# Automatic Selection of Representative Photo and Smart Thumbnailing Using Near-Duplicate Detection

Wei-Ta Chu
Dept. of Computer Science and Information Engineering
National Chung Cheng University
wtchu@cs.ccu.edu.tw

Chia-Hung Lin
Dept. of Computer Science and Information Engineering
National Chung Cheng University
lchu96m@cs.ccu.edu.tw

## ABSTRACT
This paper presents two applications about representative photo selection and smart thumbnailing using the results of near-duplicate detection. For a given photo cluster, near-duplicate photo pairs are first determined, and the relationships between them are modeled by a graph. The most typical one is then automatically selected by examining the mutual relation between them. For smart thumbnailing, we determine the region-of-interest of the selected representative photo based on locally matched feature points, which is a view different from conventional saliency-based approaches. The experiments show satisfactory performance in representative selection and promising results in ROI determination.

## Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process*, *selection process.* I.4.9 [**Image Processing and Computer Vision**]: Image Representation.

## General Terms
Algorithms, Management, Experimentation.

## Keywords
Near-duplicate detection, image selection, region-of-interest.

## 1. INTRODUCTION
Creation, display, and management of digital photos have been one of the most important activities in the digital life and in the cyberspace. People are accustomed to record their daily life or journeys by digital cameras, and share their living/travel experience on the web. For the effectiveness of managing and browsing photos, users urgently need the following functionalities.

First, users often select one representative photo for each of their albums so that visitors can understand the content inside the album at a glance. In addition, the photo owners can easily recall their life or travel experience by seeing the representative photos. Second, nowadays the browsing devices are not limited to high-definition PC monitors but also PDA or cell phones. Crudely resizing the representative photo to meet the limits of different devices would cause large information loss and diminish the advantage of "fast preview" from representative photos.

In this paper, we address these two issues by developing (1) automatic selection of representative photo and (2) smart thumbnailing based on region-of-interest (ROI). We focus on photos in journeys because the number of this kind of photo increases explosively. Moreover, these photos have clear and specific themes so that we can determine the representative photo and display the most prominent region.

Assume that we visit several scenic spots in a journey. Photos taken in the same scenic spot can be clustered together by a time-based clustering method [1]. Then, the goal of selecting the representative photo is to automatically determine which photo in a cluster best presents a scenic spot. After selecting representative photo, we want to further find the "representative region" of this photo to generate an information-rich thumbnail. The desired region can be viewed as a kind of region-of-interest (ROI), although the approach we develop is based on a viewpoint different from conventional content-based approaches.

In this paper, we advocate that both the selection of representative photos and ROI determination can be achieved by utilizing the concept of near-duplicate detection [2] (NDD). It's reasonable to assume that the most prominent landmark/view would appear several times in a time-based photo cluster. After finding the near-duplicate photos, one of them is selected as the best representation of this scene spot. Moreover, the region that mostly contributes to near-duplicate detection provides us the clues of finding the most prominent region. The result of NDD not only facilitates the selection in the inter-photo domain but also in the intra-photo domain.

The rest of this paper is organized as follows. Section 2 describes the techniques of near-duplication detection, which plays the essential role in this work. In Section 3, we model the relationships between photos as a graph, and automatically select the most representative one. In Section 4, we describe the idea of utilizing NDD to determine ROIs. Section 5 describes the experimental results and Section 6 concludes this work.

## 2. NEAR-DUPLICATE DETECTION
### 2.1 Essential Idea
The photos taken around the same place would include significant content variations. Some of them may include the most famous landmark or view, but some of them may include the shops


*MM'08*, Oct. 27–Nov.1, 2008, Vancouver, BC, Canada.

around there, pedestrians, or something that is not directly related to this scenic spot.

Figure 1 shows the content variations in the photos taken in the famous Rokuonji temple in Kyoto. From this example and many other web-based albums, we found that most travelers incline to take the landmark or famous views several times. Moreover, tourists usually take photos at some specific locations such that they can capture the canonical view as that in the postal card. According to these observations, we propose that we can approach the selection of representative photo based on near-duplicate detection, which finds the near-duplicate pairs like the fifth to the eighth photos in Figure 1.

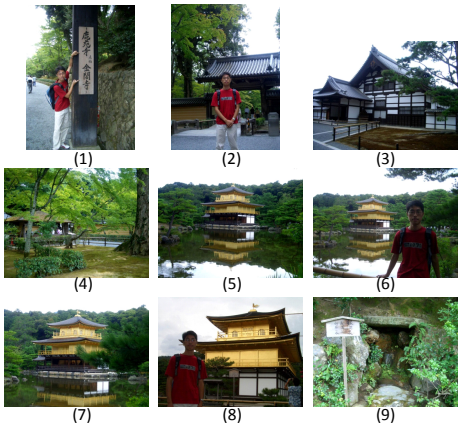



Figure 1. Photos taken around the same scenic spot.

Applications of near-duplicate detection (NDD) have been proposed for many different purposes, such as sub-image retrieval [2] and automatic image annotation [3]. In various near-duplication detection approaches, local image descriptors that capture the salient characteristics over different image scales are widely used. Among different descriptors, Lowe's SIFT (scale-invariant feature transform) feature [4] has been demonstrated to have the best performance and is used in this work.

We exploit the SIFT-based NDD method proposed by Zhao et al. [5]. This method largely reduces the false alarms caused by conventional nearest-neighbor matching approaches and increases the matching speed with a multidimensional index structure. Moreover, as the near-duplicate photos are often highly localized and spatially smooth, the correspondence of SIFT matched points have coherent patterns, which can be modeled by support vector machines (SVMs). This method obtains good balance between matching speed and matching accuracy.

## 2.2 Near-Duplication Detection Process

Given a set of photos $\mathrm{P} = \{p_1, p_2, ..., p_N\}$ that are clustered together by using the time-based clustering method [1], we determine whether a pair of photos $(p_i, p_j), i \neq j, i, j \leq N$, is near-duplicate by the following steps, as illustrated in Figure 2.

- SIFT-based matching: for any pair of photos in this cluster, the method in [5] that embeds a one-to-one symmetric criterion to filter out false matches is applied. Figure 3(b) shows the effectiveness of false alarms reduction, as compared to a conventional approach (Figure 3(a)).
- Orientation feature extraction: due to the characteristics of local coherence and spatial smoothness, the orientation of the link connecting matched points in two photos are similar. We calculate the orientation of links and quantize it into 36 levels. A 36-bin orientation histogram is then constructed. In near-duplicate pairs, the values of the orientation histogram would apparently concentrate.
- SVM-based determination model: a SVM is used to model the characteristics of the orientation histogram. We estimate the model parameters based on 40 near-duplicate pairs and non-near-duplicate pairs. At the test stage, we make a binary decision on each photo pair based on the SVM classifier.

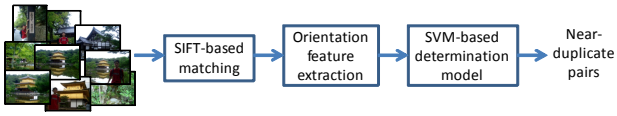



Figure 2. The process of near-duplicate detection.

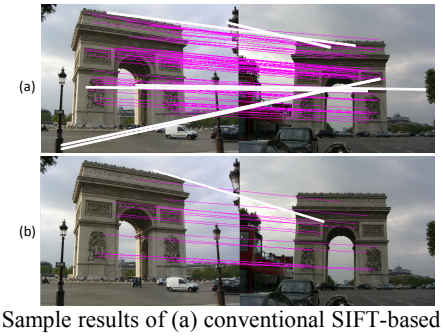



Figure 3. Sample results of (a) conventional SIFT-based matching and (b) one-to-one symmetric SIFT-based matching.

## 2.3 Sub-Clustering Before Matching

One of the critical issues in NDD is that there are tremendous pairs of photos should be examined. For example, if there are $N$ photos in a set, totally $\binom{N}{2}$ different pairs of photo are needed to be checked. To reduce the complexity, we further cluster the given set of photos based on content-based characteristics. We then perform NDD for each sub-cluster, i.e., any two photos that are in different sub-clusters would not be examined.

Because the representative landmark or view would have similar appearance, we can reasonably assume that they would be categorized in the same sub-cluster. For example, if the set of $N$ photos are categorized into $M$ sub-clusters $\{C_1, C_2, ..., C_M\}$, the total number of pairs for NDD is

$$\sum_{i=1}^{M} \binom{|C_i|}{2}, \quad (1)$$

where $|C_i|$ is the number of photos in the $i$th sub-cluster. In the case of $N = 10$, $M$=2, $|C_1| = 4$, and $|C_2| = 6$, we need originally need to check $\binom{10}{2} = 45$ photo pairs. However, we only have to evaluate $\binom{4}{2} + \binom{6}{2} = 21$ photo pairs if we perform sub-clustering first. In this work, the sub-clustering process is implemented based on RGB histograms of photos.

# 3. REPRESENTATIVE SELECTION

With loss of generality, assume that the sub-cluster $C^*$ in the set $\{C_1, C_2, ..., C_M\}$ contains the near-duplicate photos, i.e., the

photos with the landmark or specific views. Now the problem is to select one of the photos in $C^*$ to be the representative photo.

We can represent the relationship between near-duplicate photos as a non-directed, non-weighted graph $G = \langle V, E \rangle$, where $V = \{v_1, v_2, ..., v_n\}$ is a set in which any node (photo) $v_i$ is, at least one time, determined as a near-duplicate to someone else. The edge $e_{ij}$ is in $E$ if $v_i$ and $v_j$ are detected as a near-duplicate pair. Figure 4 shows an illustrative example of graphical representation of the relationships.

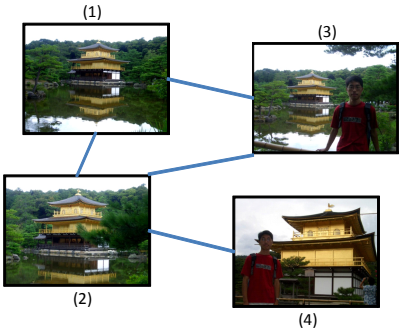



Figure 4. Relationship between near-duplicate photos.

Given this graph, we can determine the most important node by checking the "centrality value" of each node. From the idea of social network modeling, the person who is "closest" to all others plays the most important role. Similarly, we can say that the photo mostly near-duplicate to others is the most representative one. There are various measurements to evaluate the centrality value of each node. In this work, we evaluate the centrality value as the sum of in-degree of each node. Therefore, in Figure 4, the second photo would be selected as the representative photo.

# 4. SMART THUMBNAILING

In order to ease users in browsing large amounts of albums at a glance, many photo sharing platforms facilitate users to manually select a representative photo and resize it to be the epitome of each album. A user often has many albums, in each the photos in the same scenic spot are stored. We address the selection issue before. However, the resized representative photos are often suffered from severe information loss, and we may only see the rough appearance of the landmark. This situation becomes even more critical as the rapid emergence of browsing photos on low-definition mobile devices.

In this section, we further determine the "representative region" in the selected representative photo. This task is similar to finding the region-of-interest in an image. After finding the ROI, we can just extract the region and generate a better thumbnail for the representative photo.

Currently, works on ROI determination are mostly based on the bottom-up approach proposed by Itti and Koch [6]. According to human vision system, the idea is to compute the contrast of color, intensity, and orientation, and then combines these factors to construct a saliency map that describes how a photo attracts humans. In this work, we develop the determination module from a different perspective. In photos of journeys, the ROIs in representative photos are landmarks or specific views. Therefore, we advocate that it's more reasonable to find ROIs based on local feature points that contribute to near-duplicate detection, rather than color or intensity contrast.

On the basis of this idea, we can take advantage of the byproducts produced in the process of NDD. As shown in Figure 5, we found that the matched points lie on or around the most important object in photos. These points provide the foundation of linking near-duplicate objects, and the near-duplicate objects are often the landmarks or specific views that should be in ROIs.

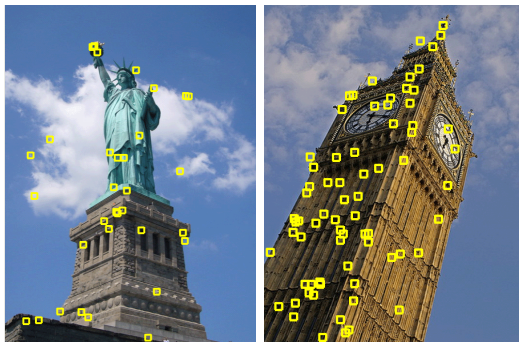


Figure 5. The matched SIFT points in representative photos.

Consider the most representative photo $p_R$ and its nearest duplicate $p_Q$. Let $\{\ell_1, \ell_2, ..., \ell_N\}$ be the set of lines connecting a pair of SIFT matched points that are in $p_R$ and $p_Q$, respectively. As described in Sec. 2.2, the orientation of these lines $o(\ell_i), 1 \le i \le N$, are gathered to construct a 36-bin orientation histogram $H$. To determine the ROI in the most representative photo, we first find the SIFT points that confidently contributes to NDD. Based on the orientation histogram, the bin with the largest histogram value is:

$$j^* = \arg\max_j H(j), \quad j = 1, 2, ..., 36. \tag{2}$$

We select the lines which orientations fall into the $i^*$-th bin or its two adjacent bins:

$$\{\ell_i | (Q(i) = j^*) \vee (Q(i) = j^* - 1) \vee (Q(i) = j^* + 1)\}, \tag{3}$$

where $Q(i)$ denotes the bin where the orientation of the line $\ell_i$ is quantized into.

Let $\{(x_1, y_1), ..., (x_M, y_M)\}, M \le N$, be the coordinates of the SIFT points that are in the representative photo and meet the eqn. (3). The left, right, top, and bottom boundaries $(x_L, x_R, y_T, y_B)$ of the desired ROI are determined by

$$x_L = \min_k x_k, \qquad x_R = \max_k x_k,$$
$$y_T = \min_k y_k, \qquad y_B = \max_k y_k,$$

where $k = 1, 2, ..., M$.

# 5. EXPERIMENTAL RESULTS

We collected 509 photos from several different users, in which 26 clusters are included. The photo sets include famous buildings like the Notre Dame and the Brooklyn Bridge, famous landmarks like the Statue of Liberty and the Eiffel Tower, and famous scenes like the Niagara Fall.

## 5.1 Performance of Representative Selection

To evaluate the performance of representative selection, which is involved with subjective judgment, we asked seven observers to give a score to each photo that is determined as a near-duplicate to others. The score ranges from one to five. Larger score is given if

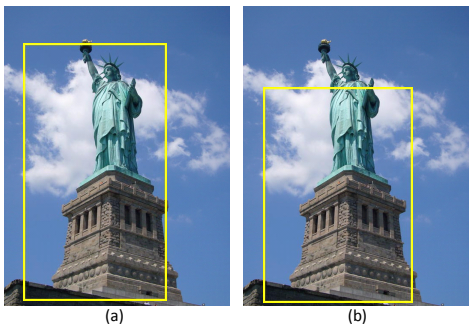
the observer thinks a photo better represents a scenic spot. To spread out the scores, observers were asked to give five or one to at least and only one of the near-duplicate photos. For each photo, the degree of representative is calculated by averaging the scores from observers.

The performance of selection is measured by the corresponding score of the selected photo. The automatic selection method obtains higher score when the selected photo better matches human's judgments. Due to the space limitation, Table 1 briefly lists the performance for five photo clusters, and the bottom row shows the overall performance for 47 photo clusters. It is not surprising that the performance would vary in different cases. Overall, the selection performance is satisfactory.

We also show the variance of human judgments. In some photo cluster, there would be many different views for the landmarks. Different observers would have varied preference for selecting the most representative view. There is a trend that in the case of larger judgment variation, the performance of selection correspondingly degrades.

Table 1. Performance of representative selection.

| Scenic spot | Score | Variance of score |
|---|---|---|
| Notre Dame | 3.14 | 0.14 |
| Statue of Liberty | 4.86 | 0.14 |
| Space Needle | 3.86 | 0.17 |
| Niagara Fall | 3.28 | 0.57 |
| Gold Gate Bridge | 2.29 | 2.90 |
| … | … | … |
| Overall | 3.21 | 0.75 |



(a) (b)

Figure 6. The results of ROI determination based on (a) filtered SIFT matched points and (b) saliency values.

## 5.2 Performance of ROI Determination

It is hardly to quantify the performance of ROI determination. Therefore, we compare the proposed method with the saliency-based approach. We use the SaliencyToolbox [7] to generate the saliency map. On the basis of saliency values, the same method described in Sec. 4 is used to determine the ROI.

Figure 6 shows the comparison of the ROI determination results. The ROI determined based on filtered SIFT matched points is notably better. The reason is that the saliency-based approach only considers the contrast in color, intensity, and orientation. On the contrary, the features used in the proposed method are directly related to the region of interest, i.e., the near-duplicate object.

## 5.3 Complexity Reduction

Table 2 shows three examples about the number of photo pairs needed to be checked in NDD with and without the sub-clustering process described in Sec. 2.3. We can see that the times of NDD is largely reduced with this process. Note that the number of reduction depends on the content characteristics of a photo cluster. If photos in the same cluster have large variations, i.e., higher entropy in this cluster, there may be more sub-clusters with similar sizes, and the number of reduction is larger.

Table 2. Number of photo pairs needed for NDD.

| Scenic spot | # photos in this cluster | # pairs w.o. sub-clustering | # pairs w. sub-clustering |
|---|---|---|---|
| Notre Dame | 19 | 171 | 40 |
| Statue of Liberty | 24 | 276 | 14 |
| Rokuonji | 15 | 105 | 48 |

# 6. CONCLUSION

With near-duplicate detection, we present automatic selection of representative photos and ROI determination. The relationships between near-duplicate photo pairs are described as a graph, and the representative photo is determined by checking the centrality value of each node. For the selected representative, the SIFT matched points are further used to locate the region of a landmark or a specific view. In the experiments, we design a scheme that not only quantifies the performance of the proposed selection method but also considers human's subjective judgments. For ROI determination, we compare the proposed method with the saliency-based approaches to show its effectiveness.

# 7. ACKNOWLEDGMENTS

This work was partially supported by the National Science Council of the Republic of China under grants NSC 96-2218-E-194-005.

# 8. REFERENCES

[1] Platt, J.C., Czerwinski, M., and Field, B.A. 2003. PhotoTOC: automating clustering for browsing personal photographs. In Proc. of IEEE Pacific Rim Conference on Multimedia, 6-10.

[2] Ke, Y., Sukthankar, R., and Huston, L. 2004. Efficient near-duplicate detection and sub-image retrieval. In Proc. of ACM International Conference on Multimedia, 869-876.

[3] Wang, X.-J., Zhang, L., Jing, F., and Ma, W.-Y. 2006 AnnoSearch: image auto-annotation by search. In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1483-1490.

[4] Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60, 2, 91-110.

[5] Zhao, W.-L., Ngo, C.-W., Tan, H,-K., and Wu, X. 2007. Near-duplicate keyframe identification with interest point matching and pattern learning. IEEE Trans. on Multimedia, 9, 5, 1037-1048.

[6] Itti, L., and Koch, C. 2001. Computational Modeling of Visual Attention, Nature Rev. Neuroscience, 2, 3, 194-203.

[7] Walther, D., and Koch, C. 2006. Modeling attention to salient proto-objects. Neural Networks, 19, 1395-1407.