1D−3

Spatial frequency analysis of movies

Marcus Csaky Ryotaro Suzuki
Media Integration and Communication Lab, ATR
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan
Tel: 774-95-1475, Fax: 774-95-1408
Email: {mcsaky,ryotaro}@mic.atr.co.jp

Abstract

Spatial frequency analysis is used to differentiate between two movie sequences. In addition, movie features such as long and close-up shots are identified within the movie segments.

1 Introduction

The method proposed here is part of the "Image Wave" study which pursues the physical attainment of rhythm information from movie clips in order to realize the automatic synchronization of multimedia content creations. With this goal in mind, we use spatial frequency analysis to sample movie data and to get an indication of the particular type of movie. This method is not the only one available but it is suitable due to its frequency domain representation which is needed for the quantitative synthesis of movies.

The two movie segments analyzed (expediently named Odessa1 and Odessa2) are scenes from the famous Russian film "Battleship Potemkin" by S.M. Eisenstein. Odessa1 is very slow "moving" with a lot of long shot camera styles whereas Odessa2 is very quickly paced and it is composed of a lot of close-ups and scene changes. The first 6 images below are two (3 images each) typical movie sequences from Odessa1 and Odessa2.

Figure 1.

Watching the contrasting movie segments creates two very different impressions in the viewer's mind. The goal of this analysis is to be able to differentiate these two selections quantitatively.

2 Method and Results

The temporal change of the images in a movie can be analyzed in a variety of different ways including spatial frequency analysis. Spatial frequency can be directly obtained from the two-dimensional (2D) Fourier transform of an image (represented by a function f(x,y) which, in this case, indicates the image brightness at a point in the plane). This spatial representation can be transformed into the spectral domain, F(u,v), via the Fourier transform. The overall normalized spatial frequency is defined as $\sqrt{(\frac{u}{X})^2 + (\frac{v}{Y})^2}$ where X and Y are the magnitudes of the horizontal and vertical image boundaries respectively[1]. Spatial frequency indicates the nature of change throughout an image or series of images. Higher spatial frequencies represent rapid image intensity changes whereas lower spatial frequencies represent smoother changes[2]. A movie selection is decomposed into its individual frames and then a 2D Fourier transform is taken of each frame's pixel brightness values. The image pixel power values are then raster scanned to associated each power value with its corresponding spatial frequency.

With this methodology in place, the spatial frequency data from the two movie segments (Odessa1 and Odessa2) is analyzed. The overall goal in this analysis is to characterize a rhythm (temporal wavelength or frequency) in a movie. The movie clips are qualitatively analyzed by the spatial frequency method and the two lowest and the two highest frequency bands are selected in order to analyze the movie trends. The high and low frequency components display the biggest difference between the two movie segments due to the fact that low frequency indicates smooth image changes while high frequency indicates rapid intensity changes.

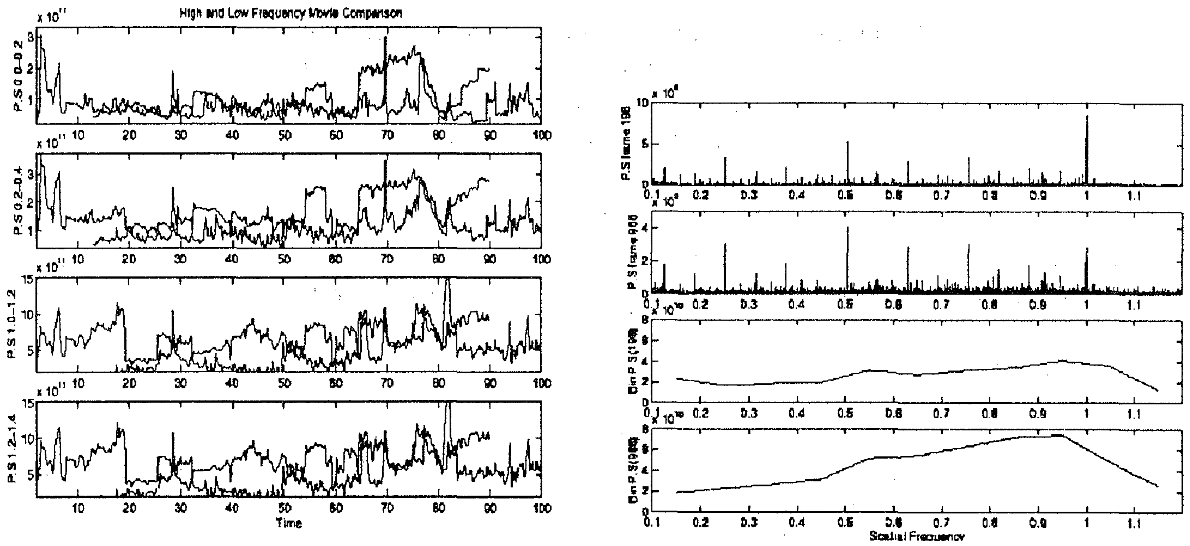

Figure 2. High and low frequency movie and long shot versus close-up comparisons

In the left plots of figure 2, the data corresponds to time series of Odessa1 and Odessa2 where Odessa2 begins at about the 2 second mark and Odessa1 begins after this time. The top two plots represent the two lowest frequency bands while the bottom two plots represent the two highest frequency bands. The x-axis is the temporal progression of the movie segment and the y-axis is the power spectra at each frequency bin. At the lowest frequency bin (0.0-0.2), the magnitude of Odessa1 and the magnitude of Odessa2 are about the same. However, at the two highest frequency ranges (1.0-1.2 and 1.2-1.4) the magnitude of the power spectra of Odessa2 is much larger than that of Odessa1. Theoretically, one would expect the low frequency component of Odessa1 to be greater than the low frequency component of Odessa2.

Therefore, it is possible to characterize the difference between Odessa1 and Odessa2 by using the spatial frequency method but it is possible to tell the difference between close-ups and long shots? The two right side images of figure 1 are an example of a close-up versus a long shot from the movie sequence. These images are compared (right plots of figure 2) using the spatial frequency method. The top two plots in the right of figure 2 display the spatial frequency along the x-axis and the power spectra for the particular close-up or long shot image along the y-axis. The bottom two plots in the right of figure 2 have frequency bins (where the frequency is binned in 0.2 width divisions) of spatial frequency along the x-axis and the power spectra along the y-axis. In the right of figure 2, the third and fourth plots from the top are the binned spatial frequency representation of the long and close-up shot respectively. One can see that the close-up image has a greater magnitude of power spectra at higher frequencies than does the long shot image.

3 Conclusion

Spatial frequency analysis is proposed to determine the composition of an image or series of images. With this method it is demonstrated that it is possible to distinguish between two different movies. It is also shown that close-up and long camera shots can be recognized.

References

[1] A. Rosenfeld and A. Kak: *Digital Picture Processing*, Orlando: Academic Press, Vol.1, 10, 1982.

[2] J. Evans and G. P. Morriss: *Digital Pictures: Representation and Compression*, New York: Plenum Press, 1998.