Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

Image Retrieval using BOVW and Relevant Feedback

Ismail El Sayad
Department of Computer and Communication Engineering
Lebanese International University
Beirut, Lebanon

Samih Abdul-Nabi
Department of Computer and Communication Engineering
Lebanese International University
Beirut, Lebanon

Zahraa Loubany
Department of Computer and Communication Engineering
Lebanese International University
Beirut, Lebanon

Hussein Kassem
Department of electrical and electronic Engineering (EENG)
Lebanese International University
Beirut, Lebanon

Dina Balchy
Department of Computer and Communication Engineering
Lebanese International University
Beirut, Lebanon

Abstract **We introduce a new methodology in Image Retrieval using Relevance Feedback. The image will be represented in a higher level depending on the users' feedback as they evaluate the veracity and accuracy of the retrieved images. The visual data (hereby known as Visual Descriptors) are extracted using TOP-SURF, whilst the textual data (known as Textual Descriptors) are extracted based on the TF-IDF values of the annotated tags of the images. The usage of the advanced process BOVW, Bag of Visual Words, is inspired by the success of Bag of Words in text classification and retrieval. The BoVW represents the image by all words that describes it or can be generated from. The empirical distribution of words is captured with a histogram making the similarities distinguishing much easier. The feedback is represented by a ranked relevancy assignment performed by the user. The system by its turn will consider the images with a high relevancy value to enhance the image presentation according to a weighted schema using the concept of BoVW.**

Keywords: BOVW, Relevance Feedback, CBIR, SURF

I. INTRODUCTION

Through the lapse of era, the amount of digital image collections has grown dramatically due to the rapid increase of online users and web applications. This considerable increase was a result of the technological breakthroughs we witnessed. Admittedly, images are the easiest way of communication used to convey information and reach the audience brains smoothly, which made the demand on using images grows significantly.

With the popularity of social media applications, it's not a surprise that images are becoming increasingly important for content sharing and viewing. Generally, audience and readers like to visualize stories not just to read them. Images illustrate the ideas for the readers so the overall experience is more tangible and less demanding on the reader's attention with a clear visual content. Visual content is content that engages and inspires. This makes the opportunity to spark the readers' interest less daunting than comprehensive texts.

The importance of web applications especially social media is undeniable. People are spending most of their time navigating browsers, checking news and keeping an eye on their friends' new feeds which are most of the time pictures shared publically. Moreover, the influence of television, old photographs and games has contributed to this growth as well. In this context, the development of appropriate systems to manage effectively these huge image collections is a necessity.

The efficient systems used for managing these collections use the so-called CBIR – Content Based Image Retrieval Systems [1]. Basically, CBIR is a common method for image retrieval which has been created to solve the main problems of the query by text which is commonly known as TBIR – Text Based Image Retrieval [2].

In TBIR [2] systems, images are manually annotated by text descriptors, which are used by a database management system to perform image retrieval. The user provides query in terms of keyword and the system by its turn will retrieve all images that are similar to the user's query or have the same textual annotation.

The TBIR system is illustrated in the Figure below:

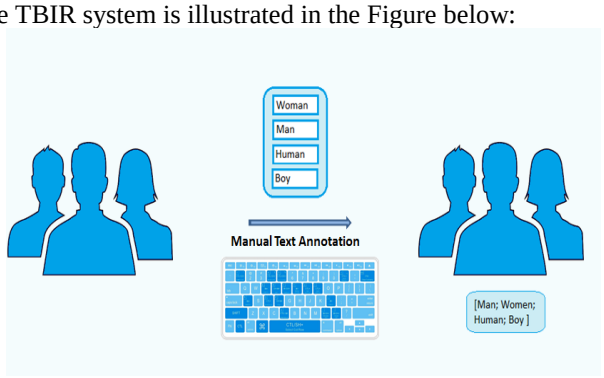

Figure 1: Text-Based Image Retrieval System

Any image found within the system database is manually annotated by textual phrases. Although this technique is known to be computationally fast in image retrieval, it has some difficulties. Firstly, the manual annotation of a large number of images requires a considerable level of human labor. Secondly, the labeled images may hold unexpressed feelings and emotions that cannot be described by text descriptors. Thirdly, the manual annotation of images may hold a significant level of inaccuracy due to subjectivity of human perception.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

The CBIR [3,4,5] was introduced to solve such problems by taking the input as a query image not as a text which in turn searches for images similar to the query image by color, texture or form. CBIR, also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases.
Figure 2 shows w typical schema of CBIR system:

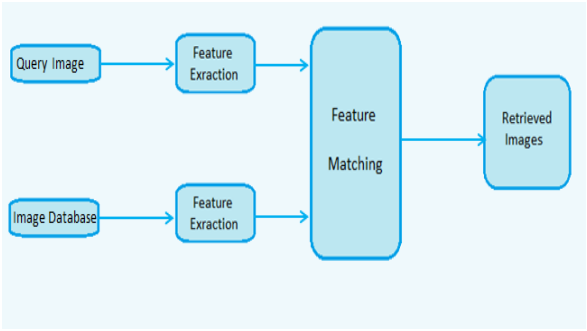

Figure 2: Content Based Image Retrieval System

The images found in the database are said to be training images. Each training image is represented as a vector of features which are also known as code words. Similarly, any query image will also be represented in the same way so that common features (similarities) are easily distinguished. The similarities detection are performed by measuring the Euclidean distance between the feature vector of query image and those of the training set images found in the database.

A great advantage can be taken from the Relevance Feedback (RF) that can be implemented within the CBIR technique. RF forms a great interaction between the system and the users after offering them the results of their search. This feedback could be an image reordering of the yielded results, word description or a rating mark for the image according to its relevancy to the query image which will actually be our main concern.

Our framework could be generally summarized by offering a higher level image representation using the users' relevance feedback which improves the system's performance by the images driven from the database to the user.

II. PROPOSED MODEL

Our proposed scenario is based on three main process. Three processes will be mainly introduced; two by which they are performed by the system, and one process by the user. As shown in Figure 3 Firstly, the system will perform the retrieval process, is followed by feedback process done by the user. Beyond the feedback submission, system will update the order of the retrieved images and rearrange them according to the users' accumulative feedbacks.

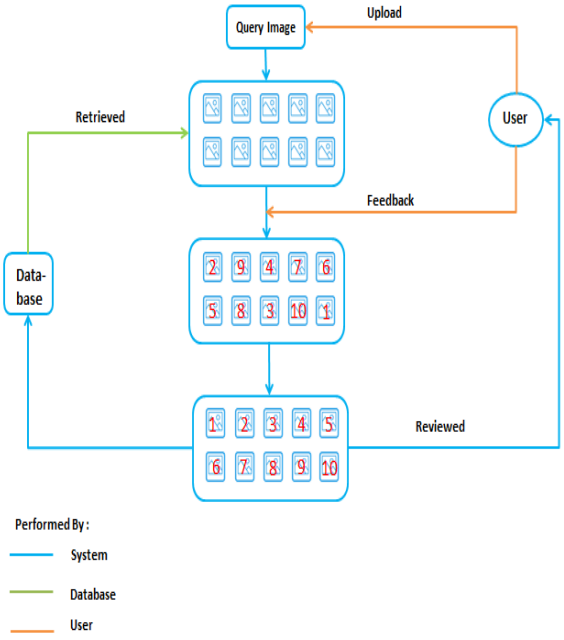

Figure 3: Architecture of the RF Proposed System

A. *Image Retireval Using BOVW*

Generally, there are four sequential steps for performing a Bag of Visual Words.

A.1 Extracting features From Training Images.

In this step, the features are extracted from the image as an interesting point. For instance, the features found in the image of the figure above might be eyes, window, mouth, or hands each as a local patch. These features are then presented as numerical vectors called "features descriptors". One of the most famous descriptors is SURF descriptor which presents each patch as 64-dimensional vector. Note that there are many key point description techniques such as Harris and SIFT [6]. SURF algorithm is similar to SIFT, but it is more simplified and computationally faster.

Figure 1: Feature Extraction Illustration

Figure 4 above shows a clarified illustration of how the system extracts the features where the interesting points are detected. Each training image has its own features that will be later on be clustered altogether to form different group of similar features.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

A.2 Clustering like-features together

This step is defined as a process of organizing objects into distinct groups with similar members. A cluster therefore is a collection of common features (patches) or features with high similarities between them. Here, patches are converted to "code words" that refers to words in text documents, producing a "code book" or a vocabulary that also refers to a dictionary of words. K-means clustering [7] could be an effective algorithm for image clustering which works as follows:

1. Select initial cluster centroids "c" at random.
2. Compute the distance between each patch and the centroids of the clusters.
3. Assign each patch to the cluster with the nearest centroid (minimum distance).
4. Recalculate each centroid as the mean of the objects assigned to it.
5. Repeat previous 2 steps until no change.

Figure 2: Feature Clustering

AS shown in Figure 5 After assigning the similar feature to the same cluster, a codebook containing all visual words can be formed. All similar features are gathered within the same code word forming indexed visual words. The figure above represents a set of different clusters (groups of features) after being clustered together.

A.3 . Representing the image as a set of weighted Visual Words

At this stage, images are no longer represented as a set of pixels. Instead, it can be represented with a higher level that is more oriented to the semantic as a set of patches or visual words known as "Bag of Visual Words". Each image can be exemplified as a vector containing all visual words found in the dictionary as vector components. Each component has its own "tf-idf" value which is used as a weighting factor as shown in the figure 6.

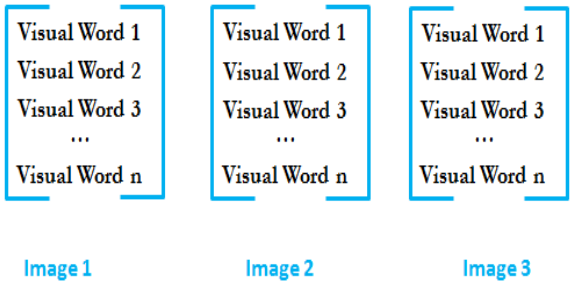

Figure 3: Image Representation as Weighted Vector

The "tf" represents the number of occurrences of a visual word in the image divided by the total number of visual words in this image which is called (Term Frequency). The other factor "idf" refers to the total number of images divided by the number of images where the visual word appears and it is termed as (Inverse Document Frequency). Accordingly, the (tf-idf) weighting factor of each component in the vector is the product of the previous two static values.

A.4 Constructing histograms of frequency offeatures

Beyond the vector presentation of the image, we can easily visualize a histogram showing how many features the image has in each cluster. In other words, a histogram illustrating the frequency of each visual word contained in this image. Note that these histograms show the frequency of occurrences and not the position.

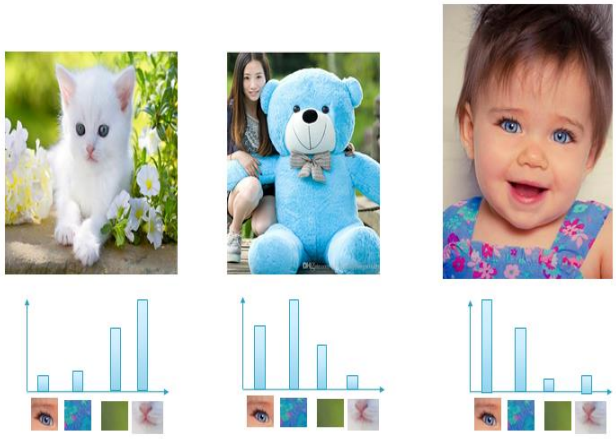

Figure 4: Constructing Histograms of Frequency Features

Figure 7 shows a set of histograms representing the frequency of features of different images. After we extract features from the tested image, and form its corresponding histogram of visual word frequencies to be compared with the ones obtained previously. Through this approach, matches can be effectively computed. A positive or true match considered to be within the same category or having a high correlation with. Figure 8 illustrates how this comparison takes place.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

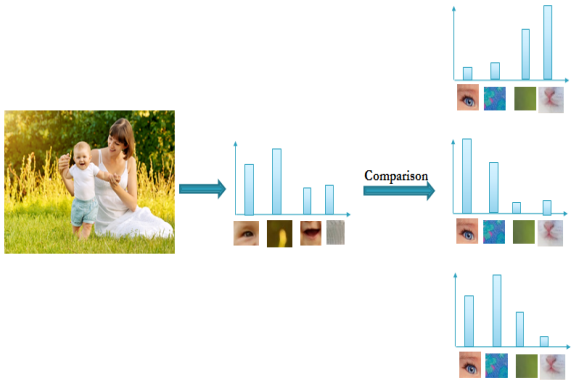

Figure 5: Evaluating Images against Obtained Histograms

B. Relevance Feedback for the retrieved Results

Lately, all recent retrieval systems embedded within their systems the user's relevance feedback to further improve the retrieval process and produce more meaningful and related retrieved images [8,9,10]. We consider the most positive image selection on each feedback iteration. Through continuous learning and interaction with end-users, as shown below

1. The system provides initial set of retrieved images through query image.
2. User evaluates the above results as to whether they are relevant (positive examples) or irrelevant (negative examples) to the query.
3. Machine learning algorithm is applied to learn the user's feedback. Then go back to (2).

Steps (2) – (3) are repeated till the user is satisfied with the results.

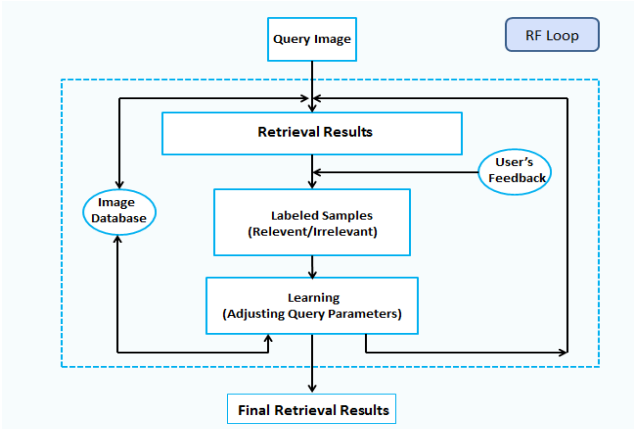

Figure 6: Relevance Feedback Algorithm

Figure 9 shows a general overview of the RF algorithm. The process performed in step (3) is to be performed by the system. That is, the burden of specifying the weight is removed from the user and held by the system instead.

C. Merging RF with BOVW

The system will rearrange the set of retrieved images according to the RF performed by various users. The feedbacks of all users are taken into account every single iteration, so that the user is able to visualize the original order of retrieved images by the system and also the order of the images after his own feedback.

III. EXPERIMETAL RESULTS

To examine the execution of our method, we used the Histogram of Magnitudes of the optical flow. We used a video recorded in our university to observe the condition. Videos are utilized to show meaningful data to the security

A. Experimental Setup

The experiments were conducted using our personal computer with Intel Core i7 using visual studio 2010 and OpenCv 3.3.0 software. After getting the video, we partition the frame into 4x4 non overlapping zones. For each zone, optical flow vector is estimated.

The design consideration of our retrieval software was the programming language that suits best our simulation. While various options such as C++ and JAVA were employed; our chosen one was C# due to the availability of libraries for image processing and MicrosoftSQL server for database due voluble dealing with large databases.

Using TOP-SURF [11] as an open source code was our starting point. TOP-SURF is an image descriptor that integrates interest points with visual words, and thus enhancing its performance. TOP-SURF offers the elasticity in descriptor size variation and supports very efficient image corresponding. Besides the visual word extraction, visualization, and comparisons it also provides a high level API and very large pre-computed codebooks [6]. TOP-SURF descriptor is a fully open source, although it depends on libraries that need different licenses. As the original SURF [12] descriptor is a closed source, we used the OpenSURF as an alternative open source, which depends on OpenCV that is released under the BSD license.

For a better database practicing, our work is connected to MicrosoftSQL server 2014 containing all tables and procedures needed for retrieval process execution and RF engagement as shown in Figure 10. Note that in the SQL software, we have used one table and 3 procedures (Retrieve, update and delete) as shown below.

Column Name	Data Type	Allow Nulls
RATING_ID	int	☐
RATE_ORIGINAL	decimal(18, 10)	☑
RATE_UPDATED	float	☑
SOURCE_IMAGE	varbinary(MAX)	☑
COMPARE_IMAGE	varbinary(MAX)	☑

Figure 10: SQL Database Table

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

B. Implementation/SimulationResults

Same interpretation as zone A, for the different zones shown

The retrieval demo of the query image with the set of images compared to it with the cosine difference is illustrated in the figure below.

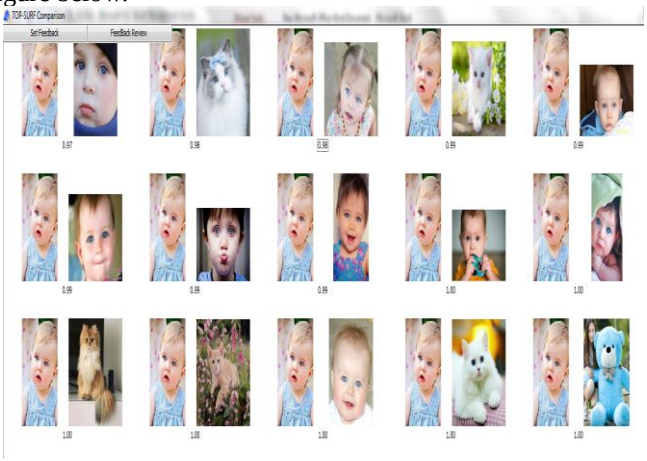

Figure 11: Retrieved Images using Cosine Difference

As shown in the above figure, the user has two different options; either he chooses to submit his feedback by rearranging the set of retrieved images or he directly view the updated rearranged form of images after engaging all previous feedbacks.

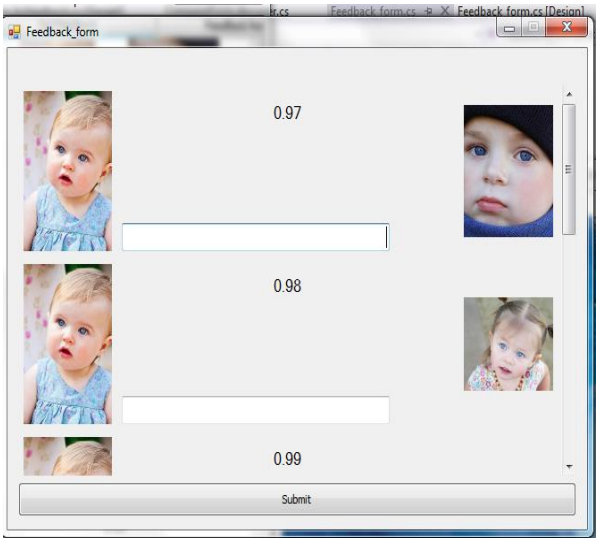

Figure 12: Users' Feedback Insertion

The above figure shows how the user is allowed to submit his feedback regarding the first top ten retrieved images. Here, the user gives a rating mark for each image ranging from 1 up to 10. The mark represents the degree of relevancy between the image and the retrieved one. Number 1 presents the image that is most relevant and closely connected to the query image, from the users own perception. However, number 10 presents the lowest degree of relevancy. Once the user submits his feedback, the system -which is automatically connected to the SQL- will update the table of images and insert the ranking mark corresponding to each image.

The user is allowed to visualise the updated order of the retrieved images related to the same query image. In other words, the system collects all feedbacks associated to the tested query image and averages them up to give the new order of each retrieved images. Notably, the images are arranged in the ascending order as the lowest number presents the most relevant. Remarkably, the user is capable of visualising the new order of the retrieved images even if hadn't submit his own feedback. This means that the step of inserting the feedback is optional, and the user is not obliged to do it if he is not interested to do so. Figure 13 shows a demo of the set of retrieved images in the new rearranged order.

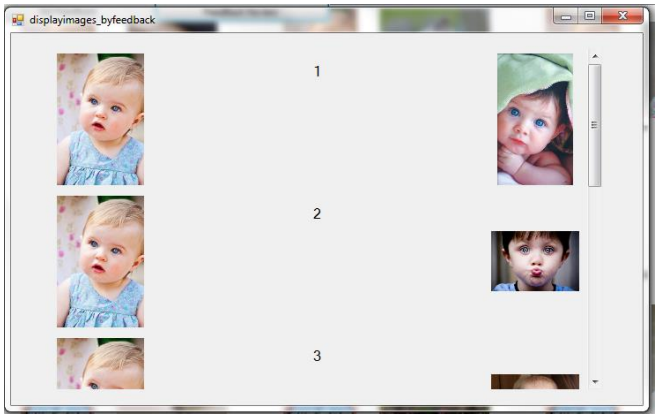

Figure 13: Images reviewing using RF

The above figure obviously shows how the images are rearranged taking into account the user's feedback.

Precision performance measure is used to calculate the performance of the retrieval system. Precision is a relation presenting the number of relevant images retrieved to the total number of retrieved images (relevant and irrelevant). This is shown in the equation below.

After each iteration, the system averages the total entries of feedbacks and returns the order of the image according to the equation below.

$$Average\ Feedback = \frac{\sum Feedback\ Entries}{Number\ of\ feedback\ submitted}$$

Equation 3 – Average Feedback

$$Precision = \frac{Number\ of\ relevant\ images\ retrieved}{Number\ of\ images\ retrieved} X\ 100$$

Equation 4 – Precision Percentage

The precision is computed for each category separately in order to evaluate the overall performance of our system before and after consecutive rounds of relevance feedback. The results are presented in the table below. The results percentage is computed taking into account the retrieved top 10 images of the set under study.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 07, July-2018

Precision (%) / Category	Before RF	After RF Round 1	After RF Round 2	After RF Round 3
Babies	50	60	70	80
[illegible]	60	70	80	100
Flowers	50	60	80	90
Musical Instruments	40	60	60	80

Table 1: Comparison of Precision Percentage

Regarding the first category (Babies), the precision of 50% before relevance means that originally the system retrieves 5 relevant images out of the top 10 retrieved images. After the first round of the feedback submission the precision percentage increases to 60% which means that 6 images out of the top 10 are now relevant to the query image. After each round,

IV. CONCLUSION

As huge image databases become a vast necessity in scientific, medical and in advertising/marketing domains, approaches for organizing a database of images and for effective retrieval have become very crucial. From here, the CBIR systems gave birth. It is the field of representing, organizing and searching images based on their visual content rather than textual tags describing it. Retrieval of images is no longer based on textual phrases and annotations but on features extracted directly from the image data. This approach retrieves digital images from large databases using the content of the images themselves without human intervention, therefore eliminating inefficient and subjective manual labeling.

The implementation of Relevance Feedback in CBIR systems proved that by testing the user's satisfaction and engaging it in the system, the retrieval process is much more efficient and precise. That is, the result of the retrieved images met the users' desires efficiently after engaging the feedbacks of several users within the retrieval process.

As for future work, we aim to increase the performance of CBIR systems by enhancing the way we search by. Besides the Relevance feedback we've added, searching would be better if we combine the two approaches (CBIR and TBIR) together instead of searching in terms of an image solely. In other words, we join textual and visual features using a certain algorithm, which fuses Visual Descriptors and Textual Descriptors to produce a multimodal global feature that helps in the retrieval process. In addition to that, we will use the content based image retrieval to create an upper level image presentation, for instance; a visual phrase. Furthermore, this model can be used to retrieve frames from videos using a query image.

REFERENCES

[1] Long, Fuhui, Hongjiang Zhang, and David Dagan Feng. "Fundamentals of content-based image retrieval." Multimedia Information Retrieval and Management. Springer, Berlin, Heidelberg, 2003. 1-26.
[2] Wilkins, Peter, et al. "Text based approaches for content-based image retrieval on large image collections." (2005): 281-288.
[3] Jain, Neha, Sumit Sharma, and Ravi Mohan Sairam. "Content Base Image Retrieval using Combination of Color, Shape and Texture Features." International Journal of Advanced Computer Research 3.1 (2013): 70-77.
[4] Liu, Ying, et al. "A survey of content-based image retrieval with high-level semantics." Pattern recognition 40.1 (2007): 262-282.
[5] Gudivada, Venkat N., and Vijay V. Raghavan. "Content based image retrieval systems." Computer 28.9 (1995): 18-22.
[6] Scovanner, Paul, Saad Ali, and Mubarak Shah. "A 3-dimensional sift descriptor and its application to action recognition." Proceedings of the 15th ACM international conference on Multimedia. ACM, 2007.
[7] Tatiraju, Suman, and Avi Mehta. "Image Segmentation using k-means clustering, EM and Normalized Cuts." University Of California Irvine (2008).
[8] Rui, Y., Huang, T. S., Ortega, M., & Mehrotra, S. (1998). Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Transactions on circuits and systems for video technology, 8(5), 644-655.
[9] Rui, Yong, Thomas S. Huang, and Sharad Mehrotra. "Content-based image retrieval with relevance feedback in MARS." Image Processing, 1997. Proceedings., International Conference on. Vol. 2. IEEE, 1997.
[10] Zhou, Xiang Sean, and Thomas S. Huang. "Relevance feedback in image retrieval: A comprehensive review." Multimedia systems 8.6 (2003): 536-544.
[11] Thomee, Bart, Erwin M. Bakker, and Michael S. Lew. "TOP-SURF: a visual words toolkit." Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010.
[12] Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110.3 (2008): 346-359.