Video Super-Resolution using Texton Substitution

Kenji Kamimura\textsuperscript{1} Norimichi Tsumura\textsuperscript{1} Toshiya Nakaguchi\textsuperscript{1} Yoichi Miyake\textsuperscript{1} Hideto Motomura\textsuperscript{2}

\textsuperscript{1}Chiba University \hspace{1cm} \textsuperscript{2}Matsushita Electric Industrial Co., Ltd.

![Figure 1: Results of super-resolution based on the proposed method. (a): low resolution frame, (b): result of super-resolution for the training frame (closed data), (c): result of super-resolution for the test frame (open data), (d): reference frame (true high resolution).](image)

**Figure 1:** Results of super-resolution based on the proposed method. (a): low resolution frame, (b): result of super-resolution for the training frame (closed data), (c): result of super-resolution for the test frame (open data), (d): reference frame (true high resolution).

1 Introduction

A video camera produces an image sequence with a specific frame rate (e.g., 30 frames per second). This imposes limits on the spatial resolution due to limited bandwidth. Image interpolation, such as bi-cubic interpolation, can increase the resolution, but yields a blurring of edges and image details. To create plausible high-frequency details in the blurred image, super-resolution technique has been a long studied area [Baker and Kanade 2002; Capel and Zisserman 2003]. However, it is difficult to apply these methods to a video sequence, since [Capel and Zisserman 2003] requires multiple images, and [Baker and Kanade 2002] requires a high computational cost. In order to resolve these problems, we proposed a method called "texton substitution" [Kamimura et al. 2006]. In this paper, we improve the method of texton substitution by using temporal connection.

2 Texton Substitution

Figure 2 shows the overview of texton substitution for video sequence. Texton substitution is a learning based method and consists of two phases: the training and the inference phase.

In the training phase, the starting point is a set of training pairs of the high and (degraded) low resolution frames. First of all, we transform the color information of input frames into the luminance (Y) and chrominance (Cb, Cr) signals. Next, we decompose each Y image into three resolution layers using stationary wavelet transform, and as a result, each layer has three orientational properties for spatial frequency. These processes result in an eleven-dimensional feature vector (two chrominance signals and nine orientational wavelet coefficients) at each pixel in the high and low resolution training pairs. The obtained feature vectors are clustered using the K-means technique to create a small set of prototypes called textons. The texton relations between the high and low resolution images are stored in the substitution table. Since the low resolution image loses the high frequency components, many textons in the high resolution image are related to a single texton in the low resolution image. Therefore, we use the spatial connections of the textons in the low resolution image to obtain a unique solution. The texton of each pixel is connected to four adjacent pixels (crossed texton). These connections increase the apparent varieties of the texton. The substitution table is constructed by using the relation between the low resolution crossed textons and the high resolution texton.

In the inference phase, the input frame is decomposed in the same manner as in the training phase. The obtained coefficient vectors of the input image are classified into the low resolution textons and substituted based on the trained substitution table to produce the high resolution output frame. To improve the output quality, the temporal connection is considered in this paper. If the appropriate textons cannot be found from the training pairs, similar texton pairs are searched from the neighboring region of the previous frame, and the best match pair is reused in the current frame. Figure 1 shows an example of the super-resolution of a video frame. The result shows that the proposed method reproduces the details of the face well.

**Figure 2:** Overview of texton substitution method

References


*e-mail: kamimura@graduate.chiba-u.jp*