Measuring Semantic Similarity between Concepts in Visual Domain

Zhiyong Wang #1, Genliang Guan #, Jiajun Wang *, Dagan Feng #$

School of Information Technologies, University of Sydney, Sydney, Australia
1 zhiyong@it.usyd.edu.au

* School of Electronics and Information Engineering, Soochow University, Suzhou, China

$ Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China

***Abstract*—Concept similarity has been intensively researched in the natural language processing domain due to its important role in many applications such as language modeling and information retrieval. There are few studies on measuring concept similarity in visual domain, though concept based multimedia information retrieval has attracted a lot of attentions. In this paper, we present a scalable framework for such a purpose, which is different from traditional approaches to exploring correlation among concepts in image/video annotation domain. For each concept, a model based on feature distribution is built using sample images collected from the Internet. And similarity between concepts is measured with the similarity between their models. Hereby, a Gaussian Mixture Model (GMM) is employed to model each concept and two similarity measurements are investigated. Experimental results on 13,974 images of 16 concepts collected through image search engines have demonstrated that the similarity between concepts is very close to human perception. In addition, the entropy of GMM cluster distributions can be a good indication of selecting concepts for image/video annotation.**

I. Introduction

Exploring semantic similarity between concepts has been indispensable in the natural language processing domain[1] and played an important role in many applications such as language modeling and information retrieval[2]. For example, concept similarity has been utilized to improve the relevance of the search in information retrieval through query expansion[3].

Recently, due to the so-called "semantic gap" between multimedia content representation based on low level features (e.g. color and shape) and high level meanings of human understanding[4][5], there is a significant shift from content based retrieval to concept based semantic retrieval in multimedia information retrieval domain. In order to achieve semantic representation of multimedia data, various annotation techniques (e.g. image annotation) have been proposed to bridge the semantic gap by attaching semantic labels to multimedia data (e.g. image)[6]. Therefore, measuring semantic similarity between concepts in multimedia domain has been an emerging issue similar to natural language processing, while multimedia data is annotated with linguistic terms.

It is also expected that many tasks such as visual information annotation and visual information retrieval can be better conducted, if such similarity information can be effectively utilized. It is believed annotation accuracy can be improved

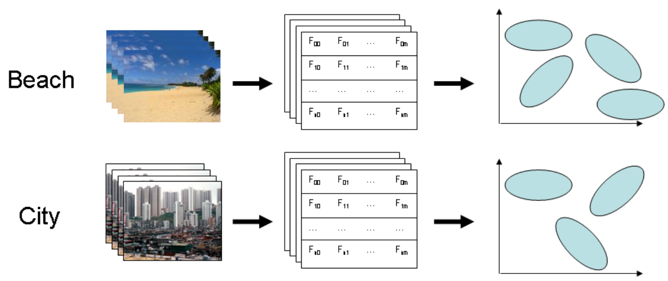

Fig. 1. Diagram of the proposed framework.

if concepts in the annotation lexicon are not similar, since similar concepts generally result in overlaps in feature space and impose more difficulties for machine learning tasks. As also shown in[7], correlation among concepts can be utilized to improve video annotation performance.

In addition, it is worthwhile to explore visual semantic similarity automatically since semantic similarity between concepts evolves over time and across domains[1], although semantic similarity information can be directly borrowed from that of the natural language processing domain and LSCOM (Large Scale Concept Ontology for Multimedia) [8] has been manually created.

Koskela *et al.* [9] proposed a clustering-based framework to analyze concept similarity by modelling semantic concepts with cluster histograms, which for the first time explicitly addressed the issue of concept similarity in multimedia domain. They also investigated various methods for assessing concept similarity by accounting for different information sources such as semantic network WordNet. However, there is information loss while visual feature vectors were quantized to build a discrete codebook of clusters by using Self-Organizing Map (SOM) approach. In addition, clustering was conducted on the images of all the concepts, which makes their method not scalable to a large number of concepts.

Hereby, we propose to model each concept individually, rather than build clusters from the image repository of all the concepts. Therefore, the proposed framework is scalable to the number of concepts, since the computational complexity is only mainly affected by the number of sample images of a specific concept. A Gaussian Mixture Model (GMM)

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.

is utilized to model each concept with low-level features directly, instead of discrete clusters. And semantic similarity between concepts are based on the distance between their GMMs. As shown in Figure 1, for a given concept, only the visual feature vectors of its sample images are directly utilized to construct the GMM. Similar to [10] where Yanai and Barnard proposed to measure visualenss of adjective terms with entropy, entropy of the weights of clusters in a GMM is utilized to measure concept "visualness" which is an indication of concept diversity in our method.

Note that concept similarity is different with concept correlation. Concept similarity characterizes content similarity between concepts (i.e. how closely two concepts are about the same semantics); while concept correlation reveals relatedness between different concepts. For example, *car* is semantically similar with *vehicle*, and is correlated with *road*. Concept correlation has been utilized to improve annotation performance through context fusion and integration[11][12][7]. The relationship between concepts is mostly modelled through co-occurrence statistics (e.g. mutual information) or joint probability, since one image (or video) is accompanied with several concepts and all the concepts share the same dataset.

In summary, the contributions of the proposed method are:

1) directly utilizing visual features for concept modeling without quantizing visual feature vectors, and measuring semantic similarity using distance between continuous distributions;
2) being scalable to the number of concepts, since the each concept is modeled individually based on its own sample images;
3) computational efficiency since clustering images of one concept is easier than clustering images of all the concepts together and the overall computational complexity is mainly affected by the modeling complexity of one concept; and
4) conducting experiments on data collected from the open web, instead of closed datasets manually annotated, which a) ensures that each concept has enough sample images, and b) is more challenging due to noise images (e.g. irrelevant images).

II. Image Representation

Many feature extraction methods have been proposed to characterize image contents[4]. Ideally, objects contained in images can be extracted and described to match human perception, which significantly relies on image segmentation techniques. As indicated in [13], annotation performance varies due to segmentation errors. Sometimes, simple uniform partition based approaches outperformed segmentation based approaches. Carneiro *et al* demonstrated that simple uniform partition can achieve best performance[14]. Therefore, each image was uniformly partitioned into 5×5 regions, which is similar to [9] denoted with SOM, Each region was represented with 32-dimension low-level features including 12-dimension color features based on color moments of R, G, B and Gray components, and 20-dimension shape and texture features. The shape and texture features have been widely used for image annotation, of which more details can be found in [15]. Note that any better image representation techniques can be incorporated into our framework.

III. Concept Modeling with Gaussian Mixture Model

We assumed that each image of a given concept is generated by a mixture of Gaussians. Therefore, the distribution of a random variable $X \in \mathbf{R}^d$, is a mixture of k Gaussians if its density function is

$$f(x|\theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\sum_j|}} \exp\{-\frac{1}{2}(x-\mu_j)^T \textstyle\sum_j^{-1}(x-\mu_j)\}, \tag{1}$$

where $\theta = \{\alpha_j, \mu_j, \sum_j\}_{j=1}^{k}$ consists of:

- $\alpha_j \geq 0$, and $\sum_{j=1}^{k} \alpha_j = 1$;
- α_j, μ_j, and $\sum_j$ are the weight, mean, and covariance matrix of the j-th Gaussian mixture, respectively.

Given a set of feature vectors $x_1, ..., x_n$, the maximum likelihood estimation is

$$\begin{aligned} \theta_{ML} &= \arg\max_{\theta} L(\theta|x_1, ..., x_n) \\ &= \arg\max_{\theta} \sum_{i=1}^{n} \log(f(x_i|\theta). \end{aligned}$$

And the Expectation-Maximization (EM) algorithm is used to determine the maximum likelihood parameters θ_{ML} [16].

Due to the fact some concepts are specific (e.g. car) and some concepts are abstract (e.g. city) which covers more diverse contents, concept diversity provides evidence of learning difficulty for a given concept. The more specific the concept is, the easier concept learning will be. Since each sample image of a given concept is represented with low-level features, each Gaussian component of the GMM represents a cluster in the low-level feature space. If a concept is illustrated with a small number of dominant clusters, the entropy of the cluster weights will be low; on the contrary, if the concept includes diverse contents contributing to a large number of random-like clusters, the entropy of the cluster weights will be high. Therefore, the entropy of the cluster distribution of a given concept w, as defined in Equation (2), can be utilized to measure the diversity of a concept.

$$H(w) = -\sum_{i=1}^{k} \alpha_i \log \alpha_i. \tag{2}$$

In addition, those clusters indicate the "aboutness" of the concept. For example, it is expected that the dominant clusters of *sunset* images will be of features representing *sky* and *cloud* and the clusters in *city* images will be of contents related to *city* such as *buildings* and *face*.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.

IV. Similarity Measurement

There are many known methods having been proposed to measure distribution distance for discrete distributions (e.g. histogram), such as Kullback-Leibler (KL) Divergence and Jensen-Shannon (JS) Divergence. However, these methods cannot be easily extended to continuous distributions. For example, there is no closed-form expression for the KL Divergence between two mixtures of Gaussians. Generally, Monte-Carlo simulations are applied to approximate the KL Divergence between two mixtures of Gaussians, f and g,

$$KL(f||g) = \int f \log \frac{f}{g} \approx \frac{1}{n} \sum_{t=1}^{n} \log \frac{f(x_t)}{g(x_t)}. \tag{3}$$

where x_1, ..., x_n are sampled from $f(x)$.

Due to high computational complexity of Monte Carlo simulations, many approaches have been proposed to achieve efficient and accurate approximation of KL Divergence. In this paper, we investigate two distances for similarity measurement, parametric distance [17] and unscented transform based distance [18].

A. Parametric Distance

Let $f(x) = \sum_{i=1}^{n} \alpha_i f_i(x)$ and $g(x) = \sum_{j=1}^{m} \beta_j g_j(x)$ be two mixture densities where $\alpha = \{\alpha_i, ..., \alpha_n\}$ and $\beta = \{\beta_i, ..., \beta_m\}$ are discrete distributions, and f_i and g_j are arbitrary continuous densities. The overall distance between $f(x)$ and $g(x)$ is defined as

$$Dp(f,g) = \min_{w=[w_{ik}]} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ik} d(f_i, g_j), \tag{4}$$

where

$$w_{ik} \geq 0, \qquad 1 \leq i \leq n, 1 \leq j \leq m,$$
$$\sum_{i=1}^{n} w_{ij} = \beta_j, 1 \leq j \leq m, \quad \textstyle\sum_{j=1}^{n} w_{ij} = \alpha_i, 1 \leq i \leq n,$$

The solution optimizing w_{ij} is posed as a linear programming problem and appropriate distance function $d(f_i, g_j)$ can be chosen in terms of specific applications. In our work, we assume that each Gaussian is of the form $N(\mu, \sigma)$, and define the following distance function

$$d(f_i, g_i) = (\sum_{k=1}^{d} |\mu_k^{f_i} - \mu_k^{g_i}|^2 + \sum_{k=1}^{d} |\sigma_k^{f_i} - \sigma_k^{g_i}|^2)^{\frac{1}{2}}. \tag{5}$$

B. Unscented Transform based Distance

Goldberger *et al.* proposed an efficient method to approximate KL Divergence between two Gaussian mixtures[18]. According to their work, KL Divergence based on the unscented transform[19] gives excellent results, with slight computational overhead.

The unscented transform is to generate a set of $2d$ "sigma" points based on a d-dimension normal random variable $x \sim t(x) = N(\mu, \sum)$ as follows

$$\begin{aligned} x_k &= \mu + \left(\sqrt{d\textstyle\sum}\right)_k, \ k = 1, ..., d \\ x_d + k &= \mu - \left(\sqrt{d\textstyle\sum}\right)_k, \ k = 1, ..., d \end{aligned} \tag{6}$$

where $(\sqrt{\sum})_k$ is the k-th column of the matrix square root of $\sum$. Given these "sigma" points, we can have the following approximation

$$\int t(x) h(x) dx \approx \frac{1}{2d} \sum_{k=1}^{2d} h(x_k). \tag{7}$$

Given two mixture densities:

$$f(x) = \sum_{i=1}^{n} \alpha_i N(\mu_{1,i}, \sum_{1,i}), \tag{8}$$

$$g(x) = \sum_{i=j}^{m} \beta_j N(\mu_{2,j}, \sum_{2,j}). \tag{9}$$

According to Equation (7), the approximation of $\int f \log g$ is

$$\frac{1}{2d} \sum_{i=1}^{n} \alpha_i \sum_{k=1}^{2d} \log g(x_{i,k}), \tag{10}$$

where

$$\begin{aligned} x_{i,k} &= \mu_{1,i} + \left(\sqrt{d \textstyle\sum_{1,i}}\right)_k, k = 1, ..., d, \\ x_{i,d+k} &= \mu_{1,i} - \left(\sqrt{d \textstyle\sum_{1,i}}\right)_k, k = 1, ..., d. \end{aligned} \tag{11}$$

KL Divergence between two mixtures of Gaussians, f and g can be rewritten as,

$$KL(f||g) = \int f \log \frac{f}{g} = \int f \log f - \int f \log g. \tag{12}$$

In general, similarity is for ranking purpose. Hence, KL Divergence between f and g is equivalent to $-\int f \log g$ which can be further approximated with Equation (10).

V. Experimental Results

Experiments were conducted on 16 concepts as shown in Table I. For each concept, we attempted to collect 1000 images (the maximal limit) by using Yahoo! image search engine[20]. Due to some dead links, in total, a set of 13,974 images were collected for these concepts.

Similar to [9] denoted with SOM, each image was uniformly partitioned into 5×5 regions. Each region was represented with 32-dimension low-level features including 12-dimension color features based on color moments of R, G, B and Gray components, and 20-dimension shape and texture features. The shape and texture features have been widely used for image annotation, of which more details can be found in [15]. In our approach denoted with GMM, each concept is modelled with a 25-component GMM, and in SOM all the visual features were clustered into 256 clusters.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DIVERSITY RANKING OF 16 CONCEPTS.

Rank	GMM		SOM	
	Concept	Entropy	Concept	Entropy
1	clouds	3.083	clouds	4.636
2	sunrise	3.124	sunset	4.764
3	sunset	3.128	sunrise	4.827
4	sky	3.147	f-16	4.899
5	face	3.160	mountain	4.917
6	mountain	3.161	sky	4.935
7	f-16	3.172	bay	4.953
8	boeing	3.176	beach	4.981
9	ships	3.180	boeing	5.012
10	helicopter	3.181	helicopter	5.038
11	birds	3.181	ocean	5.084
12	buildings	3.181	buildings	5.097
13	beach	3.183	ships	5.107
14	city	3.188	city	5.108
15	ocean	3.192	face	5.115
16	bay	3.196	birds	5.174

A. Concept Diversity

The 16 concepts were ranked in terms of their diversity defined in Equation (2), as shown in Table I. It is observed that both GMM and SOM approaches have similar ranking. That is, the top ranked concepts (e.g. *clouds*) are more specific than those bottom ranked concepts (e.g. *city*), which comply with human perception. For instance, most *clouds* images include mainly cloud and sky regions. These regions can be clearly characterized with several clusters (e.g. normal white clouds and sunset/sunrise clouds) of low-level features as shown in Figure 2(a). On the contrary, *city* images as shown in Figure 2(b) include a broad range of contents (e.g. person, street, and buildings) having different visual attributes, which results in high entropy value. In addition, such cluster information reveals the correlation between concepts such as *city* and *person*.

It is also noticed that GMM and SOM significantly disagree on concepts *bay* and *face* as shown in Figures 2(c) and 2(d). After carefully examining the images collected for these two concepts, we found that 1) *bay* images exceptionally includes a number of images which visually have nothing to do with concept *bay* (i.e. water side scenery); 2) *face* images does focus on close-up images of human heads. Therefore, our GMM based approach can more faithfully represent content diversity for these two concepts even using fewer clusters than SOM.

B. Semantic Similarity

As shown in Table II, concept similarity is ranked for each concept, where PD and UTD denote the parametric distance and unscented transform based distance of the GMM based approach, respectively. It is observed that most similar concepts comply with human understanding, such as *clouds* vs *sky*, *sunset* vs *sunrise*, since they share similar contents.

However, there are also negative examples such as *mountain* is ranked as the most similar to *city*. It is due to three facts: 1) images in these two concepts do share some contents such as *sky*. Therefore, the contribution of certain content should be weighted since *sky* information is ubiquitous; 2) visual features do not discriminate some contents semantically; and 3) noise images.

C. Computational Complexity

Both our approach and SOM based approach rely on visual feature clustering. Suppose that there N concepts, and each concept have M sample images. For SOM based approach, clustering should be conducted on $N \times M$ images; Nevertheless, clustering only M images is required for our proposed approach. Since our proposed approach models each concept individually, computation can be even speed up through parallel computing. In our experiments, clustering 13,974 images takes 1.5 hours on PC Intel 2.66GHz CPU and 3.0G Memory running Linux. It only takes about 40 minutes for our proposed approach to model all the 16 concepts.

VI. CONCLUSION

We present a method measuring semantic similarity between concepts which are modelled with Gaussian Mixture Models by using an efficient approximation between two mixtures of Gaussians. Experimental results on 13,974 images of 16 concepts obtained through image search engine has demonstrated that the proposed method can reveal semantic similarity in visual domain. In addition, the entropy based on the weights of Gaussian components can be utilized to guide the selection of annotation terms. The proposed method is also scalable to the number of concepts and computational efficient since each concept is modelled individually from its sample images. The immediate work in the future is to eliminate noise images, to select the optimal number of mixture components by using Minimum Description Length (MDL) criterion, and to conduct more comprehensive experiments with more concepts to obtain more general observations. It is also interesting to investigate the effects of different image representations in our proposed framework.

ACKNOWLEDGMENT

The work presented in this paper is partially supported by ARC and Hong Kong Polytechnic University grants.

REFERENCES

[1] D. Bollegala, Y. Matsuo, and M. Ishizuka, "Measuring semantic similarity between words using web search engines," in *International Conference on World Wide Web (WWW)*, May 2007.

[2] Y. Li, Z. A. Bandar, and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Trans. on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 871–882, August 2003.

[3] M. Mitra, A. Singhal, and C. Buckley, "Improving automatic query expansion," in *International ACM/SIGIR Conference on Research and Development in Information Retrieval*, 1998.

[4] D. Feng, W. C. Siu, and H. J. Zhang, *Multimedia Information Retrieval and Management*. Berlin Heideberg, Germany: Springer-Verlag, 2003.

[5] M. S. Lew, N. Sebe, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1–19, February 2006.

[6] M. Worring and G. Schreiber, "Semantic image and video indexing in broad domains," *IEEE Trans. on Multimedia*, vol. 9, no. 5, pp. 909–911, August 2007.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.

(a) clouds (b) city

(c) face (d) bay

Fig. 2. Image collages of four concepts.

[7] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang, "Correlative multi-label video annotation," in *The ACM International Conference on Multimedia*. Augsburg, Germany, September 2007.

[8] M. Naphade, J. R. Smith, J. Tesic, S. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis, "A large-scale concept ontology for multimedia," *IEEE Multimedia*, vol. 13, no. 3, pp. 86–91, July 2006.

[9] M. Koskela, A. F. Smeaton, and J. Laaksonen, "Measuring visual similarity between concepts," *IEEE Trans. on Multimedia*, vol. 9, no. 5, pp. 912–922, August 2007.

[10] K. Yanai and K. Barnard, "Finding visual concepts by web image mining," in *International Conference on World Wide Web (WWW)*, 2006.

[11] A. B. Benitez and S.-F. Chang, "Semantic knowledge construction from annotated image collections," in *The IEEE International Conference on Multimedia & Expo (ICME)*. Lausanne, Switzerland, August 2002.

[12] M. R. Naphade, I. Kozintsev, and T. Huang, "Factor graph framework for semantic video indexing," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 40–52, January 2002.

[13] K. Barnard, P. Duygulu, R. Guru, P. Gabbur, and D. Forsyth, "The effects of segmentation and feature choice in a translation model of object recognition," in *The IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2. Wisconsin, USA, June 2003, pp. 675–682.

[14] G. Carneiro, A. Chan, P. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394–410, March 2007.

[15] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, March 2003.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.

TABLE II

SIMILARITY RANKING FOR EACH CONCEPT.

bay	PD	ocean	sky	clouds	birds	sunset	sunrise	face	beach	ships	helicopter	boeing	buildings	f-16	mountain	city
	UTD	birds	ocean	ships	f-16	face	beach	sunrise	sky	clouds	city	sunset	mountain	buildings	boeing	helicopter
	SOM	beach	ocean	birds	buildings	mountain	city	helicopter	boeing	face	ships	sky	f-16	clouds	sunrise	sunset
beach	PD	clouds	sky	ocean	face	bay	birds	sunset	sunrise	boeing	helicopter	ships	buildings	f-16	mountain	city
	UTD	birds	ships	clouds	city	ocean	sky	sunset	f-16	mountain	bay	boeing	buildings	face	helicopter	sunrise
	SOM	bay	birds	face	ocean	city	buildings	mountain	helicopter	sky	boeing	ships	f-16	clouds	sunrise	sunset
birds	PD	clouds	sky	ocean	bay	sunset	face	sunrise	beach	boeing	helicopter	ships	buildings	f-16	mountain	city
	UTD	beach	ships	face	clouds	ocean	sunset	sunrise	city	f-16	sky	bay	boeing	buildings	mountain	helicopter
	SOM	ocean	face	buildings	city	beach	helicopter	bay	ships	mountain	boeing	sky	f-16	sunrise	clouds	sunset
boeing	PD	clouds	sky	ocean	birds	face	helicopter	sunset	bay	buildings	sunrise	f-16	ships	beach	mountain	city
	UTD	ships	birds	face	clouds	sunset	beach	f-16	ocean	sky	sunrise	city	bay	mountain	helicopter	buildings
	SOM	helicopter	ships	ocean	f-16	buildings	bay	birds	beach	sky	face	city	mountain	clouds	sunrise	sunset
build-ings	PD	clouds	sky	ocean	face	birds	sunset	boeing	sunrise	bay	helicopter	beach	ships	f-16	mountain	city
	UTD	ships	face	boeing	ocean	city	birds	sunrise	f-16	sky	mountain	beach	sunset	clouds	bay	helicopter
	SOM	ocean	helicopter	city	birds	ships	bay	face	beach	boeing	mountain	f-16	sky	sunrise	clouds	sunset
city	PD	mountain	face	sunset	ocean	sky	bay	sunrise	birds	beach	clouds	ships	buildings	boeing	f-16	helicopter
	UTD	mountain	face	sunset	sunrise	ships	ocean	beach	f-16	sky	clouds	bay	birds	boeing	buildings	helicopter
	SOM	buildings	birds	face	ocean	beach	bay	ships	mountain	helicopter	boeing	sky	f-16	sunrise	sunset	clouds
clouds	PD	sky	sunset	sunrise	ocean	face	helicopter	boeing	birds	f-16	bay	beach	ships	buildings	mountain	city
	UTD	beach	face	sunset	birds	sunrise	f-16	boeing	ships	ocean	sky	city	bay	mountain	buildings	helicopter
	SOM	sky	sunrise	f-16	sunset	ocean	boeing	helicopter	ships	beach	bay	mountain	face	birds	buildings	city
f-16	PD	clouds	sky	ocean	sunset	boeing	face	sunrise	birds	helicopter	ships	bay	beach	buildings	mountain	city
	UTD	ships	boeing	birds	face	sunrise	beach	ocean	clouds	sunset	city	buildings	bay	sky	mountain	helicopter
	SOM	helicopter	boeing	ships	ocean	buildings	birds	face	bay	sky	beach	clouds	city	mountain	sunrise	sunset
face	PD	clouds	sky	sunset	sunrise	ocean	birds	bay	beach	boeing	ships	helicopter	buildings	mountain	f-16	city
	UTD	sunrise	city	ships	sunset	birds	ocean	sky	beach	mountain	f-16	clouds	buildings	boeing	bay	helicopter
	SOM	birds	city	beach	ocean	buildings	ships	helicopter	bay	sky	mountain	f-16	boeing	sunrise	sunset	clouds
heli-copter	PD	clouds	sky	ocean	sunset	birds	sunrise	boeing	face	bay	ships	buildings	f-16	beach	mountain	city
	UTD	ships	birds	ocean	face	sunrise	boeing	buildings	sky	city	f-16	clouds	sunset	mountain	beach	bay
	SOM	ships	ocean	f-16	buildings	boeing	birds	face	bay	sky	beach	city	mountain	clouds	sunrise	sunset
moun-tain	PD	sunset	sky	city	face	clouds	sunrise	ocean	bay	birds	beach	buildings	ships	boeing	f-16	helicopter
	UTD	city	face	sunset	sunrise	ships	sky	ocean	clouds	f-16	birds	beach	bay	boeing	buildings	helicopter
	SOM	bay	ocean	birds	beach	buildings	city	sky	helicopter	face	boeing	ships	f-16	clouds	sunrise	sunset
ocean	PD	clouds	sky	sunset	sunrise	bay	face	birds	beach	helicopter	boeing	ships	buildings	f-16	mountain	city
	UTD	ships	birds	sunrise	sky	face	city	mountain	sunset	beach	clouds	f-16	buildings	boeing	bay	helicopter
	SOM	helicopter	buildings	ships	birds	bay	boeing	sky	city	face	beach	f-16	mountain	clouds	sunrise	sunset
ships	PD	clouds	sky	ocean	sunset	birds	sunrise	face	bay	boeing	helicopter	beach	f-16	buildings	mountain	city
	UTD	ocean	birds	face	sunrise	city	sky	sunset	beach	clouds	f-16	boeing	mountain	bay	buildings	helicopter
	SOM	helicopter	ocean	buildings	boeing	f-16	birds	face	city	sky	bay	beach	mountain	clouds	sunrise	sunset
sky	PD	clouds	sunset	sunrise	ocean	face	birds	bay	helicopter	beach	boeing	ships	buildings	f-16	mountain	city
	UTD	ocean	sunset	birds	face	mountain	ships	sunrise	city	clouds	beach	boeing	f-16	bay	buildings	helicopter
	SOM	ocean	clouds	sunrise	helicopter	ships	sunset	face	mountain	beach	bay	birds	buildings	boeing	f-16	city
sunrise	PD	sunset	clouds	sky	ocean	face	birds	bay	helicopter	ships	beach	boeing	mountain	buildings	f-16	city
	UTD	sunset	ships	ocean	beach	face	birds	f-16	sky	city	clouds	boeing	bay	mountain	buildings	helicopter
	SOM	sunset	sky	face	clouds	ocean	helicopter	ships	beach	birds	mountain	f-16	buildings	bay	city	boeing
sunset	PD	sunrise	clouds	sky	face	ocean	birds	bay	mountain	ships	helicopter	beach	boeing	buildings	f-16	city
	UTD	sunrise	face	ships	ocean	birds	city	sky	beach	f-16	mountain	clouds	bay	boeing	buildings	helicopter
	SOM	sunrise	sky	clouds	face	ocean	beach	helicopter	ships	mountain	birds	bay	buildings	f-16	city	boeing

[17] Z. Liu and Q. Huang, "A new distance measure for probability distribution function of mixture type," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, June 2000.

[18] J. Goldberger, S. Gordon, and H. Greenspan, "An effficient image similarity measure based on approximations of kl-divergence between two gaussian mixtures," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[19] S. Lulier and J. K. Uhlmann, "A general method for approximating non-linear transformations of probability distributions," Dept. of Engineering Science, University of Oxford, Tech. Rep., 1996.

[20] Yahoo! Image Search, "http://images.search.yahoo.com/."

Authorized licensed use limited to: Hong Kong Polytechnic University. Downloaded on June 24,2010 at 07:41:20 UTC from IEEE Xplore. Restrictions apply.