Concept-Based Image Retrieval

Omar Almoallim
Linjie Fang

Table of Contents

1. **Overview of Concept-Based Image Retrieval (CBIR)**
   - **1.1** What is CBIR?
   - **1.2** How does it relate to Information Retrieval?
   - **1.3** What is the impact on our daily life?
   - **1.4** What are typical applications or uses?
2. **Components of CBIR**
   - **2.1** Query Formation
   - **2.2** Image Representation
   - **2.3** Image Indexing
   - **2.4** Image Retrieval
3. **Algorithms and approaches to CBIR**
   - **3.1 Scale-Invariant Feature Transform**
     - **3.1.1** What is SIFT?
     - **3.1.2** Characteristic and advantage
     - **3.1.3** Major process in implement SIFT
   - **3.2 Histogram of Oriented Gradients**
     - **3.2.1** What is the HOG?
     - **3.2.2** Advantages
     - **3.3.3** Major process to implement HOG
4. **How is CBIR being used today?**
   - **4.1** Current applications and Companies
   - **4.2** Challenges and Future of CBIR
5. **Conclusion**
6. **References**

1. Overview of Concept-Based Image Retrieval

1.1 What is CBIR?

Concept-Based Image Retrieval (CBIR) allows the retrieval of images based on what is present in the photo rather than the associated tags that come with the photo, including metadata. The need for a CBIR has increasingly become more important as we move towards a more media driven society where camera-enabled phones are practically everywhere. Thus resulting in a need to search through images in an effective manner, which currently is not possible. Image retrieval by text today today works by using the text associated with the image, that generally does not provide any important, relevant or effective details and descriptions about the actual image, meaning that this associated data returns irrelevant results to the user and would generally make the search engine unreliable.

In order to implement a Concept-Based Image Retrieval, the system first and foremost requires to understand images at a high level. This presents two particular problems. Firstly, the intention-gap, a non-technical issue. That is a user intends to search something specific, however it is very broad or has many meanings[1]. A great example of the intention gap iis a user query of ‘ apple’ and in return is presented with Apple Inc. rather than the fruit, or vise-versa (even through an image query). The second problem is the image similarity is very far off the users’ query, this is called the semantic gap. This technical problem is one of the challenging aspects of Content-Based Image Retrieval which are being heavily studied and in this paper 2 models will be discussed.

The idea of an image based retrieval is not too distant from that of a text retrieval (just as what we have seen in class). Image search aims to retrieve relevant visual documents/images from a visual corpus. That means that the query can vary depending on the CBIR engine, however, an input of a visual image or sketch is generally the goal, but is not limited to those. Image preprocessing allows the image representation to be stored in the database or index that would be used for the search.

Ultimately, CBIR allows the retrieval of visual documents using the image itself rather than textual information of the image.

1.2 How does it relate to Information Retrieval?

Just as we have seen in class, Concept-Based Image Retrieval retrieves image documents rather than textual documents, and just as textual retrieval provides an information need, image retrieval provides that information need in a different method. Take for example the idea to retrieve geographic location images that are very specific with its contents[1]. Manual interaction with this type of large data is insufficient and inappropriate, especially since it is growing by the day. Thus, CBIR system becomes useful in the treatment of this type of content.

1.3 What is the impact on our daily life?

The overall impact of CBIR is positive. As it provides the proper tools to be able to effectively search, retrieve, organize and manage visual data. Thus saving precious time when searching to get reliable visual results.

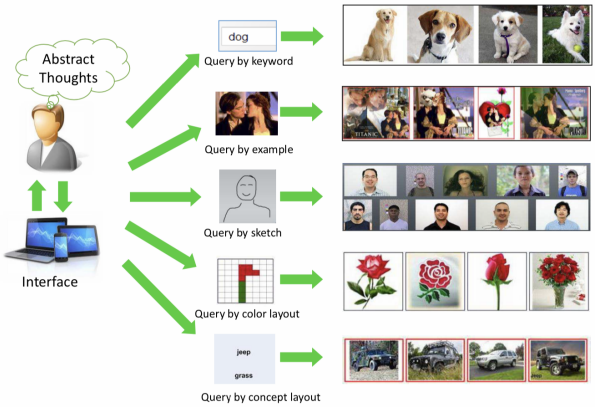
1.4 What are typical applications for CBIR?

CBIR can be used for various activities. A really important and fascinating one is tattoo image retrieval. Tattoos are becoming more common today with individuals and thus have become a method for human identification. Tattoo pigments are embedded in the skin to a level that become permanent on some level. Over years, tattoos have helped identify criminals[3] - even more so when the tattoos have some hidden meaning that becomes easier to identify when there is a dataset of featured images - and also identified victims of different attacks and disasters such as the 2004 Asian tsunami. Another application are individuals who want to find images that are online without the individuals permission.

2. Components of CBIR

2.1 Query Formation

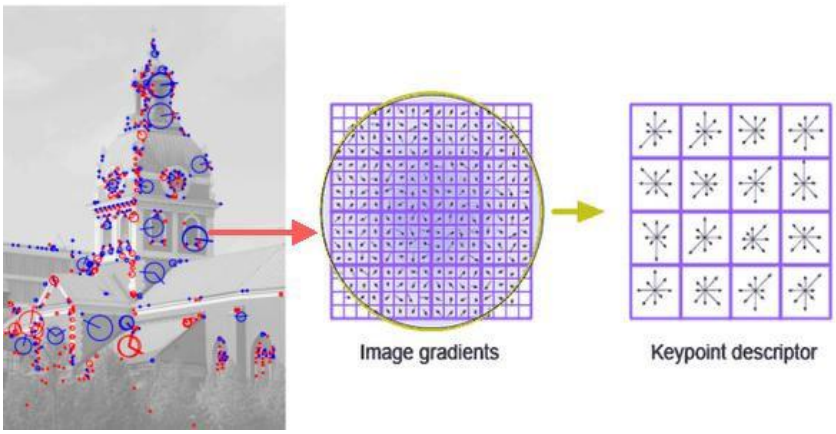
Depending on how the CBIR system is set up, there are different ways that an image retrieval can work, and that is shown in the figure below[4].



The query formation is a crucial step in image retrieval as it has a significant impact on the results. The most common query type in CBIR is a query by example (QBE), that is providing an image similar to what the user wants to find.

2.2 Image Representation

Aside from the user, image representation is another very important part of a CBIR system, ultimately this is how the images will be compared to provide the results. When analyzing images, the low-level features, such as the colors, shapes become somewhat useless to use directly for comparison with images as visual objects and scenes undergo various transformations [4]. Shown below is a way of image extraction, different parts of the image are recognized, then converted into an understandable quantitative description.



[6]

The goal of image representation is to provide a model that is able to effectively and accurately measure image similarity. This is done by extracting the high-level image features and turning them into an understandable object for the engine to understand, simply by Bag-of-Visual-Words or more robust by vectoring features. What is meant by high-level features is the semantics of the image. That is the objects, contents, and scenes in the image as close as possible to how human beings perceive it [5]. A more detail of feature extraction is explained in the next section.

2.3 Image Indexing

Database organization of the image sets is essential to provide efficient and fast retrieval. There are a few ways that the indexing can be built, two of which are inverted document indexes, just as we have seen in class, and the second is hashing based index.

Inverted document index has many different variants depending on the query types allowed. One variant is the Bag-of-Visual-Words, where each image is represented with a fixed dimensional feature vector, however, this would make components as large as 1 million of which many are zeros (image feature is extremely high-dimensional and sparse) [7], thus unsuitable to directly apply the inverted index structure. The solution to this would be the Locality Sensitive Hashing (LSH). LSH is an effective technique to

index dense features. It works by hashing the feature vectors and hashes that are similar to one another would have a more similar hash due to being more probable to have a hash collision[5].

D. Image Retrieval & Ranking

Ranking is very similar to what we have seen in class. One way would be to do a TF-IDF, in CBIR that would be the Feature-Frequency-IDF. This is common as it has been proven effective in textual retrieval [8]. Another way is to use machine learning to rank, there are three diverse models that will be discussed here. Support Vector Machines (CBIR-SVM), Genetic Programming (CBIR-GP), and Association Rules (CBIR-AR). The final analysis of this research concluded that all learning algorithms provide significant improvements when compared to a typical ranking strategy that uses solely descriptors. As well, the research concluded that, in general, CBIR-AR and CBIR-GP outperforms CBIR-SVM. The query given by the user goes through preprocessing to extract its features then compared with the data set, ranked and returned (just as we have seen in class).

3. Algorithms and approaches to CBIR

3.1 The scale-invariant feature transform (SIFT)

3.1.1 What is SIFT?

"The scale-invariant feature transform is a feature detection algorithm in computer vision, published by David Lowe in 1999." [11] It can detect and describe local features in images. Applications include image-based matching and recognition, robotic mapping and navigation, image stitching, 3D modeling, video tracking, etc.

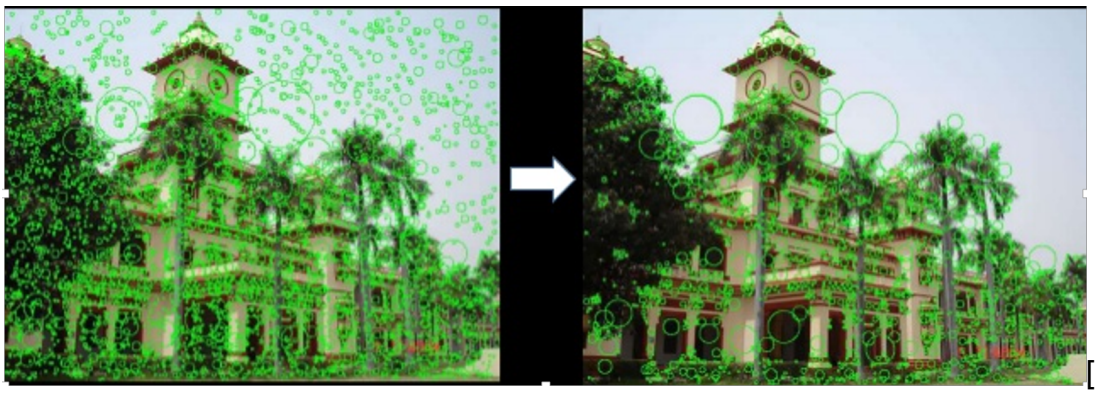
3.1.2 Characteristic and Advantage

A good image feature descriptor in SIFT should have the following characteristic:
(1) "Invariance to image scale and rotation" When the image changes in scale (shrinks, enlarges, etc.), or when the image rotates, we can ensure that the features we extract are unchanged. This is the main feature of SIFT, which is why the name of this feature is called scale invariant features.
(2) "Robustness to Affine distortion, change in 3D viewpoint, Addition of noise, change in illumination" Which means, our features will not change due to noise in the scene in the image, affine distortion, changes in brightness, or changes in perspective when we take pictures. Our features remain robust to these effects.
Base on its characteristics, the algorithm has the following advantage:
Accuracy, stability: features are local so not affected by scale & rotation
Distinctive feature: we are able to match the feature with the database that contains large amount of image

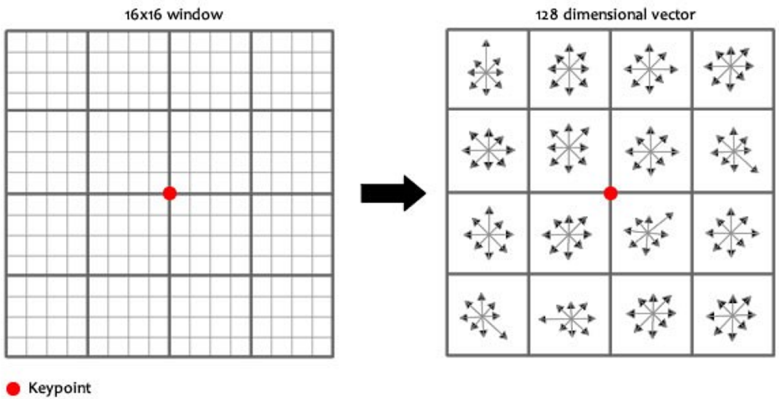
Quantity: Even a small object can generate a large number of SIFT features.
Efficiency: SIFT has close to real-time performance and high efficiency.

3.1.3 Major Process in Implement SIFT

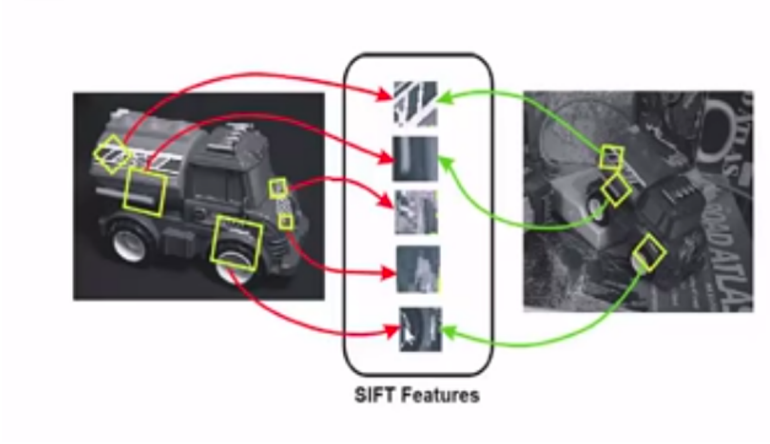
1. Feature point detection: The purpose of this step is to detect the potential key points in scale space that satisfy scale invariance and rotation


[17]

2. Feature point localization: based on the key point we found in the first step, refines the location of these feature points to more accurately locate the location of the feature key points and remove any poor features at the same time.
3. Orientation Assignment: The purpose of this step is based on the local gradient direction of key points assigning orientation to each key point.



4. Key point descriptor: At this point, each key point has a location, scale, orientation. Then we can use a high dimensional vector to describe each key point, the example in the image is to convert each key point into a high dimensional vector like the image above.
5. Key point Matching: Lastly Key points between two images are matched by identifying their nearest neighbors.



[12]

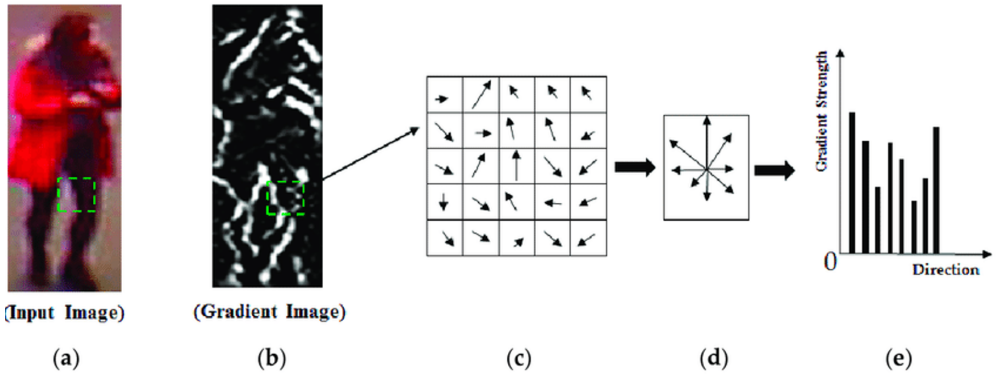
3.2 The Histogram of Oriented Gradients (HOG)
3.2.1 What is the HOG?

The Histogram of oriented gradients is also a feature descriptor used to detect objects in computer vision and image processing. “The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image”[15]. Hog + SVM has been widely used in image recognition, especially in pedestrian detection. How it works is basically First divide the image into many small connected areas, we call it the cell. Then, the gradient or edge direction histogram of each point in the cell is collected. Finally, these histograms can be combined to a feature descriptor

3.2.2 Advantage

Compared with other feature descriptions, HOG has many advantages. First of all, since HOG operates on the local grid cells of the image, it can maintain a good invariance to the geometric and optical deformations of the image. Secondly, under the different conditions as long as the pedestrian can generally maintain an upright position, the pedestrian can be allowed to have some body movements. These movements can be ignored without affecting the detection. Therefore, HOG features are particularly suitable for human detection in images.

3.2.3 Major process to implement HOG



[13]

1. Preprocessing
Grayscale a three-dimensional image of the target, because color information is usually not useful, it is usually converted into a grayscale image first. Divide the image into small cells like 6*6 pixels.

2. Calculate the Gradient Images
To calculate a HOG descriptor, for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell. The main purpose is to capture contour information while further weakening the interference of lighting.

3. Calculate Histogram of Gradients in cells
We need to first calculate the horizontal and vertical gradients; after all, we want to calculate the histogram of gradients for each cell, then histogram of gradients is calculated for each cell.

4. Block Normalization and calculate the HOG feature vector
We take a set of cells as a block. Then we can get the HOG feature descriptor of the image by concatenating the feature descriptors of all blocks in the image.

4. How is CBIR being used today?

A. Current applications and Companies

Some companies that have dominated CBIR include TinEye, whose mission is to make images searchable. Others include Ditto, which works to analyze photos shared in social media. The medical field is increasingly being digitized which means more visual data that needs to be retrieved and searched through [9].

B. Challenges and Future of CBIR

The main challenge that is faced today with Concept-Based Image Retrieval is the semantic gap between the user's needs and the capability of CBIR algorithms [10]. The advancement of CBIR might be heading towards a new query paradigm to extract the user perception, and would make the query more customizable. As well, a more intelligent system that would be able to extract more quality semantics, this is also evident due to the advancement of Artificial Intelligence.

5. Conclusion

In conclusion, Concept Based Image Retrieval allows an effective way to access visual data and retrieval of it. It begins with preprocessing to turn the image contents into vector features, that is done by various models including SIFT and HOG which can generate a high dimensional vector for each local features to represent the image or calculate and analyze the gradient histogram of each block in an image and generate the feature descriptor. Then indexing the data in a way that allows the retrieval in a fast and accurate approach, that could be by an inverted index or by a locality Sensitive Hashing structure. Most common query type is a Query by Example, meaning the query is a visual image. CBIR proves to be relevant and important in fields such as police investigation by analyzing tattoos, or by medical professions to retrieving visual documents.

References

[1] Yang, Linjun & Geng, Bo & Hanjalic, Alan & Hua, Xian-Sheng. (2012). A Unified Context Model for Web Image Retrieval. ACM Transactions on Multimedia Computing, Communications, and Applications - TOMCCAP. 8. 10.1145/2240136.2240141.
[2] Laban, Noureldin. (2014). Geospatial Image Retrieval: Proposed System For Content Based Satellite Image Retrieval.
[3] Lee, Jung-Eun & Jin, Rong & Jain, Anil & Tong, Wei. (2012). Image Retrieval in Forensics: Tattoo Image Database Application. IEEE MultiMedia. 19. 40-49. 10.1109/MMUL.2011.59.
[4] Zhou, Wengang & Li, Houqiang & Tian, Qi. (2017). Recent Advance in Content-based Image Retrieval: A Literature Survey.
[5] Zhao, Rong & Grosky, William. (2001). Bridging the Semantic Gap in Image Retrieval.
[6] https://miro.medium.com/max/1186/1*K68boX7fmtsYmyG2LlcmhQ.jpeg
[7] Zhang, Xiao & Li, Zhiwei & Zhang, Lei & Ma, Wei-Ying & Shum, Heung-Yeung. (2009). Efficient Indexing for Large Scale Visual Search. 1103-1110. 10.1109/ICCV.2009.5459354.
[8] Zheng, Qing-Fang & Gao, Wen. (2008). Constructing visual phrases for effective and efficient object-based image retrieval. TOMCCAP. 5. 10.1145/1404880.1404887.
[9] João Augusto da Silva JúniorRodiney Elias MarçalMarcos Aurélio Batista (2014). Image Retrieval: Importance and Applications.
[10] Aslam, Nida & Irfanullah, & Loo, Jonathan & Roohullah,. (2009). Limitation and Challenges: Image/Video Search & Retrieval. JDCTA. 3. 98-102. 10.4156/jdcta.vol3.issue1.aslam.
[11]D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2.
[12]https://medium.com/analytics-vidhya/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40
[13]https://www.researchgate.net/figure/Methodology-of-image-features-extraction-using-the-histogram-of-oriented-gradients-HOG_fig5_315327257
[14]Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
[15]Mostafa Ezziyyani,"Advanced Intelligent Systems for Sustainable Development",(2019)
[16]Zhang, H. and Cao, L., 2013. An Image Matching Algorithm Based on SUSAN-SIFT Algorithm. Applied Mechanics and Materials, 325-326, pp.1588-1592.
[17]https://www.slideshare.net/kislayabhi/sift-41469195
[18]Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.(2016:Google Citation: 14046)