Classification of Video Resolution for the Enhanced Display of Images on HDTV

Jewoong Ryu*, Gibak Kim†, Sang Hwa Lee*, Byungseok Min‡ and Nam Ik Cho*
*INMC, School of Electrical and Computer Engineering, Seoul Nat'l University, Seoul, Republic of Korea
E-mail: youjw@ispl.snu.ac.kr Tel: +82-2-8808480
†School of Electrical Engineering, Soongsil University, Seoul, Republic of Korea
E-mail: imkgb27@ssu.ac.kr Tel: +82-2-8287266
‡Samsung Electronics Co. Ltd., Suwon, Republic of Korea
E-mail: byungseok.min@samsung.com Tel: +82-31-2790875

Abstract—**Although video sources for HDTV broadcasting are mainly in HD resolution, some of them are still from low resolution video sources such as videos that are taken long time ago, internet-streamed or cellphone videos. When these sources are displayed on HDTV, they usually appear to be blurry and their color is not vivid. To alleviate these problems, a proper video enhancement for each source is necessary to display them on HDTV with satisfaction. However, since the decoded images on HDTV do not contain the information on the origin of sources in many cases, it needs to be classified whether their origins out from HD source or not. For this, we propose an HD/non-HD classifier based on the Support Vector Machine (SVM) with the frequency information of the selected region in the decoded image. To evaluate the performance of the proposed HD/non-HD classifier, we use a test database of 6252 HD and 6934 SD still images captured from various TV genres. The experimental results show that the proposed classifier yields high accuracy rate of 99.61%.**

I. Introduction

High-definition television system (HDTV) had started to broadcast since the late 1990s and became the mainstream in these days. Even though the HDTV is the most common TV set and a lot of video sequences are produced in HD resolution nowadays, some video sequences such as user created contents (UCCs), internet-streamed videos and TV streams taken a long time ago still have poor resolutions to be displayed on HDTV. Hence, when these video sequences are displayed on HDTV, they are interpolated to HD resolution to fit the TV screen, which results in blurry effects on images. Moreover, the color of some of theses sequences is not vivid on HDTV because they are produced for displaying on the analog CRT TV or other devices [1].

In order to address these problems, some researches on contrast enhancement of the color [1]–[3] and edge enhancement of blurry video sequences [4] have been performed for the decades. However, these researches have focused on the enhancement for the video sequences regardless of the original source information. Without considerations of the original source, theses methods can introduce over-saturated color image or unwanted high frequency artifacts on HD sequences.

Thus, in this paper, we propose a classification method for finding original resolution of video sequences, which helps the image quality enhancement algorithms have the information of the original source to apply them appropriately. One may think that the information of original source can be attained by the decoder of HDTV. However, due to the variety of digital video source devices such as smart phone, laptop and set-top box of IPTV, the decoded image itself is the only information at the enhancement device. Hence, our algorithm classifies the source resolution by exploiting the signal characteristics of the decoded images from the input video stream. That is, we analyze the differences of frequency features of HD/non-HD video stream, and train a Support Vector Machine (SVM) with the frequency features of them to provide a correct information of original source which can be utilized in further processing.

The rest of paper is organized as follows. We present the classification method of the resolutions of video source(HD/non-HD) in Section II. In Section III, experimental results on the databases constructed from TV sequences for the evaluation of classification accuracy are presented and a simple example of image enhancement by exploiting our algorithm is also provided. Finally, conclusion of this paper is followed in Section IV.

II. Classification of the Resolution of Video Source

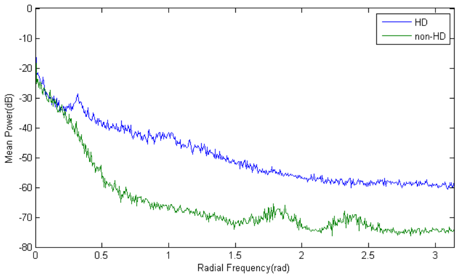
Classification of video resolution on HDTV is stated as a problem of finding different characteristics of HD/non-HD video sequences and designing a proper machine to determine the resolution of video source. We provide a brief explanation on the differences between HD and non-HD video sequences and present a training method using Support Vector Machine (SVM) that classifies the original source resolution of video sequences.

A. Extracting Features of HD/non-HD Video Sequences

Since the HDTV is designed to display a video sequence whose resolution is HD, the broadcaster(or HDTV device) interpolates the images to the size of HD when the source is not HD. However, it inevitably brings the loss of high frequency component because the interpolation filter suppresses it to avoid aliasing effect [5]. In terms of image quality, the loss of high frequency information causes smoothed edge,

(a) HD image

(b) Non-HD image



(c) Frequency power of above images

Fig. 1. Examples of HD/non-HD video sequences and their frequency components.

which is prominent difference between HD and non-HD video sequences when displayed on HDTV as shown in Fig. 1. Hence, in the training stage, we extract features containing frequency information from both video sequences to determine the resolution of each frame image.

A video sequence which is displayed on HDTV is consisted of frame images with the 1920×1080 resolution. In order to classify the original resolution of video sequences, we extract frequency information from each frame and select a certain region of the frame image that represents edge characteristics better than the others. First, we set 5 regions with $M \times N$ size as candidate regions $\mathbf{R}_i$ as shown in Fig. 2, where M is the width of the region and N is the height. In order to obtain the edge map, we exploit Sobel operators

$$\mathbf{G_x} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \mathbf{G_y} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad (1)$$

to calculate the edge image of each candidate region $\mathbf{R}_i$. The region of interest is chosen so as to include high frequency information such as edges by

$$\mathbf{R}_{\mathrm{roi}} = \arg\max_{\mathbf{R}_i} \left(\sum_{(x,y)\in\mathbf{R}_i} \left(|\mathbf{G_x} * I(x,y)| + |\mathbf{G_y} * I(x,y)|\right) \right), \quad (2)$$

where $I(x, y)$ is the pixel value of input image at (x, y) and asterisk($*$) denotes convolution operation.

Then, the frequency information of the selected region is obtained by performing 1-D Fast Fourier Transform (FFT) [6] on the horizontal lines of the image. Because many TV sequences are interlaced, which introduces false edges in vertical direction, we do not use the vertical frequency information. Since we have already determined the region that represents edge information of an image, we calculate the M-point FFT coefficients of each horizontal line in $\mathbf{R}_{\mathrm{roi}}$ as

$$\mathbf{X}_i(k) = \mathrm{FFT}(\mathbf{x}_i), \quad (3)$$

where $\mathbf{x}_i$ is the i-th horizontal line of the image in $\mathbf{R}_{\mathrm{roi}}$, $0 \le k < M$ and $0 \le i < N$. We calculate the power of FFT coefficients in each line and obtain the normalized spectral power in $\mathbf{R}_{\mathrm{roi}}$ as

$$\begin{aligned} \overline{\mathbf{X}}(k) &= \sum_{i=1}^{N} \frac{|\mathbf{X}_i(k)|^2}{\mathbf{Z}_i} \\ \mathbf{Z}_i &= \sqrt{\sum_k |\mathbf{X}_i(k)|^2}. \end{aligned} \quad (4)$$

Note that the dimension of $\overline{\mathbf{X}}(k)$ is M, which is usually a power of 2 and too large for the training. To alleviate the problem of large dimension, we reduce dimension of $\overline{\mathbf{X}}(k)$ to define the feature vector of m-th frame $\mathcal{F}_m$ by employing P non-overlapping windows $\mathbf{w}_j$ which makes $\mathcal{F}_m$ have P dimension as

$$f_{j,m} = -\mathbf{w}_j \cdot \left(\log_{10} \overline{\mathbf{X}}(k)\right), \quad (5)$$

where $j = 1, 2, \ldots, P$ and $f_{j,m}$ is j-th component of $\mathcal{F}_m$.

B. SVM-based Training

In order to solve the classification problem of HD/non-HD sequences, we employed the Support Vector Machines

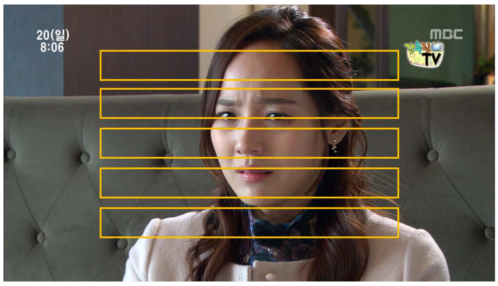


Fig. 2. Candidate regions $\mathbf{R}_i$ of an images .

(SVM) [7]. The SVM classifier produces a hyperplane which guarantees the maximum margin between two classes of data. As in Fig. 1 (c), the frequency characteristics of HD/non-HD sequences are separable due to attenuation of high-frequency components which is attributed to interpolation filter. Thus, we exploit the SVM as a classifier of original video resolution by training HD/non-HD databases with the features derived from previous section.

After the SVM training process with the training set which is constructed from the images of HD/non-HD sequences, we attain a two-class classifier which includes kernel function $K(\cdot,\cdot)$ as

$$\begin{aligned} g(\mathcal{F}_m) &= \sum_{i=1}^{N_{\text{sv}}} \alpha_i y_i K\left(\mathcal{F}_m, \mathbf{s}_i\right) + b, \\ & \sum_{i=1}^{N_{\text{sv}}} y_i \alpha_i = 0, \end{aligned} \tag{6}$$

where $\mathcal{F}_m$ is the feature vector of m-th frame, N_{sv} is the number of support vectors and $\{\mathbf{s}_i, y_i\}$ are support vectors and its ground truth label (1 for HD, -1 for non-HD) from the training database calculated during the optimization process with [8]. For the evaluation of the machine, a feature vector of m-th frame of the test sequence $\mathcal{F}_m$ is derived as in Section II-A. Then, using the classifier $g(\mathcal{F}_m)$, we can determine the classification result based on the sign of $g(\mathcal{F}_m)$. That is, if $g(\mathcal{F}_m)$ is positive, m-th frame is classified as an HD frame, otherwise, it is classified as a non-HD frame.

C. Post-processing

The classification process of HD/non-HD video source is a frame-wise decision: the SVM classifier provides a decision on every frame of video sequence. However, some frames in video sequences that have scenes with large motions can be misclassified due to the blurry artifacts. In this case, even the video is recorded in HD resolution, some of its frames are classified as non-HD which can cause unwanted flickering effects. Hence, we refine the result of SVM classifier by considering temporal aspect as well. In general, the resolution of the video source is not changed rapidly, i.e., the resolution change is usually a shot-wise. Based on these observations, a moving average of classifier in L previous frames from the current frame is used, defined as

$$\overline{g}(\mathcal{F}_m) = \frac{1}{L} \sum_{k=m-L+1}^{m} g(\mathcal{F}_k). \tag{7}$$

In addition to the refining classifier in a temporal manner, we also introduce unknown label. We make the m-th frame leave undetermined if the $|\overline{g}(\mathcal{F}_m)|$ is smaller than a threshold and classification result of that frame be unchanged from previous frame. If the threshold is set to high value, it makes the resolution label of each frame not vary frequently and prevents the flickering artifacts. On the other hand, high threshold may results false detections on some images such as blurry HD image and non-HD images with strong edges. Thus, threshold is empirically set to 2.0 in our method to guarantee high detection performance while suppressing flickering effectively.

TABLE I

EXPERIMENTAL RESULTS ON HD AND NON-HD DATABASE. †AVERAGE ACCURACY WITHOUT STILL IMAGES.

	Frames	Classification Accuracy	
		w/o Post-processing	with Post-processing
Non-HD still images	183	97.81%	N/A
HD still images	160	98.13%	N/A
Non-HD Show	1627	98.83%	100%
Non-HD Show 2	1046	100%	100%
Non-HD Sports	1934	99.95%	100%
Non-HD Drama	2357	100%	100%
HD Show	1500	99.67%	100%
HD Show 2	1324	99.67%	100%
HD Sports	1000	99.50%	100%
HD News	1500	99.93%	100%
HD Drama	928	98.81%	98.81%
Total	13216	99.61%	†99.92%

III. EXPERIMENTAL RESULTS

A. Classification Accuracy

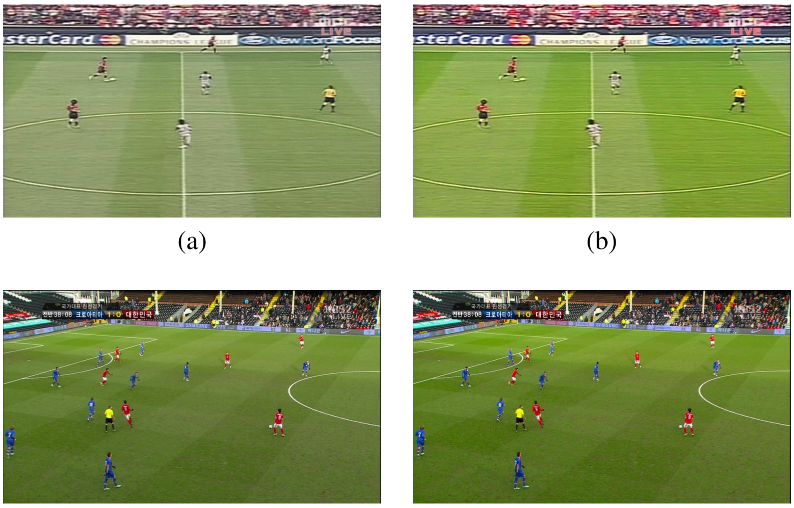
In order to evaluate the classification performance of the proposed algorithm, we conducted experiments with a database consisting of 6252 frames for HD video sequences and 6934 frames non-HD video sequences. Since the main target of the proposed method is TV sequence, the database is constructed using the digitally recorded TV sequences of sports, show, news and drama with HD and non-HD resolutions. We use 567 HD images and 643 SD images for the SVM training and exploit a radial basis function (RBF), i.e., $K(\mathbf{u},\mathbf{v}) = \exp^{-\gamma|\mathbf{u}-\mathbf{v}|^2}$.

Parameters of SVM are selected as $C = 65536$ and $\gamma = 0.2$, where C is the cost parameter of tuning the soft margin and γ is the parameter of RBF kernel function: we tuned these parameters by exploiting cross validation method [9]. The width and height of the region of interest are set to be 1024 and 64, respectively. The window $\mathbf{w}_j$ is defined as

$$\mathbf{w}_j(k) = \frac{M}{2P}\left(u\left(k - \frac{M}{2P}(j-1)\right) - u\left(k - \frac{M}{2P}\,j + 1\right)\right), \tag{8}$$

where $k = 1, 2, \ldots, M$, $j = 1, 2, \ldots, P$ and $u(\cdot)$ is the unit step function. The dimension of feature vector P is set to be 16. We normalize the feature vectors $f(\mathcal{F}_m)$ of training set to $f(\mathcal{F}_m) \in [0, 1]$ for maximizing margin between two classes in training process.

Experimental results are summarized in Table I. Our method shows reasonable classification performance on various kinds of TV sequences. As in Table I, there are some classification errors on the sequences which are originated from the frames that have a lot of motions, which are corrected by post-processing.



(a) (b) (c) (d)

Fig. 3. Examples of video enhancement by the proposed algorithm. (a) Input non-HD image. (b) Enhanced non-HD image. (c) Input HD image. (d) Enhanced HD image.

B. Application to Video Enhancement

One application of our proposed HD/non-HD video source classification algorithm is the enhancement of video sequence. Conventional image enhancement methods for TV such as changing color [1], [2] and unsharp masking [4] do not consider the video source resolution because the metadata of the video sequence is not available when it comes to the image enhancement process. By employing the proposed video source classification algorithm, a proper method for image quality enhancement of each video source can be applied. For example, since non-HD video sequence has blurry edges and low color contrast, one can apply unsharp mask [4] and contrast enhancement with large gain to the non-HD video sequences. On the other hand, as HD video sequence usually has sharp edges and high color contrast, the same strategies for non-HD sequences cannot be applied. Fig.3 shows examples of simple enhancement on video sequence using the proposed classification method. For the images classified as non-HD, we multiply 1.5 to the pixel values of chroma channel to enhance the color contrast and apply unsharp mask to enhance the sharpness of edges. And for the color contrast enhancement of HD images, we multiply chroma channel with 1.2 instead of 1.5 to prevent the images from over-saturation. Since the different enhancement strategies can be applied to both HD and non-HD video sequence by the proposed algorithm, both images are enhanced appropriately without artifacts. Note that the enhancement strategies applied in this paper are the most simple algorithms to visualize the applications of our proposed algorithm. Numerous other enhancement algorithms for HD/non-HD video sequences [2], [4] can be applied with our classification method.

IV. Conclusion

In this paper, we presented a classification algorithm for video resolutions based on the characteristics of frame images of each video sequence. Based on the difference between the high frequency characteristics, we employed the SVM classifier with the frequency-power based feature vector. Experimental results showed that the proposed algorithm has an outstanding performance of classifying the original source resolution of the video sequences.

Acknowledgment

This research was supported by Samsung Electronics Co., Ltd, and also by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013027840).

References

[1] D. G. Han, "A cost effective color gamut mapping architecture for digital tv color reproduction enhancement," *IEEE Trans. Consum. Electron.*, vol. 51, no. 1, pp. 168–174, Feb. 2005.

[2] D. G. Han, "Real-time color gamut mapping method for digital tv display quality enhancement," *IEEE Trans. Consum. Electron.*, vol. 50, no. 2, pp. 691–698, May 2004.

[3] J.H. Souk and J. S. Lee, "Recent picture quality enhancement technology based on human visual perception in lcd tvs," *IEEE/OSA J. Display Technol.*, vol. 3, no. 4, pp. 371–376, Dec. 2007.

[4] G. Scognamiglio, G. Ramponi, A. Rizzi, and L. Albani, "A rational unsharp masking method for tv applications," in *International Conference on Image Processing (ICIP)*, 1999, vol. 4, pp. 247–251 vol.4.

[5] E. Hogenauer, "An economical class of digital filters for decimation and interpolation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 2, pp. 155–162, Apr. 1981.

[6] P. Duhamel and M. Vetterli, "Fast fourier transforms: A tutorial review and a state of the art," *Signal Processing*, vol. 19, no. 4, pp. 259 – 299, Apr. 1990.

[7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995.

[8] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, Apr. 2011.

[9] C. W. Hsu, C.C. Chang, and C. J. Lin, "A practical guide to support vector classification," 2003.