

Discovering Robustness Amongst CBIR Features

Brandeis Marshall
Computer and Information Technology
Purdue University
brandeis@purdue.edu

Abstract.

Digital photography faces the challenges of image storage, retrieval and provenance at the consumer and commercial level. One major obstacle is in the computational cost of image processing. Solutions range from using high-throughput computing systems to automatic image annotation. Consumers can not dedicate computing systems to image processing and handling nor do consumers have large-scale image repositories to make automatic image annotation effective. Nevertheless, we consider an alternative approach: reducing computational cost in image processing. Using a 25,000 image collection, we consider using a sub- set of image features to evaluate image similarity. We discover several robust features displaying comparable relevancy performance with the additional benefit of reduced processing cost.

1. Introduction

The rise of digital photography has grown enormously in the past several decades due to the prevalence of digital cameras. As a byproduct, consumer photography has soared and creates a new challenge of how to organize and share personal images. The record keeping of family events, special occasions and memorable moments becomes cumbersome since storage of digital images are spread amongst folders on a personal computer, portable storage devices and photo sharing portals. No longer are photo albums purely in a physical form with dry-mount, slip-in or self-adhesive magnetic options.

Photo-taking has become relatively inexpensive since digital cameras do not require photo film. As a result, users tend to take many more photos and, some- times, these photos are of the same or similar images. An amateur photographer must then perform the photowork of downloading, filtering, storing and sharing their photo collection [15] making use of multiple storage devices. Due to the data capacities of digital cameras, users frequently opt to use their cameras as storage. If a photo is important for sharing or as a keepsake, personal computers or external storage mediums e.g. a USB drive tends to become the new storage device. For photo-sharing, images are stored on web-based servers to minimize the data load on email accounts.

The challenge that consumer photography faces is supporting a streamlined photowork process on one hardware component. Each storage device described above covers specific photowork activities. Personal computers are used as image editing and storage, USB drives serve as an archive and photo-sharing portals display the recent keepsakes. One major hinderance in properly supporting con- sumer photography is in the computational cost of image processing. Image retrieval research tends to focus on commercial photo collections which takes advantage of high-throughput computing systems. With high performance machinery, image categorization, archival and retrieval has a low computational cost and time.

Personal computers have multifaceted uses and can not dedicate all of its resources to image processing. The reduction of image processing cost can allow for improved photowork capabilities on a desktop computer. In this paper, we address the retrieval aspect by reducing the image processing cost. The feature extraction stage of image processing is costly and usually performed offline in many image retrieval research. We identify features that are common in image processing, including color histogram, color moment, homogeneous texture, texture edge and texture tamura, and test the robustness of each feature in effectively/accurately returning relevant results.

The contributions of this paper are: (1) discuss the benefits and disadvantages of common features color histogram, color moment, edge histogram, homogeneous texture, texture edge and texture tamura, and (2) compares the result accuracy of using all or a subset of these image features on a 25,000 MIRFLICKR image database.

2. Related Work

Content-based image retrieval research has three focuses: image features, data analysis methods and image datasets. Image features can be enhanced [9] or creating new features [4, 5, 16, 23] with the expectation that better semantic information can be extracted through the low-level visual descriptors. To evaluate the usefulness of the image features, clustering and/or classification techniques [4, 5, 7, 10, 22, 23] are designed to efficiently determine the most relevant images. This evaluation may be biased due to the image dataset used in the experiment. In small datasets, clustering models are applied to identify similar images where larger datasets make use of classification methods are intended to gauge the repeatability of the assignment policy for similar images. These datasets vary in their color palette (color or grayscale), number of images (1K-12K), image content (faces, landscapes, indoor/outdoor, etc.) and average image size. Ultimately, the scalability of evaluating heterogeneous large-scale image datasets has become the growing problem.

In managing digital photos, management systems tend to focus on the classification of consumer photography [1–3, 13]. The EXIF metadata embedded in jpeg images captures the timestamp and other distinctive characteristics of photos. However, if the original photograph is deleted or modified, the timestamp data is adjusted. The change in timestamp can make classification in photo management systems difficult. Boutell et al. [2, 3] consider indoor-outdoor and sunset detection using camera metadata such as timestamp data to improve classification performance. The EXIF metadata cues outperforms content-based cues, given the caveat that the EXIF metadata does not alter. EXIF metadata has also been used to cluster photographs with corresponding geographical information [19], but only newer digital cameras have GPS data in EXIF metadata.

Image collections have been the focus of a majority of image retrieval research e.g. [8,12,14]. Datta et al. [8] survey over 300 publications on image retrieval contributions, automatic image annotation, and introduce new related subfields. Huiskes et al. [12] outlines three challenges to image retrieval relevance feedback: image representation quality, cost and inconsistency of relevance feedback and topical ground truth. For consumer photography, the image quality tends to be better than commercial-use photos while relevance feedback and ground truth determinations are highly subjective making consistent evaluations and repetitive results

difficult. Ke et al. [14] finds near-duplicates for cases of forgery and copyright violations, but requires image preprocessing and numerous optimizations to achieve good performance.

To handle the complexity of image search, researchers typically either perform processing optimizations [21] or use annotations [6]. Annotations are recognized as subjective, sometimes generic, descriptions without semantic perceptions. For example, a phrase ‘little boy running away from man in park’ would be tagged as {boy, man, park} where the relationship between the boy and man and name of the boy, man and park are unknown. In consumer photography, we would like to know who are the boy and man in the picture and where the park is located. The work in [21] shows the optimization of mean and standard deviation computation through single-loop vectorization led to an average speedup of 1.85. However, the image size can be considered small (maximum size of 4096 x 4096) as consumers desire high quality pictures as keepsakes. The tradeoff observed in both approaches is the off-line preprocessing stage to control performance time and use of business-scale computing systems, which can not be replicated for personal use.

The developing subarea of consumer photography has revealed other limitations and applications that commercial-use photography do not address. For example, a human subject study in [18] documents the concern for privacy and lack of integration with current communication channels. With a reduction in image processing costs, the users gain more control of personal images without relying on photo-sharing portals. Consumer-driven applications using photography are being designed and implemented for various purposes [11]. The Bookmarkr system [11] constructs a digital-physical link between a digital photo and the print photo in order to better catalog print photos from the original digital version. Due to a heavy preprocessing load, Bookmarkr can not be easily executed on personal computers.

In these prior works, image processing is circumvented in some manner to provide better efficiency. Traditionally, multiple features represent an image digitally as more information is better for more accurate performance. To the best of our knowledge, we are the first to challenge this notion in order to make consumer photography become more accessible to real-world applications.

3. Robustness Discovery

In this section, we first discuss two existing methods, Euclidean distance and Feature Independent, in identifying image similarity using features. Then, we propose the use of a single image feature to determine closeness of an image to a query image. The Feature First relies on one image feature for all queries. For each existing and proposed method, we examine the impact of computational cost.

To compute image similarity, the image collection $C = Q \cup R_q$ where Q represents the query images and R_q denotes the set of relevant images for each query image $q \in Q$. Each image is a collection of features F . A feature captures a specific quality of an image that is represented in a numerical form. We have features F^1, \dots, F^n in which each feature F^i comprises of dimensions such as F^i_1, \dots, F^i_m . We refer to the rank order of a dimension as a *input ranker*. The input rankers serve as input into the Average method, which orders according to the average rank of each image. In our experimentation, we impose the use of the first K for the input rankers and aggregate ranker (result of Average method).

Existing Approaches

Euclidean Distance (ED).

A common approach to determining similarity is using a distance function such to maximally leverage the image contents and balance a dimension's importance and independence in relation to other dimensions. For simplicity in image processing, researchers assume that the dimensions are equally important and independent. Hence, ED is typically selected as the distance function since it is best suited for finding the similarity between two images. ED relies on magnitude comparisons of each dimension that can bias the positive and negative intermediate differences across dimensions. On a positive note, partial matching of a query image's dimension can be measured and incorporated into the final computed value.

Minkowski Form. If the dimensions are equally important and independent, then the Minkowski distance is best suited to find the similarity between two images. If $p=2$, then we have the Euclidean distance.

$$dist(q,c)=(\sum_{1 \leq f \leq n} \sum_{1 \leq d \leq m} |c_d^f - q_f^d|^p)^{1/p}$$

For ED, each dimension of the features are accessed over the entire image collection. Then the resulting similarity value for each image is sorted. The computational complexity becomes $O(c * n * m + c \log(c))$ where c is the number of images in C , n is the number of features, m is the number of dimensions per feature and cost of the mergesort algorithm.

Feature Independent (FI) [17].

Each dimension of every feature is considered individually as an independent contributor to the similarity evaluation. Thus, repeatedly observed images across features can be identified and improve their likelihood to appear at the top of the rank list. In strict contrast to more traditional methods ,such as Euclidean distance, partial matching is unavailable as only multiple appearances of exact match images can potentially elevate an image's ranking.

```

1: for feature  $f$  do
2:   for dimension  $d$  do
3:     for candidate  $c$  do
4:        $sim_d = dist(q_d^f, c_d^f)$ 
5:        $r_d = sort(sim_d)$ 
6:        $inputs \leftarrow inputs \cup r_d$ 
7:  $r_A = \mathbf{A}(inputs_K)$ 
8: return  $r_K^A$ 

```

For FI, the computational cost increases since mergesort is performed after each dimension is processed. Hence, the computational complexity becomes $O(c * n * m * c \log(c))$.

The major distinction between ED and FI is in the ranking scheme. ED is a score-based ranking function using the actual raw value of dimension values. In contrast, FI considers rank-based ordering in which dimension similarity is evaluated as a monotonically increasing function.

Score-based ordering may be more accurate in computation; however, rank-based ordering can better fulfill the information need of the request.

Features

The features commonly used in computing the Euclidean distance are color histogram, color moment, homogeneous texture, texture edge and texture tamura. There are other features such as color layout, edge histogram and scalable color, but they are not as informative. Below we discuss each selected feature in addition to its strengths and disadvantages. In our experimental evaluation, we produce the top results of the individual features according to the results observed in [17] by aggregating the dimensions using the Average method.

Color Histogram (CH). A 5-dimension image feature that aims to represent the distribution of color throughout an image using the RGB values within the pixels. We construct the color histogram from the RGB color space reducing the 256 colors into 4 bins. Bin 0 corresponds to intensities 0-63, bin 1 contains the 64-127 intensities, bin 2 contains the 128-191 intensities, and bin 3 contains the 192-255 intensities. We count the number of pixels that fall within each RGB color intensity. CH provides a general overview of the colors within an image and unaffected by image rotation; however, if the color palette is not uniform across the relevant images, then CH may not be a distinctive feature.

Color Moment (CM). The 2-dimensional feature also relies on distribution of color throughout an image. The first color moment is defined as mean:

$$E_i = \sum_{j=1}^N (1/N)(p_{ij})$$

The mean represents the average color value within the image. The second color moment is defined as standard deviation:

$$\delta_i = \sqrt{(1/N) \sum (p_{ij} - E_i)^2}$$

The second color moment represents the variance in the distribution. Image similarity between two image distributions is defined as:

$$d_{mom}(q_i, c_i) = w_1 |E_i^1 - E_i^2| + w_2 |\delta_i^1 - \delta_i^2|$$

Thus given a pair of images, they can be ranked in which higher d_{mom} values denotes less similarity and are ranked lower. CM has several variants, such as three moments (mean, variance and skewness) for each color component of RGB, that do not result in greater performance. The efficiency and effectiveness of CM tends to parallel or better that of CH with similar disadvantages.

Edge Histogram (EH). EH has 80-dimensions in order to represent the distribution of the 5 types of edges within each sub-image. For each image, it is divided into 16 equally sized sub-images, or 4x4 non-overlapping blocks as shown in Figure 1a. Within each sub-image, the edge type is classified as: vertical, horizontal, 45-degree diagonal, 135-degree diagonal and free-form edges as shown in Figure 1b. Given the 16 sub-images and 5 edge types per sub-image, we have a

total of 80 histogram bins. Hence, a sub-image's histogram denotes the frequency of each edge type revealing some of its semantics based on the sub-image's location. EH is sensitive to scaling and rotation.

Homogeneous Texture (HT). HT has 32-dimensions meant to measure the mean and standard deviation of the image, energy, and energy deviation values of Fourier transform of the image from partitioned frequency channels based on the human visual system. The human visual system approach to texture featuring relies on a sub-band representation, which divides the spatial frequency domain into channels. HT is equally partitioned in 30 degree angles along the angular direction and in an octave division along the radial direction as shown in Figure 2. This channel layout is center-symmetric resulting in sparse frequency samples in low frequency regions (higher numbered channels) and lack of texture information [20]. To increase the usefulness of texture information in this case, Radon transform scheme is applied to the images by transforming the 2D image into a 1D projection. Ideally, HT captures the frequency of repetitive patterns within an image. If relevant images do not have such patterns, HT will not provide a significant information gain.

Texture Edge (TE). With 9-dimensions, TE focuses on identifying the boundaries of objects within an image. These boundaries are located by isolating the different color patterns for a particular object as sky and tree have different color palettes. Since a color distinction is used, similarly shaped objects can be misclassified resulting in inaccurate ranking.

Texture Tamura (TT). Texture tamura, with 6-dimensions, captures the coarseness, contrast, directionality, linelikeness, regularity and roughness of an image which is based upon high level perception concepts. Coarseness computes the distance between notable grey level spatial variations to the size to the primitive elements. Contrast measures how these grey levels fluctuate within the image and assesses the level of black or white bias. Directionality compares the frequency distribution of local edges against their directional angles counterparts. Linelikeness, regularity and roughness are highly correlated to the coarseness, contrast and directionality dimensions. Roughness is a summation of the coarseness and contrast values, linelikeness is the average coincidence of edge directions and regularity computes the normalized summation of the standard deviations of coarseness, contrast, directionality and linelikeness. Unfortunately, TT tends to not be very effective with respect to finer texture distinction.

Feature First Algorithm (FF). To combat the need for a priori knowledge, we consider the

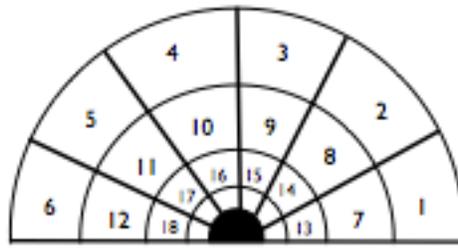
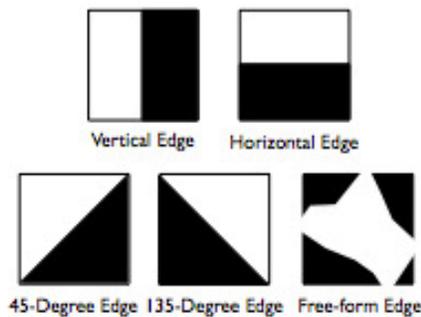


Fig. 2. Homogeneous Texture Spatial Frequency 2D Representation.



(a) sub-image block representation



(b) types of edges

Fig. 1. Edge Histogram Representation.

robustness of a single image feature. Prior work purports using multiple features to gain better information and higher performance of the images. However, multiple features also increase the computational cost of image processing. We challenge this notion through our Feature First method. In the code snippet below, we process a single feature's dimensions. We rank order the images according to a distance measure e.g. Euclidean distance. Hence, the dimensions transform into input rankers for the Average aggregation method. The result of the Average method is returned to the user.

```

1: for dimension  $F_j^i$  of feature  $F^i$  do
2:   for candidate  $c$  do
3:      $sim_d = dist(q_d^f, c_d^f)$ 
4:      $r_d = sort(sim_d)$ 
5:      $inputs \leftarrow inputs \cup r_d$ 
6:  $r_A = Average(inputs_K)$ 
7: return  $\Gamma_K^A$ 

```

When considering our Feature First algorithm, the computational cost is dramatically reduced to $O(c * i * j * c \log(c))$ where i represent one of the image feature and j denotes i 's dimensions. In the most costly case, using only feature HT would decrease the number of dimensions from 54 to

32, which is a 40% reduction in dimensionality. As consumer photography data grows, a reduction of processing time while maintaining accuracy becomes of greater importance since individuals will not have access to extensive high performance hardware and software to provide a low response time.

4. Experimental Evaluation

To discover the most descriptive feature, we test on a 25,000 image collection [12] consisting of 24 directories. In Table 1, we list each directory and the number of images containing that particular tag. These tags are very general; however, these keywords are the most popular tags identified by Flickr in 2008. These images are categorized in multiple directories. We use each image as the query image, thus executing 25,000 queries. We extract the color histogram, color moment, edge histogram, homogeneous texture, texture edge and texture tamura features for each image.

Directory	#Images	Directory	#Images	Directory	#Images
animal	3216	food	990	river	894
baby	259	indoor	8313	sea	1322
bird	742	lake	791	sky	7912
car	1177	male	6081	structures	9992
clouds	3700	night	2711	sunset	2135
dog	684	people	10373	transport	2895
female	6184	plant-life	8763	tree	4683
flower	1823	portrait	3931	water	3331

Table 1. MIRFLICKR Category Testbed

In order to check the accuracy of the query results, we compute the precision and reciprocal rank for a given K represents number of returned images ($K = 10, 20, 50$). The K -value of 10 is selected since users tend to only consider the first 10 results while the results from higher K values show the trajectory between the relevant and retrieved results. We represent the likelihood of finding relevant images in our retrieval as $l(s_K)$. In the case of the top-10, $l(s_{10}) = 10/25000 = 0.04\%$, for the top-20, $l(s_{20}) = 20/25000 = 0.08\%$ and for the top-50, $l(s_{50}) = 50/25000 = 0.20\%$. Hence the following results are based upon the retrieval of $< 1\%$ of the image collection size.

The precision P is calculated by taking as input two ranked lists m_i, m_j and finding the number of common elements in relation to the number of returned elements K . Formally, precision is defined as follows:

$$P_K(m_i, m_j) = \frac{m_i \cap m_j}{K}$$

Precision is a commonly used measure to distinguish between relevance and non-relevance. Recall, another popular performance measure, is focused on the probability of receiving relevant

results for a given query. We forego using recall since the directory sizes vary drastically e.g., 259-10373. The recall performance measure is, therefore, significantly biased in this image collection. However, neither precision nor recall indicate the degree of relevance, e.g. the rank position of relevant data. To assess relevancy based on rank position, we use the reciprocal rank measure. Formally, reciprocal rank is defined as follows:

$$RR_K(m_i, m_j) = \sum_{l=1, \dots, k} \frac{1}{l} \text{ if } m_i(l) = m_j(q)$$

where l, q ($l = q$ or $l \neq q$) refer to a position in a ranking. The RR max value at top-10, top-20 and top-50 is 2.93, 3.60 and 4.50, respectively.

Our robustness evaluation measures the quality of performance by restricting which image features contributes towards the image similarity assessment. As we limit our image features, we also examine the impact of the K-value on identifying more relevant images.

In the first experiment, we consider all image features, but we evaluate the performance using two different algorithms. The traditional Euclidean distance is compared to our Feature Independent (FI) algorithm as the performance is shown in Table 2. We bold the highest precision and reciprocal rank values for each K. The Feature Independent algorithm outperforms the Euclidean distance for all K values. Under precision, the difference between ED and FI decrease as K increases. We hypothesize that FI's image feature dimension ranking leads to more randomized order of images as K increases. The nuances of the raw image feature dimension scores are removed due to the monotonically increasing sequence of rank ordering. In contrast, FI's reciprocal rank values show that as K increases, the relevant images are appearing higher in the ordering than ED's reciprocal rank values.

	ED	FI		ED	FI
P_{10}	54.30%	58.16%	RR_{10}	1.17	1.57
P_{20}	55.38%	57.78%	RR_{20}	1.55	1.95
P_{50}	56.18%	56.89%	RR_{50}	2.07	2.47

Table 2. Euclidean Distance and Feature Independent performance

In the second series of experiments, we aim to identify robust image features that produce precision and reciprocal rank results comparable to those of FI's results. We consider image feature pairs to see if the transitivity property holds. We hypothesize that ED and FI have incorrect (noisy) image features, which is compounded by using all the image features. These feature pairs will limit and identify the noisy features from the robust features.

	P_{10}	P_{20}	P_{50}		RR_{10}	RR_{20}	RR_{50}
CHCM	59.72%	58.82%	57.15%	CHCM	1.73	2.14	2.65
CMTE	56.92%	57.65%	56.68%	CMTE	1.60	2.00	2.52
CHTE	56.85%	57.62%	56.66%	CHTE	1.58	1.98	2.50
CMTT	60.05%	58.78%	57.33%	CMTT	1.76	2.14	2.66
CHTT	59.08%	58.48%	57.10%	CHTT	1.72	2.11	2.63
TETT	57.07%	57.26%	57.00%	TETT	1.53	1.92	2.45
CMHT	58.16%	57.80%	56.91%	CMHT	1.58	1.96	2.49
CHHT	57.97%	57.74%	56.87%	CHHT	1.57	1.95	2.47
TEHT	56.86%	57.54%	56.95%	TEHT	1.55	1.94	2.45
HTTT	56.73%	57.37%	56.91%	HTTT	1.58	1.97	2.49
CMEH	56.62%	57.15%	56.90%	CMEH	1.60	1.99	2.51
CHEH	56.27%	57.17%	56.84%	CHEH	1.59	1.98	2.50
EHTE	58.27%	57.14%	57.24%	EHTE	1.58	1.97	2.50
EHTT	56.86%	57.54%	56.95%	EHTT	1.55	1.94	2.45
HTEH	57.25%	57.47%	56.70%	HTEH	1.55	1.94	2.45

Table 3. Pairwise Feature performance

Table 3 displays the precision and reciprocal rank performance for the top- 10, top-20 and top-50 results for each image feature pair. We indicate in bold the feature pair results which have better performance than Table 2, in which ED and FI methods use 134 dimensions. For each top-K, we identify the best performing image feature pair. CMTT (8 dimensions) gave the best precision of 60.05% (3.15% improvement) and 57.33% (0.77% improvement) when P_{20} and P_{50} , respectively. When P_{20} , CHCM (7 dimensions) has 58.82% precision with an 1.77% improvement over FI's top-20 precision results. The reciprocal rank results show a 9.25% improvement with CHCM image feature pair, 8.88% improvement for both CHCM and CMTT image feature pairs and 7.14% improvement for CMTT image feature pair of for the top-10, top-20 and top-50, respectively. Even though the number of relevant results increased slightly (< 4%), the newly-found relevant images can include rank positions as high as rank 11, 12 or 13 resulting in 7% – 9% performance improvement. Rank 11 would produce reciprocal rank value of $1/11 = 0.91$, rank 12 would produce reciprocal rank value of $1/12 = 0.83$, rank 13 would produce reciprocal rank value of $1/13 = 0.77$.

	P_{10}	P_{20}	P_{50}		RR_{10}	RR_{20}	RR_{50}
CM	50.98%	59.66%	57.37%	CM	1.78	2.16	2.68
CH	60.02%	59.00%	57.20%	CH	1.76	2.15	2.66
TE	56.91%	57.55%	56.89%	TE	1.61	2.00	2.53
TT	60.41%	58.86%	57.74%	TT	1.77	2.15	2.67
HT	58.16%	57.81%	56.90%	HT	1.59	1.97	2.49
EH	57.66%	58.33%	59.07%	EH	1.64	2.03	2.57

Table 4. Individual Feature performance

Given the performance of CMTT and CHCM image feature pairs, we test if a single image feature outperforms the image feature pairs and thus the FI approach. As expected, CM, CH and TT image features performed well for the top-10 and top-20 on both performance measures. However, when P_{50} , EH outperforms CMTT by 2.95%, but did not transfer to significant improvements in terms of reciprocal rank (with a value of 2.57). We hypothesize that 80-dimension EH produces a vastly different rank ordering of the images from the other image features. Thus, the EH's pairing with other image features can not rival those of CMTT and CHCM.

Through the experimental evaluation, we support the findings in [17] that the Feature Independent algorithm has better performance than the Euclidean distance with a larger image collection and more queries. Of the 6 image features, we discover CH (color histogram), CM (color moment) and TT (texture tamura) as the most robust features by providing consistently better performance over the other image feature pairs, FI and ED approaches. With all 134 dimensions needed to properly execute ED and FI approaches, time complexity (and computational cost) becomes a bottleneck; however, we discover that at most 8 dimensions can produce similar or better performance using image feature pairs or the individual image features themselves. The image feature pairs and individual image features with lower precision and reciprocal rank may be acceptable given the quicker query response time. The potential to determine and exploit the image feature robustness can allow for greater flexibility in image storage, retrieval and provenance.

5. Conclusion

The challenge that consumer photography faces is supporting a streamlined photowork process on one hardware component. Personal computers have multi-faceted uses and can not dedicate all of its resources to image processing. We identify features that are common in image processing and test the robustness of each feature in effectively returning relevant results. We executed 25K queries using each image in the MIRFLICKR image database as a query image. We compared the precision and reciprocal rank of using all the image features (through the Euclidean distance and Feature Independent methods) with those that considered image feature pairs or an individual image feature. Given the limited number of relevant images for a category with a retrieval of < 1% of the image collection, we discover that CHCM, CMTT, CH, CM and TT consistently provide good accuracy and, in most cases, better precision values than the Euclidean distance and

Feature Independent method. Hence, we can reduce our image feature dimensionality 95% (134-dimensions to maximum 8-dimensions) while maintaining the performance.

In the future, we plan to merge the image processing and image annotation concepts by investigating their level of commonality. Using the MIRFLICKR dataset, which has image annotations for most images, we want to compare similar images based on their annotations and based on their image processing for a given query image. The image processing image results may find similar which is not recorded within the image annotations. In contrast, image annotations may contain semantics that are difficult to detect using image processing. Hence, we can construct a hybrid mechanism to select image features for processing and isolate the semantic image annotations to retrieve more relevant images higher in a ranked list.

References

1. Boll, S., Sandhaus, P., Scherp, A., and Westermann, U. Semantics, content, and structure of many for the creation of personal photo albums. In ACM International Conference on Multimedia (2007), pp. 641–650.
2. Boutell, M., and Luo, J. Bayesian fusion of camera metadata cues in semantic scene classification. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2004), pp. 623 – 630.
3. Boutell, M., and Luo, J. Photo classification by integrating image content and camera metadata. In IEEE International Conference on Pattern Recognition (2004), pp. 901 – 904.
4. Carneiro, G., and Jepson, A. D. The distinctiveness, detectability and robustness of local image features. In Proceedings of European Conference on Computer Vision (2002), pp. 282 – 296.
5. Carneiro, G., and Jepson, A. D. The distinctiveness, detectability and robustness of local image features. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2005), pp. 296 – 302.
6. Chen, H.-M., Ming-Hsiu, C., Chang, P.-C., Tien, M.-C., Hsu, W. H., and Wu, J.-L. Sheepdog - group and tag recommendation for flickr photos by automatic search-based learning. In ACM Multimedia (2008), pp. 737–740.
7. Crosier, M., and Griffin, L. Using basic image features for texture classification. International Journal of Computer Vision 88, 3 (2010), 447–460.
8. Datta, R., Joshi, D., Li, J., and Wang, J. Z. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40, 1 (2008), 1–60.
9. Deselaers, T., Keysers, D., and Ney, H. Features for image retrieval: an experimental comparison. Springer-Verlag Information Retrieval 11, 2 (2008), 77–107.
10. Fu, Y., Cao, L., Gup, G., and Huang, T. S. Multiple feature fusion by subspace learning. In Proceedings of ACM International Conference on Content-based Image and Video Retrieval (2008), pp. 127 – 134.
11. Henze, N., and Boll, S. Snap and share your photobooks. In Proceeding of the ACM international conference on Multimedia (MM) (2008), pp. 409–418.

12. Huiskes, M., Thomee, B., and Lew, M. S. New trends and ideas in visual concept detection: the mir flickr retrieval evaluation initiative. In *Proceedings of the International Conference on Multimedia Information Retrieval (2010)*, pp. 527– 536.
13. Jang, C.-J., Lee, J.-Y., Lee, J.-W., and Cho, H.-G. Smart management system for digital photographs using temporal and spatial features with exif metadata. In *IEEE International Conference on Digital Information Management (2007)*, pp. 110 – 115.
14. Ke, Y., Sukthankar, R., and Huston, L. An efficient parts-based near- duplicate and sub-image retrieval system. In *ACM Multimedia (2004)*, pp. 869–876.
15. Kirk, D. S., Sellen, A. J., Rother, C., and Wood, K. R. Understanding photowork. In *In Proc. CHI 2006 (2006)*, pp. 761–770.
16. Lin, C.-H., Chen, R.-T., and Chan, Y.-K. A smart content-based image re- trieval system based on color and texture feature. *Image and Vision Computing* 27, 6 (2009), 658–665.
17. Marshall, B., and Wilson, D.-M. Applying aggregation concepts for image search. In *IEEE International Symposium on Multimedia (ISM) (2008)*, pp. 328– 333.
18. Miller, A., and Edwards, W. K. Give and take: A study of consumer photo- sharing culture and practice. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI) (2007)*, pp. 347–356.
19. Naaman, M., Song, Y. J., Paepcke, A., and Garcia-Molina, H. Automatic organization for digital photographs with geographic coordinates. In *ACM/IEEE International Conference on Digital Libraries (2004)*, pp. 53 – 62.
20. Ro, Y. M., Kim, M., Kang, H. K., Manjunath, B. S., and Kim, J. Mpeg-7 homogeneous texture descriptor. *ETRI Journal* 23, 2 (2001), 41–51.
21. Shahbahrami, A., and Juurlink, B. Optimization of content-based image re- trieval functions. In *IEEE International Symposium on Multimedia (ISM) (2008)*, pp. 607–612.
22. Wang, X.-Y., Wu, J.-F., and Yang, H.-Y. Robust image retrieval based on color histogram of local feature regions. *Multimedia Tools and Applications* 49, 2 (2010), 323–345.
23. Ziou, D., Hamri, T., and Boutemedjet, S. A hybrid probabilistic framework for content-based image retrieval with feature weighting. *Pattern Recognition* 42, 7 (2009), 1511–1519.