A Comparative Review of Aspect Ratio Conversion methods

Zhongliang Deng[1], Yandong Guo[1], Xiaodong Gu[2], Zhibo Chen[2], Quqing Chen[2], Charles Wang[2]
[1]*Beijing University of Posts and Telecommunications*
[2]*Thomson Corporate Research, Beijing*
guo1984@gmail.com

Abstract

The Aspect ratio is the fractional relation of the width of a video image compared to its height. Because of the historical and some other factors, there are so many video formats with different aspect ratios and the conversion of aspect ratio is required more and more imminently. The paper introduced all the aspect ratio conversion methods published in literatures, including cropping method, expanding method and so on, and presented experiment results to demonstrate the advantages and disadvantages of each conversion method.

1. Introduction

Aspect ratio is the fractional relation of the width of a video image compared to its height. Because of the historical and some other factors, there are so many video formats with different aspect ratios while there are also many kinds of devices used for display videos with different aspect ratios. For example, NTSC has the same aspect ratio (4:3) with the film introduced by Edison decades ago while HDTV has the aspect ratio at 16:9. Moreover, the movie companies such as Panavision introduced new format with different aspect ratio, (CinemaScope at 2.5:1, Panavision at 2.35:1 and widescreen at 1.8:1) and the newly developed lap-top widescreen computers have their own aspect ratio at 1.6:1.

Because the aspect ratio of the screen is unchangeable and it is impossible to prepare so many devices with different aspect ratios according to the videos presented, the conversion of aspect ratio for video is required more and more imminently with the appearance of new video formats with new aspect ratio.

Generally speaking, there have been four traditional methods to convert the aspect ratio. They accomplish the conversion by expanding, cropping, or adding black bars directly and separately. Fig 1(a) shows the frame from the original video. Fig 1(b) shows the frame from the video expanded directly (traditional expanding mode) and the distortion is obvious. Fig 1(c) shows the frame from video by cropping directly (traditional cropping mode) and the top and the bottom parts of the video are lost. Fig 1(d) shows the result by adding black bars to the left and right sides of the video (letter-box mode), and the advantage of the wide screen wastes.

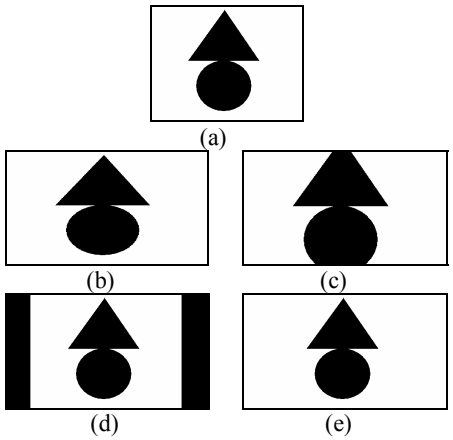


Figure 1. (a) is the frame of the original video (b) is the frame get by expand mode (c) is the frame get by crop mode (d) is the frame get by letter-box mode (e) is the frame get by our algorithm

Shih in [6] split the original picture into four regions, and use small expanding ratio to the region in the center while use large expanding ratio to the region in the border. However, the area attracting people's attention most is not always located in the center of the video and the texture on the boundary between two regions (center region and border region defined by Shih) would suffer a serious distortion.

The development of theory in saliency model acknowledged us that the area attracting people most is not always at the accurate center of the picture and it can offer information about the location of the attention area. There have been some methods for adaptive video presentation based on saliency model. The methods can be divided into two categories: cropping and re-composition, the former one is to crop the region of interest while discarding the partial

978-0-7695-3134-2/08 $25.00 © 2008 IEEE
DOI 10.1109/MUE.2008.112

environment [4] [11] and the latter one is to extract the user-interest objects and down-sample the background. [10]. Both the methods can not be extended to aspect ratio conversion. After all, they are focused on the browse video or image on mobile devices. When both the top and bottom parts of the frames are attractive the mode in [11] can not solve the problem well (for example the situation in Figure 1(c), Figure 3 (c)) and the segmentation in the latter algorithm, which is the key technique in their approach, [10] is not accurate enough for converting aspect ratio for video browsers with higher resolution than mobile devices such as TV set.

We have proposed a novel algorithm called ARCBS for aspect ratio conversion in the former paper. In ARCBS we introduced saliency model to the aspect ratio conversion for the first time and the model takes human attention into account. Another contribution in this method is that we proposed a novel cropping mode and a novel un-unique expanding mode and made use of the results of saliency analysis to adjust the parameters of cropping process and expanding process. Moreover, the cropping mode and expanding mode are combined together according to the principles from media aesthetics. [13] [14]. In this paper, we designed the evaluation system taking the subjective opinion into account to compare and analysis the effectiveness and effectiveness of the approaches.

2. Framework of ARCBS

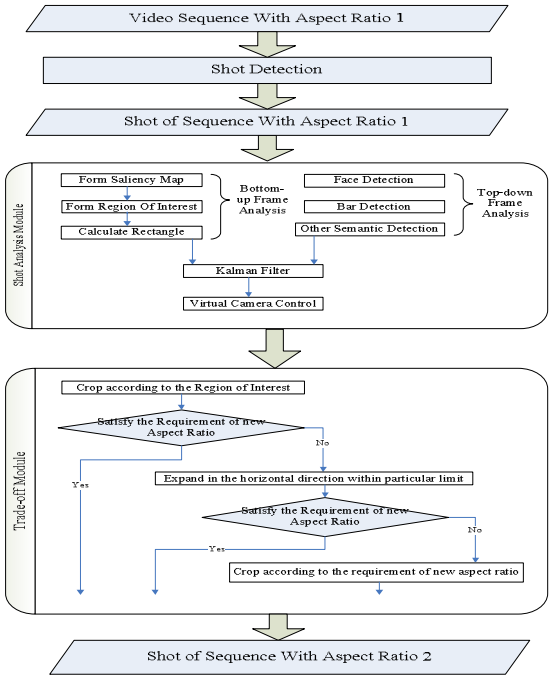


Figure 2. The framework of our algorithm

The complete framework for ARCBS is shown in Figure 2.

There are two key technologies in the framework, SA (shot analysis) and TOCE (trade-off between cropping and expanding). The shot analysis provides the information of the position of the area where people are inclined to pay attention to, and the TOCE which makes use of the results of shot analysis and obeys the principles from media aesthetics can assure the pleasant watching experience while preserving the video information as much as possible. The state machine of TOCE is shown in Figure 2.

3. Comparison and Analysis

We build up demos to validate the methods mentioned above.

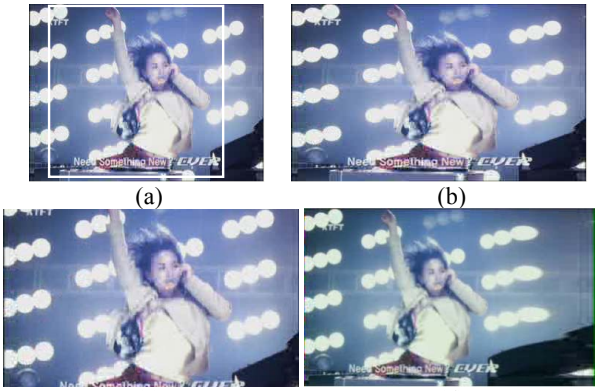
There are some examples from the experiments we made. The Fig 3 and 4 show the comparison between our approach and the traditional methods. As shown in the figure, the person in the video by expanding looks fatter than she actually is and the video by cropping lose details. (Fig 3. (c) loses the forehead of the singer and the Fig 4. (c) loses the hand and part of the letters in the frame which has a bad affection to the video comprehension). Experiment results prove that the output of the framework we proposed can achieve a more pleasing watching experience because the intact content is remained and the distortion is un-conspicuous.

The third example from our experiments is shown in Figure 5. This example is an extremely situation and does not occur frequently. It indicates that when the RecOI covers almost the whole region of the frame, our approach locates the center of cropping window at the center of RecOI and some parts of RecOI lost in the result sequence. This is the situation in which ARCBS does not perform perfect. However, ARCBS is still better than the traditional ones. In the output sequence by the traditional expanding mode the red rounded ball which is attractive becomes elliptical. And benefited from the expanding in the horizontal direction in our approach, though limited, less region of interest is cropped off than the traditional cropping mode.

The Figure 3 and 5 is the conversion from 4:3 to 16:9. Figure 4 is the conversion from 4:3 to 5:3. They show the ability of ARCBS to be adapted to various aspect ratio conversions.

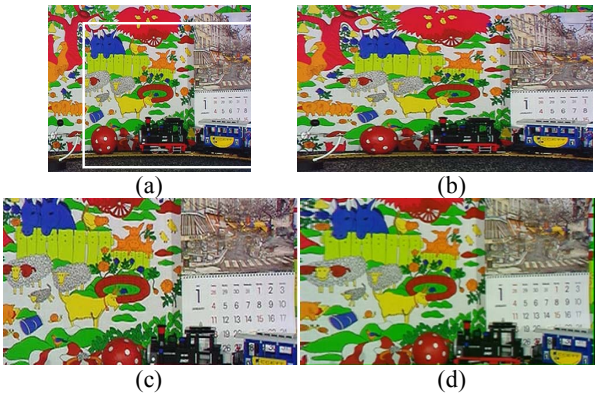
(a) (b) (c) (d)

Figure 3 (a) the original frame – the white color rectangle indicates the RecOI (b) the expanding mode (c) the cropping mode (d) our ARCBS



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

Figure 4 (a) the original frame – the white color rectangle indicates the RecOI (b) the expanding mode (c) the cropping mode (d) our ARCBS



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

Figure 5 (a) the original frame – the white color rectangle indicates the RecOI (b) the expanding mode (c) the cropping mode (d) our ARCBS

Because the evaluation of the aspect ratio conversion by different methods is subjective, we invited sixteen volunteers including five females and eleven males to answer some subjective questions. All of the participants are students from Beijing University of Posts and Telecommunications or colleagues of other research projects, and have no idea about our research work in the paper. We choose test videos from various types of videos with different aspect ratio, which were obtained from our video library or popular Web sites, and then we convert the video to have an aspect ratio at 16:9 or 2.35:1 by our demo software.

We divided our experiments into two groups, one group is to compare the traditional cropping mode with our ARCBS, and the other group is to compare the traditional expanding mode with our ARCBS. There are four questions for both groups:

Q1: Which one of the two sequences is better in term of less distortion?

Q2: Which one of the two sequences is better in terms of content preserving?

Q3: Which one of the two sequences is better in terms of the scene continuity?

Q4: Generally speaking, which one of the two sequences do you prefer?

Every participant is asked to give an assessment that ARCBS is better, worse, or the same to the compared converting mode.

The first question is concentrated in whether the key technology, expanding in our algorithm would cause visual distortion to the video sequence. The next question is to testify the ability of preserving content by cropping based on saliency model. The third question is to investigate the efficiency and effectiveness of smoothing mechanism. The last question offers a general evaluation to the conversion.

Table 1. ARCBS versus the cropping mode

	Better	The same	Worse
Q1	2.24%	79.01%	18.75%
Q2	66.64%	31.11%	2.25%
Q3	4.95%	83.00%	12.05%
Q4	88.45%	11.25%	0.30%

Table 2. ARCBS versus the expanding mode

	Better	The same	Worse
Q1	88.80%	5.57%	5.63%
Q2	4.12%	76.54%	19.34%
Q3	6.36%	82.11%	11.53%
Q4	91.78%	6.22%	2.00%

According to the statistic results in Table 1 and 2 ARCBS preponderates over the traditional conversion modes obviously.

The results of Q1 indicate that the viewers incline to ignore the distortion caused by the expanding in our

approach. One reason is that the aspect ratio of region of interest (ROI) is retained and the expanding ratio is not unique. Another reason is that by the help of cropping, the average expanding ratio is much less than that in the traditional expanding mode. Little distortion appears in only a small number of video sequences for the ROI detection is not mature enough.

The results of Q2 attest the ability of preserving content in our approach. This is because the cropping window is moved by the region of interest and the area viewers pay attention to is hardly cropped off. Moreover, the expanding horizontally reduces the area of the region to be cropped off. However, this advantage is weakened when the important information spreads all over the frame, which is not very familiar.

The results of Q3 show the efficiency and effectiveness of our smoothing mechanism.

We define PR to measure the presentation rationality of an algorithm compared with another. Using the formula below, we can get the scores to describe the presentation rationality (PR),

$$PR = \frac{1}{4 \times N} \times \sum_{j=1}^{4} \sum_{i=1}^{3} W_i \times n_{i,j} \qquad (1)$$

Where i indicates the serial number of answer, for example, W_1 denotes the answer to the question is "Better" and has a value of +1, W_2 which valued zero denotes the answer to the question is "The same", and we use W_3 to denote the answer "Worse" and value it with -1. j indicates the serial number of our questions. $n_{i,j}$ is the number of viewer who offer the ith answer to the jth question while the N is sum of participants.

We have sixteen participants in our research and calculate PR for each mode by the data in Table 1, and Table 2.

Compared with the traditional cropping mode,
The PR of our approach is 0.430.
Compared with the traditional expanding mode,
The PR of our approach is 0.509.

4. Future Work

Because the more precise the attention area determination is, the more pleasing output video sequence is, we will ameliorate the saliency model and add some other semantic models according to the users' preference by learning user feed back. Moreover, we can apply our approach to construct a browser by which users can set any aspect ratio to the video they want to watch. We also plan to accomplish the algorithm by hardware and propel the application in HDTV system. We will continue to investigate these directions in our future work.

5. Reference

[1] L. Itti, C. Koch, E. Niebur, "A Model of saliency-based visual attention for rapid scene analysis", *Journal*, IEEE Transactions on PAMI, 1998, 20(1), pp.1254-1259.

[2] X. Sun, D. Kimber, and B.S. Manjunath, ''Region of interest extraction and virtual camera control based on panoramic video capturing'', *Journal,* IEEE Transactions on PAMI, 2005, 7(5), pp.981-989.

[3] Y. F. Ma, H. J. Zhang, ''Contrast-based image attention analysis by using fuzzy growing'', *Proc,* MM'03, , Berkeley, California, USA, Nov, 2-8, 2003, pp. 374-380.

[4] X. Fan, X. Xie, H. Q. Zhou, W. Y. Ma, ''Looking into video frames on small displays'', *Proc,* MM' 03, Nov. 2-8, 2003, Berkeley, California, USA. pp. 247-250.

[5] R. S. Aygun, A. Zhang, ''Integrating virtual camera controls into digital video'', *Proc,* Multimedia and Expo, 2004. ICME '04. pp. 1503-1506.

[6] S.C. Hsia, B.D. Liu, ''An NTSC to HDTV video conversion system by using the block processing concept'', *Journal,* IEEE Transactions on consumer electronics. 1994, August, pp. 216-224.

[7] S.C. Hsia, B.D. Liu, ''A parallel video converter for display 4:3 images on 16:9 HDTV receivers'', *Journal,* IEEE Transactions on CSVT. 1996, pp. 695-699.

[8] S.K.Karungaru, M.Fukumi, N. akamatsu, ''Detection of human faces in visual scenes'', *Conference,* Australia and New Zealand intelligent information systems. 2001, Nov. pp.165-170.

[9] Z.Z. Ping, L.Kun, P.J.Hui, "Shot Boundary Detection Based on Histogram of Mismatching-Pixel Count of FMB" Conference, Industrial Electronics and applications, 2006,May, pp.1-5.

[10] W.H. Cehng, C.W.Wang, J.L.Wu, "Video Adaptation for Small Display Based on Content Recomposition" *Journal,* IEEE Transactions on CSVT, 2007, Jan, pp. 43-57.

[11] Y. D. Guo, X. D. Gu, Z. B. Chen, Q. Q. Chen, C. Wang, ''Adaptive Video Presentation for Small Display while Maximize Visual Information'', LNCS, Springer. Visual 2007, Shanghai.

[12] Y. D. Guo, X. D. Gu, Z. B. Chen, Q. Q. Chen, C. Wang, ''Denoising Saliency Map for Region of Interest Extraction'', LNCS, Springer. Visual 2007, Shanghai.

[13] C. Dorai, S.Venkatesh, "Computational media aesthetics: Finding meaning beautiful," *Journal,* IEEE Multimedia, vol.8, no. 4, Oct. 2001, pp. 10-12.

[14] D.Bordwell, K.Thompson, "Film Art: An Introduction" New York: McCraw-Hill, 2001.