Content-Based Summarization for Personal Image Library

Joo-Hwee Lim1, Jun Li1, Philippe Mulhem2, Qi Tian1
1Institute for Infocomm Research 2IPAL-CNRS
21 Heng Mui Keng Terrace, Singapore 119613
{jooheew,lijun,mulhem,tian}@i2r.a-star.edu.sg

Abstract

With the accumulation of consumer’s personal image library, the problem of managing, browsing, querying and presenting photos effectively and efficiently would become critical. We propose a framework for automatic organization of personal image libraries based on analysis of image creation time stamps and image contents to facilitate browsing and summarization of images.

1. Summarization Framework

1.1. Photo Sequence Partitioning

Essentially, photo summarization is a mapping \( \theta \) of \( n \) photos \( P = \{p_1, p_2, ..., p_n\} \) to \( m \) photos \( S = \{s_1, s_2, ..., s_m\} \) (\( m < n \)),

\[
\theta : p_1, p_2, ..., p_n \rightarrow s_1, s_2, ..., s_m \tag{1}
\]

Here \( P \) is assumed as a sequence of photos ordered by their acquisition time. The two-level partitioning steps are seen as a mapping \( \Phi \) to group photos in \( P \) into smaller subsequences \( q_j \), and then into finer subsequences \( r_k \).

\[
\Phi : p_1, p_2, ..., p_n \rightarrow q_1, q_2, ..., q_v \rightarrow r_1, r_2, ..., r_v \tag{2}
\]

Time-based Content-based

The subsets \( q_j \) and \( r_k \) are respectively formed such that

\[
Sim_t(\text{time(last}(q_j)), \text{time(first}(q_{j+1})) > \lambda_t \tag{3}
\]

\[
Sim_c(\text{content(last}(r_k)), \text{content(first}(r_{k+1})) > \lambda_c \tag{4}
\]

1.2. Key Photo Selection

The key photo selection step maps each subsequence \( r_k = \{p_{x+1}, p_{x+2}, ..., p_{x+y}\} \) to one representative photo, called \( p_k \), of \( r_k \).

\[
\Psi : p_{x+1}, p_{x+2}, ..., p_{x+y} \rightarrow p_k \tag{5}
\]

There are various selection criteria. In our case, we consider \( \Psi \) as a function that depends on the following two aspects. First, with the information of the presence of face and the location and size of the detected face in each photo, we decide which photo contains larger and better-positioned frontal faces in the photo and that photo is selected as the key photo for the partition \( r_k \). Next we define the function \( f_k \) from \( N^+ \) (going from the time of \( p_{x+1} \) to the time of \( p_{x+y} \)) to \( N^+ \) (image numbers going from \( l \) to \( y \)), as an interpolation of the cumulative function indicating for one time point \( t \), the number of images taken before (or at) \( t \) in the current image sequence. Such a function \( f_k \) is generated using cubic-spline interpolation. The temporal representative value of one image \( p \) is then computed as the absolute value of the derivative \( f'_k = df_k/dt \) for the value \( t_p \). The key photo is the one with the maximum temporal representative value. Finally all key photos are concatenated into a sequence as a photo summary.

We have implemented and applied our photo organization and summarization framework to genuine personal photos that demonstrates the effectiveness and usefulness of our approach. We will work on more systematic user studies, automatic algorithm for setting the threshold and more powerful content-based representation for content-based partitioning.