Image and Vision Computing xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

# Event-enabled intelligent asset selection and grouping for photobook creation☆

Mark D. Wood, Madirakshi Das, Peter O. Stubler, Alexander C. Loui *

*Kodak Alaris Inc., Rochester, NY 14615, USA*

ARTICLE INFO

*Article history:*
Received 29 January 2015
Received in revised form 18 November 2015
Accepted 22 December 2015
Available online xxxx

*Keywords:*
Metadata generation
Event detection
Asset selection
Image quality
Semantic understanding
Consumer image collection
Photobook

ABSTRACT

The process of creating a photo product, such as a photobook, calendar or collage, from a large personal image collection requires intensive user effort. The primary goal of the current research was to develop an end-to-end solution to the problem of photo product generation that enables the user to complete the process with minimal edits, where the system intelligently selects assets and groups them before presenting the output to the user. The automation is driven by metadata extracted both from individual images as well as from sets of assets in a collection. In particular, we use an automatically detected event hierarchy to establish meaningful groupings in the assets, and to determine an appropriate grouping and pagination for the final product. We propose a novel intermediate construct, called a storyboard, which can be translated to different product types without recomputing the underlying metadata. In addition to chronological storyboards, we also describe a novel hybrid storyboard that joins chronological image presentation with groups of images of a common theme. A pagination algorithm uses the information in the storyboard and the product constraints to generate a product. Finally, the user is provided with a metadata-driven editing mechanism that makes it easy to change the auto-populated product. Given that the proposed system envisions user interaction in creating the final product, user studies are conducted to judge the usefulness of the system, where consumers use the system to generate a photobook with their own images.



## 1. Introduction

The proliferation of low cost, high quality digital capture devices such as digital cameras, smart phones, and wearable cameras has enhanced the opportunities for picture taking while expanding the availability of capture metadata, usage data and image-based social interactions. Since users are now amassing vast collections of media assets—both digital images and video—browsing these assets has become increasingly difficult due to the sheer volume of content. In addition, with collections that include thousands of still image and video files, selecting and presenting desired subsets of the collections becomes a formidable task for the average consumer interested in generating a photo product such as a collage, photobook, or calendar. Moreover, the large collections of personal and social media assets make it increasingly difficult for consumers to retrieve specific images or videos from various events or to select interesting scenes for creating multimedia stories or slideshows for sharing.

There are a number of commercial offerings for photobook creation (e.g., Shutterfly, Google Auto-awesome, Flickr photobook, Kodak Moments app, CeWe) for automatically taking a set of multimedia assets and creating a photobook of a specified size making use of some metadata. A precursor [1] to the system described here used a rule-based system to select assets to automatically create different types of story albums. The problem of asset selection is also described in [2]. However, asset selection is only part of the problem; a complete solution must not only select the appropriate assets, but also group them in a product-appropriate manner. Research interest in the problem of automatically creating multimedia products from consumer images was spurred on by a challenge proposed during the 2009 ACM Conference on Multimedia. Gao et al. [3], Chu and Lin [4], and Sinha et al. [5] were among the papers that addressed this challenge to various degrees. Gao et al. [3] demonstrated a complete photobook creation system combining user interaction with some automated components. Chu and Lin [4] and Sinha et al. [5] focus on summarization using temporal characteristics and duplicate detection respectively. Much of the recent research in enabling automated or semi-automated storytelling from consumer collections has been in the context of online social networks [6–8], as these can provide a rich source of metadata and contextual information. In [7], Obrador et al. describe a system that uses a consumer's online photo albums to learn the users' social context and assist them in creating a photo album for sharing. Their work includes face and image esthetic ranking in order to identify the best images to use. In [8], Saini et al. also use information from a subset of a user's

☆ This paper has been recommended for acceptance by Bogdan Ionescu.

* Corresponding author at: Imaging R&D, Kodak Alaris Inc., 2400 Mt. Read Blvd., Rochester NY 14615, USA.
*E-mail address:* alexander.loui@kodakalaris.com (A.C. Loui).

http://dx.doi.org/10.1016/j.imavis.2015.12.003
0262-8856/

friends from their social networks to enrich the available assets and metadata for creating a meaningful story. Using a limited definition of an album (i.e., small number of images and no pagination), Sadeghi et al. [9] propose context-sensitive image selection to represent vacation photo collections. Another approach being followed recently is to focus on providing a supportive user interface to help the user find and select assets to use in the product, rather than trying to automate the asset selection process. In [10], Chen et al. describe a human-centric interface for creating story narratives. In [11], Karlsson et al. describe a multiscale timeline that helps the user view the assets in the collection efficiently, even on a mobile device with a small screen.

In general, a user may spend a lot of time editing the auto-populated products generated by these systems due to the poor quality of asset selection and image placement. These systems also typically require the consumer to re-do the asset selection process all over if they wish to go from one output modality to another, such as from an $8 \times 10$ photobook to an $8 \times 12$ photobook, or from a photobook to a calendar. To address this problem, we present a novel reusable, product-independent 'story' representation that can be used to generate many different product types. Moreover, current solutions typically only order photos sequentially, either based on chronology or upload order. However, consumers often like to create photobooks that may only loosely follow chronological order. For example, a manually created photobook may often have one or more pages that are dedicated to a particular theme, where the multimedia assets associated with the theme were captured at various times. Our system provides theme-based image groupings that can be incorporated into a chronologically ordered image product. While much of the published research in this area focuses on the use of contextual information from social networks to produce meaningful albums, our work focuses on using derived metadata from image understanding algorithms. Therefore our method is applicable to unannotated images in the user's primary collection. While our approach benefits from the additional metadata available from social media platforms, it does not require it.

In this paper, we introduce an end-to-end, metadata-driven system with an easy-to-use user interface for automatically creating and editing multi-page photobooks from a consumer image collection. The distinctive contributions of our system include a) a flexible intermediate representation of a story from which multiple product modalities may be created; b) the story features an ability to include thematic groupings in addition to the traditional event-based (chronological) groupings; c) a new pagination algorithm for mapping selected images onto photobook pages that respects boundaries between groupings; and d) a metadata-aided user interface that makes it easy to find alternative images for addition or replacement.

The paper is organized as follows. In Section 2, we give an overview of the automated photobook creation system. Section 3 describes the key algorithms and component technologies employed by the system. In Sections 4 and 5, we present the storyboard generator and pagination algorithm respectively. Section 6 summarizes user testing results of the system for creating a photobook product. Concluding remarks and future work are in Section 7.

## 2. System overview

Our target user is one with a large collection of images from which they would like to make a photo product such as a photobook or calendar. This collection may consist hundreds of pictures from a single event such as a vacation or a special family gathering; or it may encompass many events over time. It can include pictures gathered from their online social networks and mobile devices. The typical manual workflow requires the user to browse through their collection and select a set of images to use for the photo product. The selection step is time consuming and usually needs further refinement during the actual photo product generation step. The user must manually lay out their set of images in the photo product using a computer, web, mobile or kiosk-based software. The layout step is extremely time consuming as well, and requires decisions on the final images to use in each page. This process then needs to be repeated if a different photo product is subsequently desired.

The main area of focus in this paper is to provide an end-to-end solution to the problem of photo product generation that presents an automatically generated photo product to the user, as well as a metadata-enhanced editing mechanism that enables them to create a final product with minimal effort.

Fig. 1 shows an overview of the system with its components. The key to making the problem manageable is metadata. Some metadata is simply extracted from the camera-generated EXIF data. Other metadata is algorithmically derived, either on an individual image basis, or from sets of images. For example, [12] face detection and color analysis are performed on individual images; event clustering, near duplicate detection, and facial clustering/people recognition are performed on sets of images. The metadata is stored in a metadata repository using the flexible RDF (Resource Description Framework) model [13].

We propose a new intermediate construct, called a storyboard, that maps the asset metadata into a form that can be translated efficiently into any product. A storyboard represents a particular way of grouping, ordering, and prioritizing the media assets in a multimedia collection. The behavior of the storyboard generator is specific to the type of storyboard that is being generated. Our storyboard generator supports multiple grouping and ordering paradigms, each with its own way for prioritizing objects. For example, the hierarchical grouping may reflect the event structure of the collection with the importance of individual images encoded in a priority score. The storyboard representation is expressed in an XML format and does not contain product-specific layout information. A given storyboard representation can be easily mapped to any number of output products, e.g., photobooks of different sizes, a collage print or a DVD.

This paper also describes the generation of a new story format called a hybrid story. This story type combines a chronological presentation of images following the underlying event structure with a thematic presentation that gathers images of a common theme together to produce a more creative output product. The themes are automatically detected based on commonality of features derived from metadata.

Each photo product provides constraints specifying the number of pages, pictures per page, etc. A pagination algorithm creates a product from the storyboard representation by selecting an appropriate subset of images using the priority scores, generating page breaks based upon the story hierarchy and image emphasis scores.

The layout of images on the pages and the selection of background images are not described in this paper, but also form a part of the overall system. The photo viewer presents the user with images and video keyframes laid out on virtual pages and presents a metadata-assisted editing capability.

## 3. Component algorithms

### 3.1. Event clustering

The temporal event clustering algorithm [12] is a fundamental image organization tool for efficiently clustering a user's images into separate events and sub-events. The algorithm leverages capture date/time metadata and color histogram information of the images for clustering. Events are detected by performing a two-means clustering on a histogram of time differences between chronologically adjacent images or videos. The output of the algorithm determines whether the separation between two adjacent images belongs to one of two classes: 1) an event boundary that corresponds to a relatively large time difference between two chronologically adjacent images, or 2) a non-event boundary that corresponds to a relatively small time difference between two chronologically adjacent images. Then within each detected event,

Please cite this article as: M.D. Wood, et al., Event-enabled intelligent asset selection and grouping for photobook creation, Image Vis. Comput. (2016), http://dx.doi.org/10.1016/j.imavis.2015.12.003