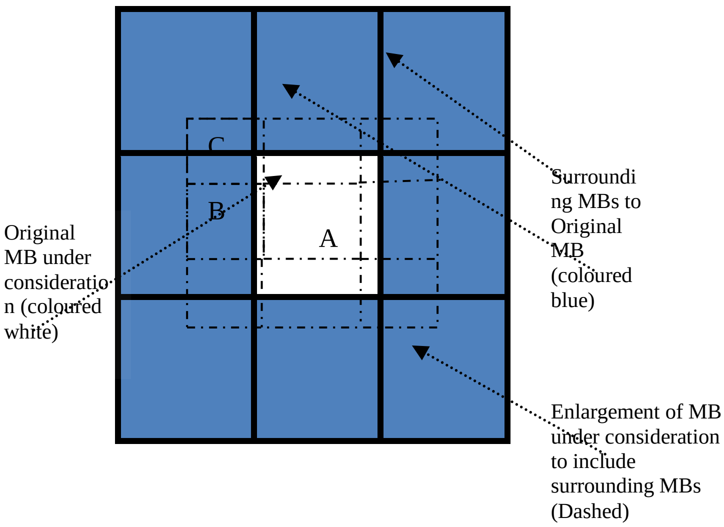

Figure 1: Illustration of sample overlapped blocks for AOBMC.

2.3. Empirical Mode Switching

To experiment with the idea of combining MBs produced by HOPTTI and HOPTTI-AOBMC, a simple MS mechanism was introduced. It uses the spatial-temporal properties of the video sequence to govern switching between MBs of the frames that will benefit from the application of AOBMC and those that should retain HOPTTI. Experimental results show that improved SI are achievable.

The spatial-temporal properties of the video sequences are measured by SBAD, which defined earlier and SMAD [10], which measures the temporal error of the MB under consideration between the previous and current HOPTTI frames. SBAD and SMAD are used to form the matching criterion (MC) for the empirical *mode switching*. The MC is thus defined as:

$$MC = \lambda * SBAD + (1-\lambda) * SMAD \tag{7}$$

where λ is a predefined weighting factor. The MS mechanism applies a threshold T as follows:

Calculate *MC using* (7)
If $MC \geq T$ THEN apply HOPTTI with AOBMC
ELSE use HOPTII alone
END

It can be noted that fixing the weights λ and threshold T by empirical experimentation limits the actual improvement that is achievable as this does not change with the variation in video content properties and underlies the fact that the mode switching if conducted intelligently, for example using RST that learns from spatial-temporal video content data, then SI can be further improved.

2.4. RST and Intelligent Mode Switching Algorithm

Intelligently switching between the HOPTTI and HOPTTI-AOBMC can not only improve performance but also eliminating the need for empirically determining the thresholds when new sequence is to be processed. This section hence discuss the implementation of an intelligent switching method, IMS. RST is a mathematical framework for inducing rules through supervised training. The resulting rules can subsequently be used to classify objects or patterns. It has been applied in video sequence property based switching used previously in [13] for de-interlacing. RST was therefore be chosen as *intelligent mode switching* (IMS) mechanism for DVC SI generation.

RST is used to intelligently decide which MBs in the frame that will benefit from the incorporation of AOBMC and which do not. Based on the temporal and spatial characteristics of the video sequence measured by SMAD and SBAD, the MBs of the intermediate SI produced by HOPTTI and HOPPTI-AOBMC along with their characterization and classification decision are represented by a two dimension decision table. In RST term, these MBs are known as objects and their characterizations are known as attributes.

The principle of RST can be illustrated in Figure 2 [13], where MBS is a set of the actual imprecise (unknown) boundary and $MBS \subseteq X$ - the universe of discourse comprises of MBs of various video sequences, there two classes namely;

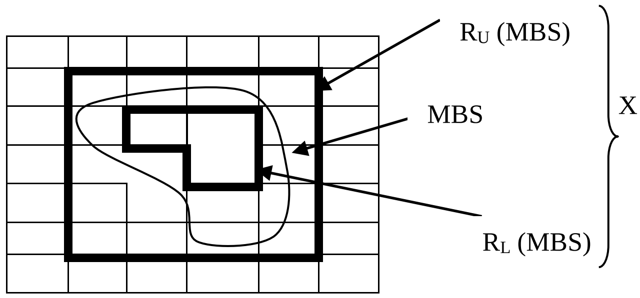

Figure 2: Schematic Illustration of the rough set theory [13].

(a) class of MBs that are known to be benefit from the incorporation of AOBMC;

(b) class of MBs that will not.

$R_L(MBS)$ is the lower approximation corresponding to MBs that we are sure they belong to class (a) and therefore used as the training set, $R_U(MBS)$ is the upper approximation corresponding to all MBs that possibly belong to class (a) but require further testing. The rough set theory states that there is an indiscernible relation in the universe of discourse which may be induced by a given set of attributes ascribed to the objects of the domain which corresponds to each MB being examined. During training, the objects in the decision table will be compared. The ones with a unique set of attributes will be retained and those are not will be eliminated.The remaining objects in the decision table can be seen as a set of rules that will be used to classify unseen objects

2.4.1. Rough Set Object, attributes and Decision Tables

Most specifically, a decision table is used for describing the objects of the universe of discourse and it consists of a two dimensional table where each row is an object and each column is an attribute, except the last column, which contains the decision. For this work, each row represents a different MB of a video sequence while each column contains a different attribute describing the MB. While some attributes are raw data, others can be obtained through some rules or based on prior knowledge from empirical experimentation. For example, the outputs of SBAD and SMAD can be further divided into three categories, Small (Sm), Medium (Me) and Large (La) to give extra attributes to describe the objects as shown in Table 1 and 2. These extra attributes can make an object (MB) more distinguishable and reduce the classification difficulty. The RST for the HOPTTI and HOPTTI-AOBMC switching is formulated into a system of attributes and objects in a decision table as shown in Table 2, where Id the identification number of the MB, Raw pixel is the average pixel intensive of the MB, Raw SMAD and Raw SBAD are the respective measures of the temporal and spatial errors, and Cond. SMAD and Cond. SBAD are extra attributes generated by the attribute filtering rules, and the DECISION column indicates whether the MB should be chosen from HOPTTI or

HOPTTI-AOBMC. These resemble a human expert completing a puzzle by placing the appropriate pieces together, but in the context of DVC, it is the reconstruction of the SI by intelligently selecting MBs from HOPTTI and HOPTTI-AOBMC frames based on the categorization of SBAD and SMAD.

Label		***Rules***		***Output***
Cond. SMAD	IF	SMAD ≤ 8	THEN	Sm
Cond. SMAD	IF	SMAD $\geq 8 \leq 16$	THEN	Me
Cond. SMAD	IF	SMAD ≥ 16	THEN	La
Cond. SBAD	IF	SBAD ≤ 8	THEN	Sm
Cond. SBAD	IF	SBAD $\geq 8 \leq 16$	THEN	Me
Cond. SBAD	IF	SBAD ≥ 16	THEN	La

Table 1: Filtration rules for attributes discretization.

AMERICAN FOOTBALL						
Id	***Raw pixel***	***Raw SMAD***	***Raw SBAD***	***Cond. SMAD***	***Cond. SBAD***	***DECISION***
7	110.3	19.77	38.02	La	La	HOPTTI
99	107.7	10.36	21.80	Me	La	HOPTTI-AOBMC

Table 2: Sample attributes of American Football sequence.

3 Results

In this study, the universe of discourse composes of the set of objects that are the MBs of HOPTTI and HOPTTI-AOBMC frames generated from the Coastguard, Hall and the American football sequences. These sequences are chosen because they contains a wide variety of global and local motions of different speeds. They are very commonly used for evaluating video coding performances [refs]. The RST based IMS algorithm is developed using the publicly available WEKA® command line RST classification software developed by the University of Waikato [14]. To evaluate the classification performance of IMS a ground truth is produced. The ground truth is obtained by artificially making the original WZ frame available such that PSNR of the MBs of the intermediate SI frames produced by both HOPTTI–AOBMC and HOPTTI algorithms can be calculated and compared. For each MB, the algorithm that produces the higher PSNR is chosen as the desired algorithm for that MB. Therefore, the ground truth is a table containing the desired algorithm for each MB. The training of IMS is conducted using the ground truth data from the coastguard sequence as it exhibits both global and multiple object motions. After the training, IMS is used to predict the desired algorithm of the MBs of the all three test sequences. An object is classified based on the similarity of its attributes to that of the trained objects.

Table 3 shows the PSNR of the sequences generated by HOPTTI, the empirical mode switching algorithm and the RST based IMS. Table 4 shows the switching (classification) accuracy and the PSNR comparisons between mode switching, IMS and the ground truth. As can be seen IMS consistently outperforms the mode switching algorithm both in terms of percentages of correctly mode switched and PSNR, with improvement of up to 3dB over HOPTTI.

Sequences	***HOPTTI [10]***	***MODE Switching***	***IMS***
Coastguard	36.4	38.5	39.45
Hall	38.5	40.7	41.42
American Football	24.5	26.6	27.04

Table 3: Average PSNR (dB) for mode switching, IMS and HOPTTI for selected test sequences.

Sequences		***Mode Switching***	***IMS***	***Ground Truth***
Coastguard	% correct switch	78.1%	94%	100%
	PSNR dB	38.5	39.45	40.2
Hall	% correct switch	75.6%	87%	100%
	PSNR dB	40.7	41.42	41.8
American Football	% correct switch	88.2%	95.5%	100%
	PSNR dB	26.6	27.04	27.7

Table 4: Block based analysis for mode switching (empirical) versus IMS for the selected test sequences.

Sample frames of the American Football sequences showing perceptual improvements are illustrated in Figure 3, where both qualitative and quantitative performance are improved as we go from all the MBs being HOPTTI to AOBMC and from basic mode switching to IMS with the ghosting disappearing and PSNR increasing.

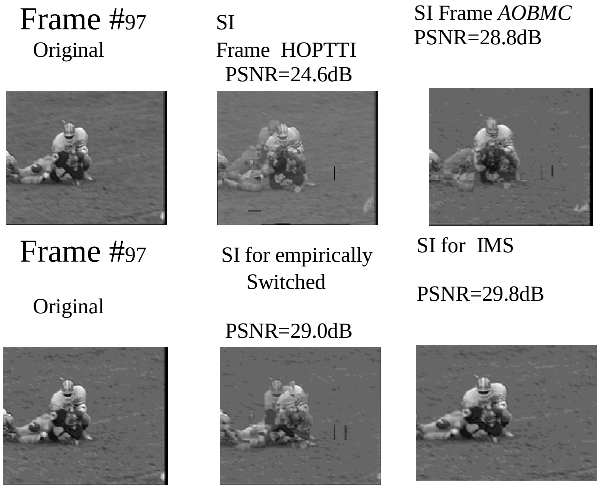

Figure 3: Sample frames for *American Football* showing the SI quality obtained using HOPTTI [10], AOBMC [10], empirically *Switched HOPTTI-AOBMC and IMS*.

In term of rate distortion (RD) performance results show that not all the improvement in SI from the SI generation module is carried to the final codec output. While SI improvement is up to 3dB the final codec output is only up to 2dB improvement over HOPTTI. The overall RD results for Hall sequence chosen to accommodate both multiple object sequence and objects in motion in a sequence is shown in Figure 4. The result is that RST based IMS outperforms HOPTTI, H.264 No Motion and H.264 intra while the H.264 inter remains the upper limit that outperforms Switched RST. The RD results for the Hall sequence show that at low bit rates RST based IMS outperforms H.264 inter. This is mainly due to the fact that the residue in H.264 which accounts for major performance only kicks in at medium to high bit rates and DVC is therefore more competitive at low bit rates.

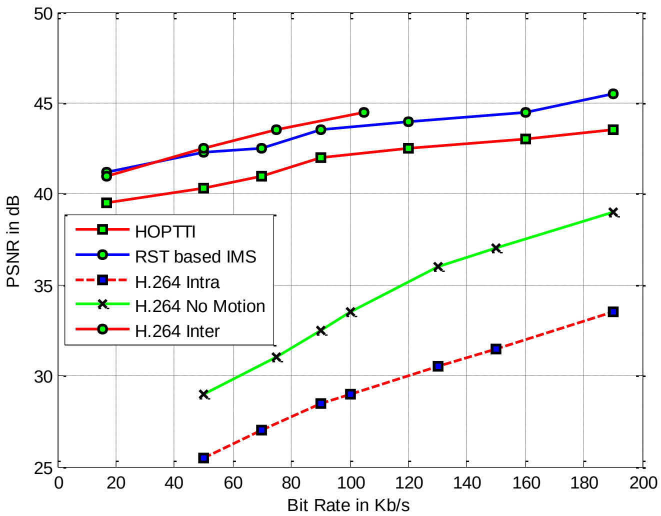

Figure 4: RD Curves showing RST based IMS PSNR performance for *Hall* sequence @ 15f/s

4 Conclusion

This paper presents a novel RST based intelligent mode switching algorithm for DVC SI generation. It can intelligently combine the best MBs from the SI generated by two algorithms addressing two SI generating issues. The IMS employed further improves the PSNR produced the empirical mode switching and HOPTTI by up to an additional 0.95dB and 3.1dB respectively and demonstrated the ability to remove overlapping and blocking artefact in SI. The switching performance analysis further shows that intelligent switching approach improves the classification accuracy by up to 16 percentage point over empirically switched MB based switching. Furthermore, qualitative (visual) results show that SI produced by IMS is significantly sharper than that produced by HOPPTI, HOPTTI-AOBMC and the basic mode switching algorithm. It can be concluded that improved quality of SI can be achieved by intelligently combining MBs from SI generated by algorithms that were aided to addressing different SI generating issues.

References

[1] D. Slepian, and J. Wolf., "Noiseless coding of correlated information sources," in IEEE Transactions on Information Theory, vol. 19, pp.471-480, 1973.
[2] A. D. Wyner, and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," in IEEE Transaction on Information Theory, vol. 22, pp.1-10, 1973.
[3] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: selecting the most promising application scenarios," in Elsevier, Signal Processing: Image Communication, vol. 23, pp. 339-352, 2008.
[4] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in Proc. of Picture Coding Symposium, vol 6, pp 14496-14410, Lisbon, Portugal, 2007.
[5] X. Li, X., "On the importance of source classification in Wyner-Ziv video coding," in SPIE conf. on Visual Communications and Image Processing, Jan. Vol. 5308, pp. 520-528, 2008.
[6] R. Liu, Z. Yue, and C. Chen, "Side information generation based on hierarchical motion estimation in distributed video coding," in Chinese Journal of Aeronautics, Volume 22, Issue 2, pp. 167-173, 2009.
[7] S. Ye., M. Ouaret., F. Dufaux, and T. Ebrahimi, "Improved side information generation for distributed video coding by exploiting spatial and temporal correlations," in EURASIP journal on Image and Video Processing, Vol. 2009, Article ID 683510, 15 pages., 2009.
[8] M. Chahine, and J. Konrad, "Estimation and compensation of accelerated motion for temporal sequence interpolation," in Signal Processing. Image Communication, vol. 7, pp. 503--527, Nov. 1995.
[9] M. Chahine, "Motion compensation interpolation using trajectories with acceleration," in IS&T/SPIE Symposium on

Electronic Imaging Digital Video Compression, San Jose, CA, USA vol.2419, Feb. 1995.

[10] M. O. Akinola, L. S. Dooley, and K. C. P. Wong, “Wyner-Ziv Side Information Generation Using a Higher Order Piecewise Trajectory Temporal Interpolation Algorithm,” International Conf. on Graphic and Image Processing (ICGIP’10), Manila, Philippines, 2010, pp. 116-121.

[11] M. O. Akinola, L. S. Dooley, and K. C. P. Wong, “Improved Side Information Generation Using Adaptive Overlapped block Motion Compensation and Higher-Order Interpolation,” 18th International Conference on Systems, Signals and Image Processing (IWSSIP 2011), Sarajevo, Bosnia & Herzegovina, 2011, pp. 97-100.

[12] B-D. Choi, J-W. Han, C-S. Kim, and S-J. Ko, “Motion Compensated Frame Interpolation and Adaptive Overlapped Block Motion Compensation” IEEE Transaction on Circuits and Systems for Video Technology, Vol. 17 (4), April 2007, pp. 407-416.

[13] G. Jeon, J. Won, R. Lee and J Jeong, “A rough set approach for video deinterlacing” IEEE International conference on Multimedia and Expo (ICME) Beijing, China, July 2007, pp. 1942-1945.

[14] H. Ian, E. Frank and M. A. Hal, “Data Mining: Practical Machine Learning Tools and Techniques” Morgan Kaufmann publishers, Burligton, MA, USA, 2011.