

MSRA-MM: Bridging Research and Industrial Societies for Multimedia Information Retrieval

Meng Wang, Linjun Yang, and Xian-Sheng Hua
 Microsoft Research Asia, Beijing, P.R. China
 {mengwang, linjuny, xshua}@microsoft.com

Abstract

This report introduces a dataset named Microsoft Research Asia Multimedia (MSRA-MM)¹ that aims to encourage research in multimedia information retrieval and the related areas. The images and videos in the dataset are collected from Microsoft Live Search and the performance of state-of-the-art industrial techniques can be evaluated accordingly. Therefore, conducting research using the dataset is able to demonstrate the practical usefulness of the studied algorithms. The data have been comprehensively annotated, such as their relevance, concept, category and quality, and different research scenarios can be implemented on the dataset.

Index Terms

Multimedia information retrieval, Benchmark.

I. INTRODUCTION

The last decade has witnessed great advances in image and video information retrieval. However, although many encouraging results have been reported in research society, there is actually a gap between the research and industrial societies on multimedia information retrieval. One case is example-based image retrieval that aims to retrieve images based on users' provided examples [7][11]. It keeps being a hot topic for years in research society, but nowadays most popular image search engines, such as Google/Yahoo/Live image search engines still adopt text-based image search. Another example is that in most of the research works on video search, clips such as shots or sub-shots are regarded as the basic unit (such as in TREC Video Retrieval Evaluation [1][10], which is the benchmark of video information retrieval). But actually most popular video search sites, such as Google/Yahoo/Live video search engines regard the whole video as search unit. How to evaluate the practical usefulness of the algorithms proposed in research society is a problem. In this report, we introduce a dataset named Microsoft Research Asia Multimedia (MSRA-MM) which aims to bridge research and industrial societies for multimedia information retrieval.

MSRA-MM actually consists of two sub-datasets, i.e., an image dataset and a video dataset that are collected from the image and video search engines of Microsoft Live Search, respectively. For image dataset, we have selected 68 representative queries based on the query log of Microsoft Live search and then collect about 1000 images for each query. There are 65443 images in all. For video dataset, we have selected 165 representative queries from query log and collect 10277 videos accordingly². The distribution of the video durations is illustrated in Figure 1. In Appendix A and B, we have illustrated the queries, their descriptions and the number of collected images/videos for each query in the image and video datasets, respectively.

II. WHAT DO WE PROVIDE?

Due to the copyright issue, we cannot share the original images and videos, and thus we resort to several alternative approaches. For image dataset, we provide a set of features, including: (1) 225D block-wise color moment; (2) 64D HSV color histogram; (3) 256D RGB color histogram; (4) 144D color correlogram; (5) 75D edge distribution histogram; (6) 128D wavelet texture; and (7) 7D face features. We also provide the thumbnails of the images (i.e., the small images displayed on Microsoft Live Search). For video dataset, we provide keyframes and motion-thumbnails. For each video, we implement shot detection and then extract a keyframe from each shot. The

¹This is Version 1.0. In the next versions we will further increase the size of the dataset.

²Video crawling is more difficult than image. The links of many videos are overdue, and several crawled videos cannot be successfully parsed. Thus the videos per query are much fewer in comparison with the image dataset.

keyframes can be regarded as a compact representation of the video content. We also provide a motion-thumbnail for each video. The motion-thumbnail is a short summarization of the original video, and it consists of a series of clips that are extracted from different shots. By viewing the motion thumbnails, users can get a general perception of the video content. The image and video data have been manually annotated in different ways, including:

(1) Relevance. For each image or video, its relevance to the corresponding query is labeled with three levels: very relevant, relevant and irrelevant. These three levels are indicated by scores 2, 1 and 0, respectively.

(2) Concepts. We have constructed a concept set that includes 50 concepts. From the original 65443 images, we randomly selected 10000 images for labeling. Then the selected images are manually labeled as “positive” or “negative” with respect to each concept. In Appendix C, we have illustrated the concepts, their descriptions and the corresponding numbers of positive samples.

(3) Category. For each video, it has been judged if it belongs to the following nine non-exclusive categories: (a) professional videos; (b) amateur videos; (c) handy videos; (d) fixed videos; (e) movie videos; (f) cartoon videos; (g) game videos; (h) MTV videos; and (i) Photo2Video (P2V) videos. So, each video has nine category flag values. If a flag value is 1, it indicates the video belongs to the corresponding category. The numbers of videos belonging to these categories are illustrated in Fig. 2.

(4) Quality. We also provide the quality information about the videos. For each video, its quality is manually labeled by three annotators with three levels: high quality (score 1), middle quality (score 0) and low quality (score -1). We average the three scores as the quality scores of the videos, and the distribution is illustrated in Fig. 3.

III. WHAT CAN WE DO ON THE DATASET?

A. Ranking and Reranking in Image and video Search

We have kept the original order of the crawled images and videos. Therefore, the performance of Microsoft Live Search can be directly measured, and this can be considered as the baseline which indicates the state-of-the-art performance of industrial society. Users can conduct research on ranking and reranking [6][12] techniques on the dataset and the works can directly be compared against industrial results.

B. Image Annotation

We can conduct annotation research on the image dataset. Different from most of the current popular image benchmark datasets that only support categorization, such as Caltech 101 [3] and Caltech 256 [4], our concept set is non-exclusive, i.e., the concepts may co-occur in an image, and thus we can investigate multi-label annotation [5][9] techniques on the dataset.

C. Video Classification

We can conduct study on video classification [2] based on the provided video category information. For example, we can classify the videos into professional videos and amateur videos and classify amateur videos into fixed videos and handy videos. We can also perform detections for certain categories, such as P2V video detection.

D. Video Quality Assessment

Web video quality assessment [8] is a topic that receives less attention, but actually it is useful in video search. For example, we can filter out low-quality videos in the search results or perform reranking to prioritize high-quality videos in ranking lists. Based on the provided quality information, we can conduct study on automatic video quality assessment.

IV. DATA STRUCTURE

Figure 4 shows the folder structure of all the data. Under the “MSRA-MM” folder, there are two folders: “ImageData” and “VideoData”. These two folders contain the image data and video data, respectively.

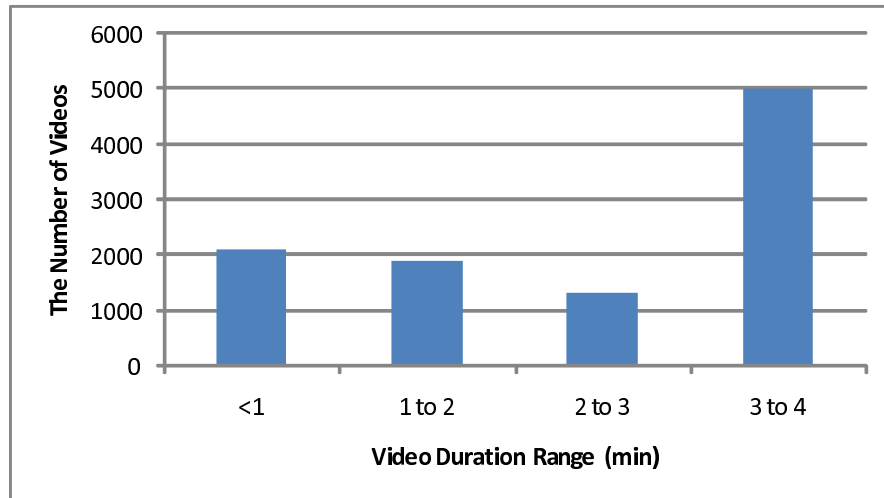


Fig. 1. The duration distribution (in minutes) of the videos.

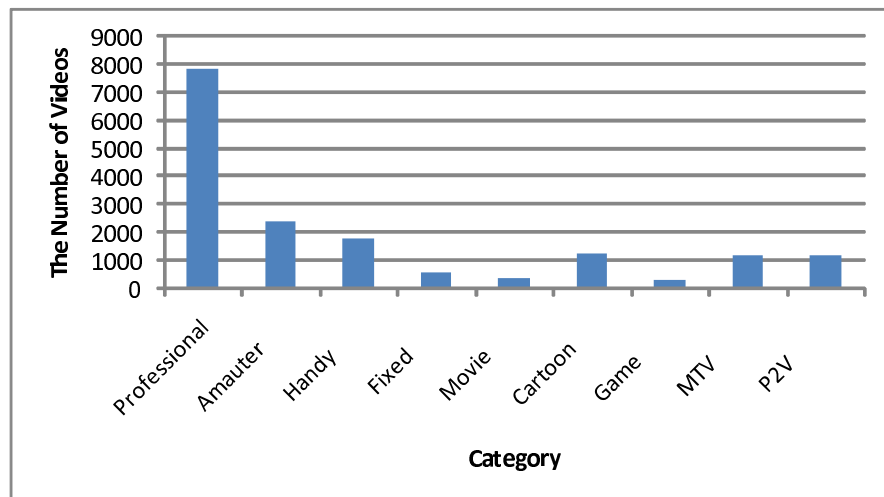


Fig. 2. The number of videos belonging to different categories.

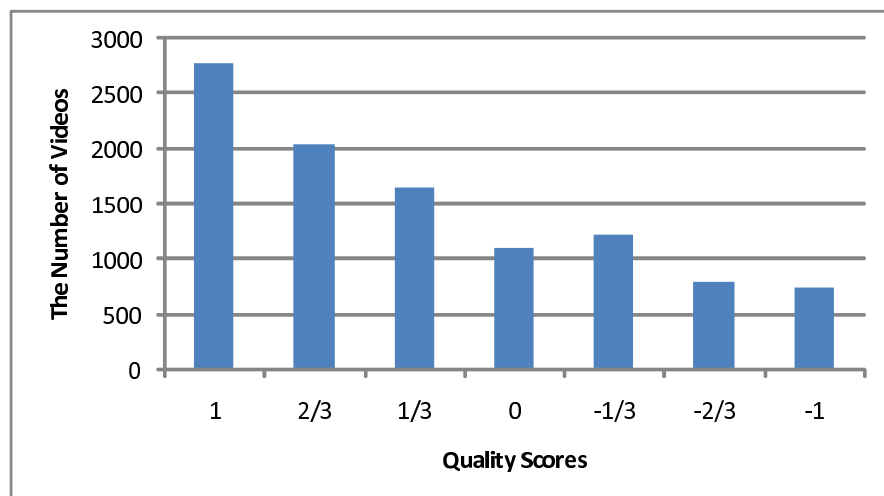


Fig. 3. The quality distribution of the videos.

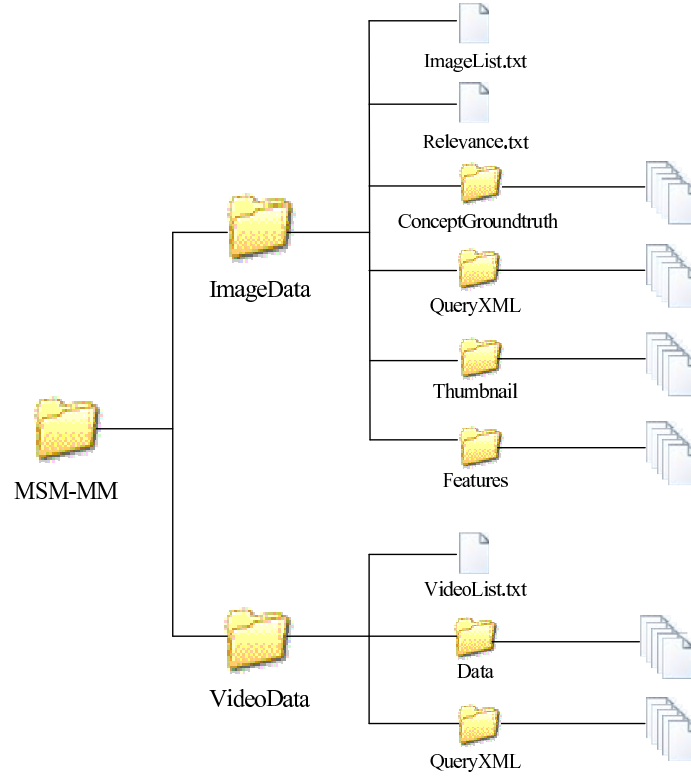


Fig. 4. Structure of the data contained in the MSRA-MM distribution

A. The structure of image data

In the “ImageData” folder, there are two text files and four subfolders: “ImageList.txt”, “Relevance.txt”, “ConceptGroundtruth” folder, “QueryXML” folder, “Thumbnail” folder and “Features” folder. The file “ImageList.txt” illustrates the names of all the 65443 images. The image file names have been defined following the style of “*query name + rank value*”. For example, “Angel_1.jpg” and “Angel_32.jpg” indicate the 1st and 32nd results in the image search ranking list of query “Angel”. The file “Relevance.txt” records the manually labeled relevance score of each image with respect to the corresponding query. As introduced in Section II, the score may be 2, 1 or 0, which indicate Very Relevant, Relevant and Irrelevant, respectively.

In the “ConceptGroundtruth” folder, there are 51 text files in all, including one file named “SelectedImageList.txt” that records the names of the selected 10000 images (note that we have only selected 10000 images from the whole 65443 images for labeling) and 50 files that illustrate the relevance of the images with respect to the 50 concepts (see Appendix C). The names of these files follow the style “*concept.txt*”, such as “beach.txt” and “bird.txt”. In each file, there are 10000 numbers that may be 1 or 0, which indicate the corresponding image is positive or negative with respect to the concept.

The “Thumbnail” folder contains the thumbnails of the original web images. The sizes of the thumbnails have been limited to be smaller than 160×160 (in pixel).

Under the “Features” folder, there are seven text files: “ColorMoment225.txt”, “RGB256.txt”, “HSV64.txt”, “Correlogram144.txt”, “EDH75.txt”, “WaveletTexture128.txt”, and “Face7.txt”. These seven files record the 225D block-wise color moment, 64D HSV color histogram, 256D RGB color histogram, 144D color correlogram, 75D edge distribution histogram, 128D wavelet texture and 7D face features, respectively.

In the “QueryXML” folder, we have collected 68 XML files that record the information about the web images of the corresponding 68 queries. As illustrated in Fig. 5, the XML files record the titles and addresses of the images, the formats and sizes of the thumbnails, etc.

```

<Properties>
  <Query>Animals</Query>
  <RawQuery>Animals</RawQuery>
</Properties>
<DocumentSet Start="0" Count="970" Retrievable="994">
  <Document Id="1842">
    <Title>Mister U @ AOSR</Title>
    <Url>http://www.misteru.com/categories/myStudents/</Url>
    <DisplayUrl>http://www.misteru.com/categories/myStudents/</DisplayUrl>
    <Thumbnail>
      <Url>http://ts1.images.live.com/images/thumbnail.aspx?q=2660293747700&id=b02bad311e406e48fe8bb278ba9d16c3</Url>
      <Format>image/jpeg</Format>
      <Width>160</Width>
      <Height>122</Height>
      <Filesize>4210</Filesize>
    </Thumbnail>
    <MediaUrl>http://www.misteru.com/images/animals.jpg</MediaUrl>
    <Width>426</Width>
    <Height>326</Height>
    <MediaFilesize>42647</MediaFilesize>
    <Id>Animals_1</Id>
  </Document>

```

Fig. 5. The format of the XML files describing the web images of different queries

```

<VideoSearchResult xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema" version="1.0">
  <Properties>
    <Copyright>Copyright © 2008 Microsoft. All rights reserved. These XML results may not be used, reproduced or transmitted in any form without the prior written permission of Microsoft Corporation.
    <Query>zoey 101</Query>
    <RawQuery>zoey 101</RawQuery>
    <Trace>0fb2d0bbafcb4d6a97d5dc88c76ffadc</Trace>
  </Properties>
  <DocumentSet Source="FEDERATOR_MONARCH" Start="0" Count="127" Total="2020">
    <Document Title="Zoey 101 - Zoey/Chase" Gping="/LS/GLinkPing.aspx?/_1_9SE/1?http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=TrOR0G_NQiw</Ui
    <Url>http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=TrOR0G_NQiw</Ui
    <DisplayUrl>http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=TrOR0G_NQiw</Ui
    <SourceTitle>YouTube</SourceTitle>
    <PlayLinkType>Embed</PlayLinkType>
    <Format>flv</Format>
    <Id>zoey 101_1</Id>
  </Document>
  <Document Title="Zoey 101" Gping="/LS/GLinkPing.aspx?/_1_9SE/1?http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=J0Uz4EJLHzs</Ui
    <Url>http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=J0Uz4EJLHzs</Ui
    <DisplayUrl>http://www.youtube.com/api2_rest?method=youtube.videos.get_details&dev_id=wtjx4nIgsx8&video_id=J0Uz4EJLHzs</Ui
    <SourceTitle>YouTube</SourceTitle>
    <PlayLinkType>Embed</PlayLinkType>
    <Format>flv</Format>
    <Id>zoey 101_2</Id>
  </Document>

```

Fig. 6. The format of the XML files describing the web videos of different queries

TABLE I
THE FILES IN THE SUBFOLDER OF EACH VIDEO

File Type	File List
Keyframes	A series of keyframes that are in jpeg format and named by their frame ID. For example, "14.jpeg", "92.jpeg", and "177.jpeg" indicate three keyframes that are extracted from the 14th, 92nd and 177th frames of the video.
Motion-thumbnail	A short video clip which has the same format of the original video, e.g., "zoey 101_80.flv".
Quality information	A text file named "Quality.txt". It records the manually labeled quality score of the video that may be 1, 0 or -1, which indicate high, middle and low quality, respectively.
Category information	Nine text files: "ProfessionalFlag.txt", "AmateurFlag.txt", "HandyFlag.txt", "FixedFlag.txt", "MovieFlag.txt", "CartoonFlag.txt", "GameFlag.txt", "MTVFlag.txt" and "P2VFlag.txt". As introduced in Section II, each file records a flag value. If the flag value is 1, it indicates the video belongs to the corresponding category.
Relevance information	A text file named "Relevance.txt". The file records a score that may be 2, 1 or 0, which indicate <i>Very Relevant</i> , <i>Relevant</i> and <i>Irrelevant</i> , respectively.
Structure information	An XML file, e.g., "zoey 101_80.xml". It illustrates the structure information of the video, such as the begin time, end time and keyframe of each shot.

B. The structure of video data

In the “VideoData” folder, there are a text file and two subfolders: “VideoList.txt”, “QueryXML” folder and “Data” folder. The file “VideoList.txt” records the ID of all the videos which also adopt the format “query name + rank value”. For example, “funny videos_1” and “funny videos_24” indicate the 1st and 24rd results in the video search ranking list of query “funny videos”.

Analogous to image data, the “QueryXML” folder contains 165 XML files that record the information about the web videos of the corresponding 165 queries (see Fig. 6), such as their titles, URL addresses, formats, etc.

In the “Data” folder, there are 10277 subfolders, i.e., a subfolder for each video. The subfolder is named by the ID of the video and it contains a lot of related files, such as the keyframes and motion-thumbnail of the video, its structure information and category information. Table 1 illustrates the file list in each subfolder.

REFERENCES

- [1] TRECVID: TREC Video Retrieval Evaluation, <http://www-nlpir.nist.gov/projects/trecvid>.
- [2] D. Brezeale and D. J. Cook, “Automatic video classification: a survey of the literature,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 38, no. 3, 2008.
- [3] L. Fei-Fei, R. Fergus, and P. Perona, “One-shot learning of object categories,” *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, 2006.
- [4] G. Griffin, A. Holub, and P. Perona, *Caltech-256 object category dataset*. California Institute of Technology Technical Report Number 7694, 2007.
- [5] W. Jiang, S. F. Chang, and A. Loui, “Active concept-based concept fusion with partial user labels,” in *Proceedings of International Conference on Image Processing*, 2006.
- [6] Y. Jing and S. Baluja, “Pagerank for product image search,” in *Proceedings of International World Wide Web Conference*, 2008.
- [7] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, “Content-based multimedia information retrieval: state of the art and challenges,” *ACM transactions on Multimedia Computing, Communications and Applications*, 2006.
- [8] T. Mei, X. S. Hua, C. Z. Zhu, H. Q. Zhou, and S. Li, “Home video visual quality assessment with spatiotemporal factors,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 6, 2007.
- [9] G. J. Qi, X. S. Hua, Y. Rui, J. Tang, T. Mei, and H. J. Zhang, “Correlative multi-label video annotation,” in *Proceedings of ACM Multimedia*, 2007.
- [10] A. F. Smeaton, P. Over, and W. Kraaij, “Evaluation campaigns and trecvid,” in *Proceedings of ACM Workshop on Multimedia Information Retrieval*, 2007.
- [11] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content based image retrieval at the end of the early years,” *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, 2000.
- [12] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X. S. Hua, “Bayesian video search reranking,” in *Proceedings of ACM Multimedia*, 2008.

APPENDIX A
IMAGE QUERIES, DESCRIPTIONS AND THE CORRESPONDING NUMBERS OF CRAWLED IMAGES

Query	Number	Description
party	958	A social gathering for entertainment and fun.
cat	929	A domesticated species of feline animal, commonly kept as a house pet.
panda	965	A black and white bear-like animal.
earth	961	Our planet, third out from the sun; (uncountable) Soil.
dog	976	An animal, member of the genus Canis.
snake	965	A legless reptile of the sub-order Serpentes with a long, thin body and a fork-shaped tongue.
cartoon	979	A humorous drawing, often with a caption, or a strip of such drawings.
background	966	A part of the picture that depicts scenery to the rear or behind the main subject.
Ronaldinho	959	A football star that comes from Brazil.
horse	948	A large hoofed mammal having a short-haired coat, a long mane, and a long tail, used for riding and for drawing or carrying loads.
woman	959	An adult female human being.
dragon	966	(Western) A gigantic beast, typically reptilian with leathery bat-like wings, lion-like claws, scaly skin and a serpent-like tail, often with fiery breath; (Eastern) A large snake-like lizard with the eyes of a hare, the horns of a stag, and the claws of a tiger.
spider	944	Any of various eight-legged, predatory arthropods, of the order Araneae, most of which spin webs to catch prey.
car	977	A wheeled vehicle that moves independently, steered by a driver mostly for personal transportation.
fish	955	A animal that lives in water, moving with the help of fins and breathing with gills.
boy	964	A young male.
ghost	953	The disembodied soul.
love	968	An intense feeling of affection and care towards another person.
youtube	967	A video sharing website where users can upload, view and share video clips.
bird	958	A member of the class of animals Aves in the phylum Chordata, having feathers and wings usually capable of flight, and laying eggs.
animal	978	A multicellular organism that is usually mobile, whose cells are not encased in a rigid cell wall and which derives energy solely from the consumption of other organisms.
flower	970	A reproductive structure in angiosperms, typically including sepals, petals, stamens, and ovaries.
angel	975	A divine and supernatural messenger from a deity, or other divine entity.
turtle	969	A land or marine reptile having a protective shell enclosing its body.
people	953	A group of two or more persons.
heart	968	A muscular organ that pumps blood through the body; A conventional shape or symbol used to represent the heart, love, or emotion.
frog	973	A small hopping amphibian.
chocolate	974	A food made from ground roasted cocoa beans.
cake	960	A rich, sweet dessert food, typically made of flour, sugar, eggs, and baked in an oven.
star	952	A luminous celestial body, made up of plasma and having a spherical shape; (geometry) A concave polygon with regular, pointy protrusions and indentations, generally with five or six points.
baby	951	A very young human being, from birth to a couple of years old.
beach	947	A horizontal strip of land, usually sandy, adjoining water.
wolf	965	A large wild canid, closely related to, and at times consanguineous to the domestic dog.
weather	964	The state of the atmosphere.
batman	981	Batman is a fictional comic book superhero co-created by artist Bob Kane and writer Bill Finger, appearing in publications by DC Comics.
email	969	Electronic communication among users of computer networks.
hairstyle	960	The style in which someone's hair has been cut and arranged.
tree	979	A large plant, typically over four meters in height, a single trunk which grows in girth with age and branches.
lion	971	A large cat, Panthera leo, native to Africa, India and formerly to much of Europe.
children	967	More than one young person.
Hawaii	959	A state of the United States of America.
food	962	Any substance that is or can be consumed by living organisms, especially by eating, in order to sustain life.
tiger	961	A large carnivorous animal of the cat family indigenous to Asia.
waterpark	955	An amusement park that features water play areas.
Indians	969	Plural form of Indian; Native Americans in the United States.
school	958	An institution dedicated to teaching and learning; an educational institution.
sport	965	Any athletic activity that uses physical skills, often competitive.
military	963	Having to do with armed forces such as the army, Marine Corps, navy and air force.

bee	966	A flying insect, of the order Hymenoptera, superfamily Apoidea.
medical	964	Of or pertaining to the practice of medicine.
plant	959	An organism that is not an animal, especially an organism capable of photosynthesis.
pig	974	Any of several mammals of the family Suidae, having short legs, cloven hoofs, bristly hair, and a cartilaginous snout.
cow	963	A female domesticated ox or other bovine, especially an adult after she has had a calf.
Disney	970	Well known for producing animated films and television series, performing stage musicals, and operating theme parks.
flag	962	A piece of cloth, often decorated with an emblem, used as a visual signal or symbol.
rose	969	A shrub of the genus Rosa, with red, pink, white or yellow flowers.
baseball	954	A sport in which the object is to strike a ball.
football	948	A game in which two teams of eleven players each contend to get a round ball into the other team's goal primarily by kicking the ball with their feet.
game	948	A pursuit or activity with rules performed either alone or with others, for the purpose of entertainment.
police	965	An organisation granted the legal authority to enforce the law.
fruit	962	The seed-bearing part of a plant, often edible.
nokia	960	Nokia is the world's leading mobile phone supplier and a leading supplier of mobile and fixed telecom networks including related customer services.
war	964	A conflict involving the organized use of arms and physical force between countries or other large-scale armed groups.
jesus	957	Jesus of Nazareth, a Jewish man whom Christians name Jesus Christ in belief that he is the Messiah.
golf	953	A ball game played by individuals competing against one another in which the object is to hit a ball into each of a series of holes in the minimum number of strokes.
map	961	A visual representation of an area, whether real or imaginary.
cowboy	957	A person who tends cattle; A person who dresses like a US cowboy, especially one who wears a cowboy hat.
hotel	952	An establishment that provides accommodation and other services for paying guests.

APPENDIX B
VIDEO QUERIES, DESCRIPTIONS AND THE CORRESPONDING NUMBERS OF CRAWLED VIDEOS

Query	Number	Description
you tube videos	165	An American video sharing website where users can upload, view and share video clips.
naruto	84	Naruto is an ongoing Japanese manga series written and illustrated by Masashi Kishimoto with an anime adaptation.
google	99	An American public corporation, and provides Internet search services.
wwe	3	A publicly traded, privately controlled integrated media and sports entertainment company dealing primarily in the professional wrestling industry.
video	120	Television show, movie
yahoo	53	An American public corporation, and provides Internet services worldwide.
leave britney alone video	156	A Britney's fan lashes out at gossip columnists such as Perez Hilton, and at reality TV star Simon Cowell, who criticized Britney Spears' onstage music performance at the 2007 MTV Video Music Awards in Las Vegas.
girls	6	A girl is any female human from birth through childhood and adolescence to attainment of adulthood. The term may also be used to mean a young woman.
paris hilton	54	An American celebutante, television personality, actress, singer and model.
music	111	An art form in which the medium is sound organized in time.
barney	169	A children's television show.
funny videos	96	A media franchise owned by the video game company Nintendo.
pokemon	129	A media franchise owned by the video game company Nintendo.
exploding whale oregon	2	A dead sperm whale was blown up by the Oregon Highway Division in an attempt to dispose of its rotting carcass.
britney spears	63	An American recording artist and entertainer.
zoey 101	93	Zoey 101 was an Emmy Award-nominated American comedy Television series.
soulja boy	140	An American rapper, dancer, and record producer.
family guy	106	Family Guy is an animated American television sitcom created by Seth MacFarlane.
chris brown	100	A Grammy-nominated American R&B and pop singer-songwriter, dancer, music video director and actor.
lil wayne	143	An American rapper.
music videos	110	A short film or video that accompanies a complete piece of music, most commonly a song with lyrics.
dog trick videos	141	Funny videos in which dog tricks are taught to a dogs for training and entertainment purposes.
akon	121	A Wolof Senegalese-American hip hop and R&B singer-songwriter, rapper, and record producer.
high school musical 2	151	The second film in the High School Musical film franchise.
barbie	151	Barbie is a best-selling fashion doll launched in 1959.
myspace	116	MySpace is a social networking website offering an interactive, user-submitted network of friends, personal profiles, blogs, groups, photos, music and videos for teenagers and adults internationally.
disney	98	Well known for producing animated films and television series, performing stage musicals, and operating theme parks.
hannah montana	128	An American television series.
tickling	124	The act of touching a part of the body, so as to cause involuntary laughter.
shakira	131	A Colombian singer-songwriter, musician, record producer, dancer and philanthropist.
anime	122	Animation in Japan and considered to be "Japanese animation" in the rest of the world.
50 cent	93	Curtis James ' stage name, is an American rapper.
alvin and the chipmunks	136	Alvin and the Chipmunks is a six-time Grammy Award-winning animated music group created by Ross Bagdasarian.
love	58	An intense feeling of affection and care towards another person.
fight	116	Purposeful violent conflict intended to establish dominance over the opposition.
ufo	92	A popular term for any aerial phenomenon that cannot immediately be identified.
movies	57	Film encompasses individual motion pictures.
beyonce	58	An American R&B singer-songwriter, record producer, and actress.
wwe divas	77	Diva is a term used by the professional wrestling promotion World Wrestling Entertainment (WWE) to refer to its female talent.
webcam	32	Internet users enjoyed viewing images from cameras set up by others elsewhere in the world.
bratz	149	A live-action feature film based on cartoon characters.
ghost	104	A disk cloning program.

lego	74	Kind of plastic toys have consisted of colorful interlocking plastic bricks and an accompanying array of gears, minifigures and various other parts.
msn	75	A collection of Internet services provided by Microsoft.
high school musical	119	An American television film, and the first in the High School Musical film franchise.
jennifer lopez	91	An American actress, singer, record producer, dancer, fashion designer and television producer.
giant octopus at seattle aquarium	1	A YouTube video about a visit to Seattle aquarium.
angelina jolie	108	An American film actor and a Goodwill Ambassador for the UN Refugee Agency.
bollywood	77	The largest film producer in India and one of the largest in the world.
teletubbies	102	A BBC BAFTA winning children's television series, primarily aimed at pre-school viewers.
babes	43	An American situation comedy.
stupid videos	58	An American website which presents funny and stupid videos from around the web and television.
eminem	90	Marshall Bruce Mathers III 's primary stage name.
john cena	18	An American actor, hip hop musician, and wrestler.
teen	46	An American teen lifestyle magazine for preteen and early teenage girls, ages 10 to 15.
dora	64	Animation series about a young child who goes on adventures with a talking monkey named Boots, backpack, map and many small insects.
benazir bhutto	135	A Pakistani politician who chaired the Pakistan People's Party (PPP).
noelia	131	A Puerto Rican pop singer who rose to fame in 1999.
rbd	41	RBD is a two times Latin Grammy nominated Mexican pop group.
inuyasha	7	A Japanese manga series written and illustrated by Rumiko Takahashi.
ronaldinho	115	A Brazilian footballer.
jennifer aniston	103	An American film and television actress.
ben 10	138	An American animated television.
carmen electra	41	An American glamour model, television personality, entertainer and sex symbol.
www	168	The World Wide Web is a system of interlinked hypertext documents accessed via the Internet.
jessica simpson	101	An American singer.
original star wars kid	61	An Internet phenomenon which started when a fourteen-year-old Quebecois high school student filmed himself wielding a golf ball retriever.
godzilla	161	A fictional Japanese monster from the Godzilla series of science fiction films.
women	64	The term woman (irregular plural: women) usually is used for an adult.
tila tequila	118	A Singapore-born Vietnamese American internet celebrity, reality TV performer, model, entertainer and singer.
dragon ball z	65	A Japanese manga series written and illustrated by Akira Toriyama.
michael jackson	26	An American recording artist, entertainer, and businessman.
undertaker	6	Someone involved in the business of funeral rites.
7 wonders world	139	A well known list of seven remarkable constructions of classical antiquity.
alicia keys	51	A well known list of seven remarkable constructions of classical antiquity.
titanic	101	A disaster film directed, written, co-produced and co-edited by James Cameron about the sinking of the RMS Titanic.
arab	13	An Arab is a person who identifies as such on genealogical, linguistic, or cultural grounds.
jeff dunham	111	An American ventriloquist and stand-up comedian.
bangbros	22	A network of adult websites.
You tube videos	127	A video sharing website where users can upload, view and share video clips.
big	15	Of a great size; large; the weakest sense of great size; Mr. Big was an American hard rock super group.
madonna	31	An American recording artist and entertainer.
vida guerra	39	A Cuban-born American model, actress, and singer.
cars	66	Cars is a 2006 animated feature film produced by Pixar and directed by both John Lasseter and Joe Ranft.
internet explorer	106	A series of graphical web browsers developed by Microsoft.
mario	58	A fictional character created by game designer Shigeru Miyamoto.
car	47	An automobile or motor car is a wheeled motor vehicle for transporting passengers, which also carries its own engine or motor.
snl	88	A weekly late-night 90-minute American sketch comedy/variety show based in New York City.
asian	19	The world's largest and most populous continent.
amy fisher	36	An American woman writer.
50cent	126	An American recording artist and actress.

mariah carey	135	An American recording artist and actress.
wrestling	153	The act of physical engagement between two people in which each wrestler strives to get an advantage over, or control of, the opponent.
rihanna	154	A Barbadian singer, model and fashion designer.
jessica alba	1	An American television and film actress.
virgin mary	145	A Jewish resident and known from the New Testament as the mother of Jesus of Nazareth; Virgin Mary is the name of a popular non-alcoholic cocktail.
you tube scandal video	186	Videos regarding to scandal in the website YouTube.
sesame street	148	An American educational children's television series.
aventura	129	A Bachata music group based in The Bronx.
hidden cam	12	A still or video camera used to film people without their knowledge.
plies	128	Algernod Lanier Washington' stage name Plies, is an American rapper.
movie trailers	72	Film advertisements for films that will be exhibited in the future at a cinema, on whose screen they are shown.
dance	41	An art form that generally refers to movement of the body, usually rhythmic and to music, used as a form of expression, social interaction or presented in a spiritual or performance setting.
images	67	An artifact, usually two-dimensional, that has a similar appearance to some subject-usually a physical object or a person.
avril lavign	80	A Canadian pop rock/pop punk singer, songwriter, fashion designer, and actress.
south park	67	An animated American television comedy series created and written by Trey Parker and Matt Stone.
ufc	98	Ultimate Fighting Championship (UFC) is a U.S.-based mixed martial arts (MMA) organization, currently recognized as the largest MMA promotion in the world.
film	14	Film encompasses individual motion pictures, the field of film as an art form, and the motion picture industry.
cricket	121	A bat-and-ball team sport that originated in England.
bleach	1	A chemical that removes color or whitens, often via oxidation; A Japanese manga series written and illustrated by Tite Kubo.
spongebob	106	The main character of the fictional Nickelodeon animated series SpongeBob SquarePants.
free	3	Free is the absence of interference with the sovereignty of an individual by the use of coercion or aggression; or is the process of providing goods or services without monetary compensation.
pamela anderson	3	A Canadian-born actress, sex symbol, glamour model, producer, author, and celebrity.
mary apparitions	3	A Jewish resident of Nazareth in Galilee and known from the New Testament as the mother of Jesus of Nazareth.
pregnant	7	The carrying of one or more offspring, known as a fetus or embryo, inside the uterus of a female.
latina	2	Latina is the capital of the province of Latina; Latina, a magazine for Hispanic women; Latina is the most common type northern Indian Sikhara
cartoons	2	Creative visual work for print media, for electronic media, and even animated films and animated digital media.
salma hayek	11	A Mexican and American actress, director, television and film producer.
street fights	6	Denote unsanctioned, usually illegal, hand-to-hand fighting in public places, between individuals or groups of people.
harry potter	6	A series of seven fantasy novels written by British author J. K. Rowling.
all	11	Every individual or anything of the given class, with no exceptions.
wwf	2	An international non-governmental organization for the conservation, research and restoration of the environment.
news	25	Reports of current events broadcast via media such as newspapers or television; A Japanese club consists of boys, which belongs to Johnny firm.
kimbo	4	A Bahamian American mixed martial arts fighter.
free videos	4	Free video refers to video content that is free to use for any purpose, or licensed under a free and open license to such an effect, at least for distribution, and at most for modification and commercial usage.
beach	14	A geological landform along the shoreline of a body of water.
oops	4	An interjection on realizing that the speaker has made a mistake or slipped.
hidden camera	9	A still or video camera used to film people without their knowledge.
musica	1	A famous music website; musica means "music" in Latin language.
shakeela	10	A popular South Indian actress from India who has appeared in Indian softcore porn movies.
facial	3	A facial is a cosmetic treatment of the face, commonly involving a variety of skin treatments.

nicole kidman	6	An Academy Award-winning actress, a UN Citizen of the World award winning humanitarian and a UNIFEM and UNICEF Goodwill Ambassador.
cats	27	A small predatory carnivorous species of crepuscular mammal. The Cats is also an award-winning musical composed by Andrew Lloyd Webber.
boondocks	5	An American slang term for “the country” or any rural/wilderness location in general that is isolated. “Boondocks” is a song released by American country music group Little Big Town. The Boondocks is a Peabody Award winning American animated series created by Aaron.
anna nicole smith	1	An American model and television personality.
paris	3	The capital of France and the country’s largest city.
catfight	3	Catfight is a term for an altercation between two women; Catfight is a compilation of unreleased songs by the British indie rock band Hefner.
college girls	1	A Channel 4 documentary series.
boxing	11	A combat sport in which two participants, generally of similar weight, fight each other with their fists.
mixed wrestling	4	A type of professional wrestling match between a man and a woman, and may also refer to tag team matches with both men and women on each team.
nice	9	Nice is a city in southern France located on the Mediterranean coast, between Marseille, France, and Genoa, Italy.
metacafe	105	A community based video sharing web site, that specializes in short-form original entertainment, where users upload, view and share video clips.
chipmunks	1	The common name for any small squirrel-like rodent species of the genus <i>Tamias</i> in the family <i>Sciuridae</i> .
spm	4	A branch of microscopy that forms images of surfaces using a physical probe that scans the specimen. A Mexican-American rapper. Statistical parametric mapping or SPM is a statistical technique for examining differences in brain activity recorded during functional neuroimaging experiments.
mujra	2	A form/adaptation of Kathak dance originated by north Indian Tawaif during the Mughal era.
windows vista	2	A line of operating systems developed by Microsoft for use on personal computers.
halle berry	6	An American actress, former fashion model, and beauty queen.
birth	4	An American actor and singer.
kane	1	A Spanish-born American professional wrestler and actor.
christmas	13	An annual holiday celebrated on December 25 that marks and honors the birth of Jesus of Nazareth.
fight	41	Purposeful violent conflict intended to establish dominance over the opposition. Fight Club is a 1999 American feature film adaptation of the 1996 novel of the same name by Chuck Palahniuk.
buy book online	25	To buy books by means of internet service
sharon stone	12	An American Golden Globe- and Emmy-winning actress, producer, and former fashion model.
howard stern	6	An American radio host and television personality.
robot chicken	28	An Emmy-Award Winning American stop motion animated television series created by Seth Green and Matthew Senreich.
games	12	A video game is a game that involves interaction with a user interface to generate visual feedback on a video device.
cristiano ronaldo	2	A Portuguese footballer who plays as a Winger for English Premier League club Manchester United and the Portuguese national team.
wet	1	A famous American Lubricant; Wet Wet Wet are a Scottish pop band that formed in the 1980s.
football	7	Football is the word given to a number of similar team sports, all of which involve (to varying degrees) kicking a ball with the foot in an attempt to score a goal.
dogs	9	A domesticated subspecies of the gray wolf, a mammal of the <i>Canidae</i> family of the order <i>Carnivora</i> .
songs	6	A song is a musical composition.
pink	2	A two-time Grammy-winning American singer-songwriter.
japanese	8	An island country in East Asia; Japanese is a language.
transformers	1	Fictional alien robots and the title characters of a popular Hasbro toy line and its spin-offs.

APPENDIX C
IMAGE LABELS, DESCRIPTIONS AND THE CORRESPONDING NUMBERS OF POSITIVE SAMPLES

Label	Number of Positive Samples	Description
indoor	4185	Situated in, or designed to be used in, or carried on within the interior of a building.
outdoor	3499	Situated in, designed to be used in, or carried on in the open air.
woman	1206	An adult female human being.
man	1864	An adult male human.
people	3067	Plural form of person. A body of human beings considered generally or collectively; a group of two or more persons.
building	562	A closed structure with walls and a roof.
sky	1033	The atmosphere above a given point, especially as visible from the ground during the day.
water	622	A clear liquid potable by humans and animals.
animal	2397	A multicellular organism that is usually mobile, whose cells are not encased in a rigid cell wall and which derives energy solely from the consumption of other organisms.
flower	490	A reproductive structure in angiosperms, typically including sepals, petals, stamens, and ovaries.
street	210	A paved part of road, usually in a village or a town.
mountain	207	A large mass of earth and rock, rising above the common level of the earth or adjacent land.
cityscape	617	The view of the buildings of a city, usually referring to a pictured landscape.
landscape	550	A portion of land or territory which the eye can comprehend in a single view, including all the objects it contains.
car	386	A wheeled vehicle that moves independently, steered by a driver mostly for personal transportation.
flag	168	A piece of cloth, often decorated with an emblem, used as a visual signal or symbol.
vegetation	1728	Plants collectively.
map	193	A visual representation of an area, whether real or imaginary.
crowd	1120	A number of things collected or closely pressed together; also, a number of things adjacent to each other.
beach	228	A horizontal strip of land, usually sandy, adjoining water.
cloud	491	A visible mass of water droplets suspended in the air.
baseball	107	A sport in which the object is to strike a ball so that one of a nine-person team can run counter-clockwise among four bases, causing a run. The team with the most runs after nine or more innings wins.
baseball	107	A sport in which the object is to strike a ball so that one of a nine-person team can run counter-clockwise among four bases, causing a run. The team with the most runs after nine or more innings wins.
basketball	10	A sport in which two opposing teams of five players strive to put a ball through a hoop.
boat	94	A craft used for transportation of goods, fishing, racing, recreational cruising, or military use on or in the water, propelled by oars or outboard motor or inboard motor or by wind.
cartoon	1097	A humorous drawing, often with a caption, or a strip of such drawings.
dog	229	An animal, member of the genus <i>Canis</i> that has been domesticated by man for thousands of years.
cat	106	A domesticated species of feline animal, commonly kept as a house pet.
food	356	Any substance that is or can be consumed by living organisms, especially by eating, in order to sustain life.
weather	487	The state of the atmosphere, mainly with respect to its effects upon life and human activities.
horse	190	A large hooved mammal having a short-haired coat, a long mane, and a long tail, used for riding and for drawing or carrying loads.
athlete	515	(US) A person who actively participates in physical sports, possibly highly skilled in sports.
snow	445	The frozen, crystalline state of water that falls as precipitation.
swimming	259	The activity of being moving oneself through water using one's arms and legs while buoyed up by the water.
earth	117	Our planet, third out from the sun; Soil.
smoke	70	The visible vapor, gases, and fine particles given off by burning or smoldering material.
baby	195	A very young human being, from birth to a couple of years old.
computer	199	A programmable device that performs mathematical calculations and logical operations, especially one that can process, store and retrieve large amounts of data very quickly.
nighttime	599	The hours of darkness between sunset and sunrise.
ocean	581	One of the five large bodies of water separating the continents.
room	714	A separate part of a building, enclosed by walls, a floor and a ceiling.

sun	72	The star which the Earth revolves around and receives light and warmth from.
heart	137	A muscular organ that pumps blood through the body; A conventional shape or symbol used to represent the heart, love, or emotion.
bird	132	A member of the class of animals Aves in the phylum Chordata, having feathers and wings usually capable of flight, and laying eggs.
road	300	A narrow strip of land made suitable for travel between places.
phone	193	A device for transmitting sound in real time across distances.
fire	53	Fire is the heat and light energy released during a chemical reaction.
studio	78	An artist's or photographer's workshop or the room in which an artist works.
police	31	An organisation granted the legal authority to enforce the law.
apple	11	A common round fruit produced by the tree <i>Malus domestica</i> , cultivated in temperate climates.