A Faster Color-based Clustering Method For Summarizing Photos in Smartphone

Kwanghwi Kim*, Seong-Min Yoon†, Sang-Hoon Kang‡ and Hwan-Gue Cho§
Dept. of Computer Science and Engineering, Pusan National University, Korea
Email: {kwanghwi*, hgcho§}@pusan.ac.kr
†Dept. of Economics, Pusan National University, Korea
Email: smyoon@pusan.ac.kr
‡Dept. of Business and Administration, Pusan National University, Korea
Email: kang.sanghoon@gmail.com

Abstract—Many people use smartphones which feature cameras and large storage capacities. They have comfortable features that allow the recording of daily life and encourage them to take hundreds of photos. However, their limited screen sizes cause browsing problems with large photo set. In addition, organizing photos by manual is a hard and tedious work. Therefore, a way to classify photos to events and a better visualization method are needed. Much research has automatically classified photos to easily present and manage collections. However, most algorithm were designed for computers, not for smartphones which have limited computing power. This paper describes a fast and lightweight spatial clustering method to summarize and visualize smartphone photos by event. We defined a nearly identical photos which are taken by people in order to get better quality photos. These photos do not need to visualize duplicately since they show same event. We calculated each photo's similarity and focused on eliminating inefficient steps to measure it. The CIELAB color space was selected to accurately measure color differences and quantize colors. We extracted nearly identical photos and cut events by optimal matching with each photo's color histogram. Our method changed the color quantization technique and accelerated clustering speed by >20 while retaining precision and recall.

I. INTRODUCTION

Smartphones are rapidly becoming very popular devices. They affects many topics of computer science such as mobile computing, intelligent information processing, interaction system and photography. In the field of photography, these phones make it easy for users to record daily life and encourage them to take more pictures. Smartphone employ global positioning system, gyroscope and accelerometer technology to better deduce information and estimate context more than typical compact cameras. However, increased photos cause some problems. Organizing these photos by manual is tedious and time-consuming work, and even smartphones’ default application do not provide these function. Also, the combination of hundreds of photos and limited screen size causes browsing problems. Since we cannot see ≥10 pictures at a time, searching for photos or determining an event’s context becomes difficult. Therefore, new methods for automatically classifying photos and effectively visualizing many photos are needed.

Although smartphone camera performance is improving, some limitations remain, such as small sensor size and lack of a manual setting. These features make people take more photos of the same scene to get a better quality photo [1]. We define these photos as nearly identical photos (NIP) and try to extract them because they do not have to show all. Figure 1 shows photos in iPhone. Photos in the red box and the blue box are taken at the same place and same scene. They are taken continuously and duplicately in order to get better quality photos such as less blurred and timely photos. These photos take a lot of space and do not need to visualize duplicately. If we find these kinds of photos and grouping, we can display more photos on a single screen. Also if we evaluate quality of each photo, we can select representative images of events. Therefore, a lightweight clustering methods for smartphone platforms is needed.

Fig. 1. An example of nearly identical photos in iPhone. Photos in the red and blue box are same scene photos. They are unnecessary to present duplicately.

To calculate the photos’ similarities, we compared the color information. We had previously researched the spatial clustering method [2] using human perceptual 25 colors [3] before. In this paper, we present the acceleration technique that selects selecting different kind of colors and pre-computed color distance information.
II. RELATED WORK

Much research has automatically classified photos to easily present and manage collections. Most research deals with how to classify each event by the photographic information: photo timestamp, color information, feature point, and geographic information. Table I shows several related works that clustered and managed photos. Cooper [4] proposed a clustering method that find each event boundary by photo taken time. Platt [5] introduced PhotoTOC, which clustered by creation time and color using an overview-detail design. Ryu [6] introduced a layout system called Photoland considered temporal-spatial information. Jang [7] introduced a clustering method for concurrent photos obtained from multiple cameras. Blazica [8] proposed a measure of the user’s affinity for a picture, the time spent viewing a picture.

Prasad [3] proposed region-based image retrieval using 25 perceptual colors. Here we employ these 25 perceptual colors in the clustering process. There is also a Color-based clustering method coupled with pyramid matching [9]. Although this approach showed a good result, it is not compatible with a smartphone’s computing power. Therefore, we needed to develop a lightweight algorithm for photo management in the smartphone environment.

![Image of Table I](image)

Table I: Several Methods and Tools for Clustering Photos

<table>
<thead>
<tr>
<th>Method</th>
<th>Clustering Criteria</th>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooper</td>
<td>Temporal and Spatial</td>
<td>Using logistic function</td>
</tr>
<tr>
<td>Jang</td>
<td>Temporal and Spatial</td>
<td>Clustering photos from multiple cameras’ photos</td>
</tr>
<tr>
<td>PhotoTOC</td>
<td>Temporal and Spatial</td>
<td>Overview-detail design</td>
</tr>
<tr>
<td>PhotoLand</td>
<td>Temporal and Spatial</td>
<td>Attaching photos with spatial information</td>
</tr>
<tr>
<td>Shoebox</td>
<td>User Affinities</td>
<td>Measuring user affinities</td>
</tr>
</tbody>
</table>

The CIELAB system is an important international standard for measuring color difference [10]. As uniform changes of components in the CIELAB color aim to uniform changes in perceived color, the relative perceptual differences between two colors in L*a*b* can be approximated by the Euclidean distance. Figure 2 demonstrates how to calculate the similarity of each photo. The input image data are converted from red-green-blue color space to the CIELAB color space to calculate human perceptual color differences. As uniform changes of components in the CIELAB color indicate uniform changes in perceived color, the relative perceptual differences between two colors can be approximated by the Euclidean distance.

![Image of Figure 2](image)

Fig. 2. A flow-chart showing spatial similarity computation. First, we converted each photo’s color model from RGB to CIELAB. Second, we quantized colors by using uniform selected colors and built color histogram. Finally, we calculated similarity values between two photos.

The color quantization method was time-consuming task. Since the old method’s 25 colors were not uniform, we should calculate each pixel’s color distance. To improve quantization speed, we partitioned CIELAB color space using a uniform cube. Figure 3 presents our partition cubes. Each color in the cubes can be quantized according to their center point. As a result, we do not have to calculate the Euclidean distance for every pixel. If a* and b* are divided 10 segments and L* into 5 segments, 500 major colors are selected. We also computed the center points distance in advance, and used this information when processing the optimal matching.

After color quantization, we built color histogram for quantized colors. The histogram is divided into dominant color areas and specific color areas. The dominant color makes the photos atmosphere and the specific color express small regions of interest. We applied optimal matching between each photo in the same window. To calculate the final similarity between two photos, we defined the similarity as:

\[
S(P_i, P_j) = w_d \times S_d(P_i, P_j) + w_s \times S_s(P_i, P_j)
\]

where \(S_d(P_i, P_j)\) denotes similarity with dominant colors and \(S_s(P_i, P_j)\) denotes similarity with specific colors. Weight

III. PROPOSED CLUSTERING METHOD

We propose a lightweight clustering methods for smartphone platforms. There are two goals in designing the spatial clustering algorithm. First, we would like to accelerate color quantization and optimal matching speed, the most time consuming portion of the process. Second, we would like to be consistent with the spatial clustering method using the 25 perceptual colors. In addition, we propose a simple quality evaluation technique which extract features by using Surf(Speeded up robust features).
values were applied to each similarity to adjust their importance. We calculated $S_d$ and $S_s$ by optimal matching between two photos histograms. The preprocessed color distance were used and reduced computational complexity.

For the final step, we predicted the optimal threshold value of similarity experimentally, and made boundaries of events. If two photos similarity is lower than threshold, a boundary is located between them. On the other hand, if two photos similarity is higher then threshold, they can be a same group, and there is no boundary.

Figure 4 shows a target photo set and photo sets clustered by manual operation and proposed method. Figure 4 (a) is a target photo set, taken by iphone, has many redundant photos. Figure 4 (b) is summarized photo set by hand considering NIP. We can notice many redundant photos are removed. Figure 4 (c) presents a photo set clustered by our method. It generated similar result to man-made.

B. Selecting Representative Photo

After clustering NIPs, we need to select representative photos of each NIP’s event. Generally, high quality photos satisfy three principles: a clear topic, a focus of attention on the subject, and the removal of objects that distract attention from the subject [11]. Most studies related to photo assessment consider color contrast and blur caused by camera shaking.

Smartphone cameras usually do not support various manual settings and people generally don’t change them. Most of their photos are taken with auto-white balance, auto-focus and auto-exposure. Also, these photos tends to have a lot of blurs because smartphone is easy to be shaken. Short focal length and small-diameter lens are unsuitable for making shallow DoF(Depth of Field) which is a technique to emphasize regions of interest. For these reasons, we focused on measure blurs, one of the most common features of photo evaluation.
We propose a simple quality evaluation technique which extract features by using Surf(Speeded up robust features) [12]. Figure 2 demonstrates how to select less blured photo. First, we extracted subject area from Exif of a photo which had recorded when people had taken a photo with auto focus or manual focus. We tried to find that region from other photos of same NIP group. Second, when corresponding region was found, we conducted Surf algorithm and detected feature descriptors. A photo which has more robust feature points is considered as more clear than others.

First, to make ground truth boundaries, we asked participants to organize photo sets. We compared our new method with the older one using various input photo sets. The input photos are described in Table II. The input photo sets were taken with a smartphone camera in different places and time periods. Each photo set were selected from participants’ photos, because they knew their context well. The number of clusters refers to groups that participants selected as NIP.

<table>
<thead>
<tr>
<th>Photo set</th>
<th>No. of photos</th>
<th>No. of clusters</th>
<th>Period(hour)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>25</td>
<td>13</td>
<td>27</td>
<td>Taiwan</td>
</tr>
<tr>
<td>B</td>
<td>50</td>
<td>24</td>
<td>2</td>
<td>Taiwan</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>69</td>
<td>51</td>
<td>Malaysia</td>
</tr>
<tr>
<td>D</td>
<td>200</td>
<td>113</td>
<td>45</td>
<td>Korea</td>
</tr>
<tr>
<td>E</td>
<td>300</td>
<td>183</td>
<td>101</td>
<td>Cyprus</td>
</tr>
</tbody>
</table>

Second, we calculated precision and recall [4]. Precision measures the proportion of falsely labelled boundaries.

\[
\text{precision} = \frac{\text{correctly detected boundaries}}{\text{total number of detected boundaries}} \tag{2}
\]

Recall measures the proportion of true boundaries detected.

\[
\text{recall} = \frac{\text{correctly detected boundaries}}{\text{total number of ground truth boundaries}} \tag{3}
\]

In this experiment we noticed that precision is more important than recall. False positive result caused information loss because different photos were classified into same group and one occluded another. False negative photos are shown seperately and easy to be corrected by simple interations. Table III presents proposed method’s precision and recall with above equations. Average precision(0.815) and recall(0.858) showed our new method retaining the old method’s precision and recall values.

<table>
<thead>
<tr>
<th>Photo set</th>
<th>No. of photos</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>25</td>
<td>0.867</td>
<td>0.867</td>
</tr>
<tr>
<td>B</td>
<td>50</td>
<td>0.800</td>
<td>0.945</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>0.806</td>
<td>0.783</td>
</tr>
<tr>
<td>D</td>
<td>200</td>
<td>0.760</td>
<td>0.867</td>
</tr>
<tr>
<td>E</td>
<td>300</td>
<td>0.844</td>
<td>0.831</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>0.815</td>
<td>0.858</td>
</tr>
</tbody>
</table>

We also compared our new method speed to the old. Figure 6 describes how much time clustering with different sizes of photo sets. This result shows that our new method increases clustering speed by about 20× while retaining precision and recall values. Since even 300 photos can be clustered in 5 seconds, our method is fast enough to use in the smartphone platform.

IV. Experiment

We presented an improved algorithms for NIP detection. In this section, we demonstrate the effectiveness of our method by comparing it to the old [2]. To compare two method, we measure precision, recall [4] and process time. We expect our new algorithm retains the old’s precision and recall with faster speed.
V. CONCLUSION

Smartphones are rapidly becoming very popular devices and encourage users to take more photos. Many photos in smartphones cause managing and visualizing problem. Therefore, new methods for automatically classifying photos and effectively visualizing many photos are needed. In this paper, we proposed a technique to improve the spatial clustering method. The method has three main parts.

1) **The CIELAB color space is partitioned uniformly.**
   We partitioned the CIELAB color space using a uniform cube, and built preprocessed data.

2) **Process speed is accelerated with preprocessed data.**
   The color quantization method was time-consuming task. Each color in the cubes can be quantized according to their center point. As a result, we do not have to calculate the Euclidean distance for every pixel and it become 20× faster.

3) **Representative photo are simply selected by using Surf.**
   We got subject area from the Exif information and compared each photo’s feature point detected by the Surf algorithm.

We could eliminate the process to calculate the Euclidean distance for color quantization and optimal matching between histograms. As a result, our new method shows an almost 20× faster than the old algorithm while retaining precision and recall.

In future works, we will develop feature-based clustering since some photos cannot cluster well with color information. We found some photos that could not clustered well by using color-based clustering method. These photos will be clustered by matching a object in the photos.

REFERENCES