



UNIVERSITY OF GOTHENBURG

# CENTERED FEATURES FROM INTEGRAL IMAGES

**ZOHAIB KHAN**

**Bachelor of Applied Information Technology Thesis**

**Report No. 2009:070**

**ISSN: 1651-4769**

## **Abstract**

*An essential tool for fast image processing is the computation of image features. Integral Images are a popular approach to obtain this goal. A drawback of this technique is that it only allows for rectangular image features. In this paper, the goal is to investigate if circular image features outperform rectangular image features using integral image approach. Additionally, experimental work for the purpose of comparison between different and new descriptor methods has also been presented in this paper.*

## **1. Introduction**

Visual object recognition is a branch of computer vision and one of the hot topics in computer vision. It is a process of finding a visual object in an image or a video sequence. Recognizing an object requires significant competence for any automated system. For normal humans it is very easy to recognize most object categories no matter if it is different in size or shape or even rotated, but this process is still challenging in the field of computer vision.

Images have become an important part of our daily life and are being used for daily communication as well. There are huge numbers of images available digitally which humans can no longer manage themselves (Chen and Wang, 2002). For this reason, we need to categorize the images e.g. bus, car, cycle etc which is a difficult task itself due to the vast variation between images of same class and also among several classes. Therefore, object recognition techniques are used to automatically generate the content descriptions for images (Grosky and Mehrotra, 1990).

BoW (Bag-of-Words), originally from document analysis domain, is a popular method for document representation which ignores the order of the words/sentence. In computer vision this method is being used for image representation where an image is taken as an object and the features extracted from the image are known as features. Using this method, we make a visual vocabulary which can be used to search through and recognize an object visually seen. Basically, it can be taken as a database of features extracted from different images. This method is known as “*Bag of Visual Words*” or “*Bag of Features*” in computer vision.

In image processing, features are referred to interesting points of an image whereas feature detection is a process of finding interesting points in an image. Feature detection is generally taken as a first operation being performed on an image. The purpose of feature detection is to examine every pixel of an image and locate if there is any point of interest is present at that pixel. There are also several algorithms available for feature detection and algorithms often only examine image in the region of the features.

For the purpose of object detection Viola and Jones (2001) proposed rectangular features (RF's) which has showed high accuracy and efficiency. Rectangular features are used for the purpose of feature extraction from an image by removing noise using

rectangular path (area).

## 1.1 Related Work

A significant success has been achieved by using “bag of visual words” framework for the purpose of object recognition and classification (Dorko and Schmid, 2003, Fei-Fei and Perona, 2005). This method starts with detecting key-points of interest from an image. Key-Point detection is a process of finding areas of interest in an image. The detection is followed by the representation of key-points using local descriptors. This step involves description of local features detected from an image.

The local features are similar to textual words in document analysis techniques (Baeza-Yates and Ribeiro-Neto, 1999). In image processing the local features are taken as visual words (Zhang, Marszalek, Lazebnik, and Schmid, 2007). After this, a classifier is trained to recognize the categories of different images and the images are being classified according to their categories based on the visual words. The purpose of image categorization is to extract features and locate the relevant visual words so that a classifier is applied to these visual words representing an image.

Therefore, Bag of Visual Words framework is being used for this project due to its success in this area of field. Integral images approach can compute rectangular features swiftly and quickly (Viola and Jones, 2001). Therefore, Integral images approach was applied for the purpose of image computation in bag of visual words framework to test the efficiency of this approach at an image description step.

In this paper, experimental work at feature description level has also been revealed. Hue Descriptor (Weijer and Schmid, 2006), Color Name Descriptor (Weijer and Schmid, 2007) and Dense Derivative Histogram methods are three different feature description methods which were tested and compared using three types of shapes square, circle and Gaussian in bag of visual words framework. The results are presented in the results section.

## 1.2 Organization

After this section, the paper has been organized in four sections. In section 2 the approach to solve this problem has been discussed. Section 3 presents the experimental details such as framework, its details and dataset used. Section 4 presents results of this project. Finally, summary and possible future work is concluded in section 5.

## 2 Integral Images

Integral image (Viola and Jones, 2001) which has close relation to summed area tables (CROW1984) in graphics was first presented by Voila and P Jones in 2001. The approach is used to sum up the features at any area in an image. In integral image,

every point (Fig 01) has a sum of its preceding points to the above and to the left sides (x-axis and y-axis).

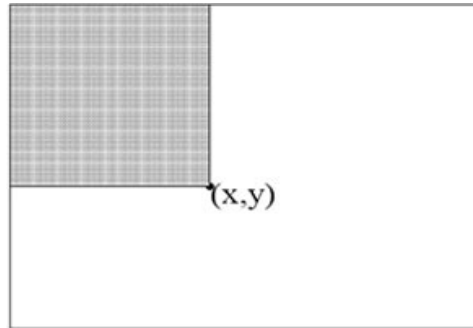


Fig 01 - Displaying a point holding sum

$$ii(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y'),$$

The integral image at position x, y holds the sum of the pixels whereas,  $ii(x, y)$  is the integral image and  $i(x, y)$  is the original image.

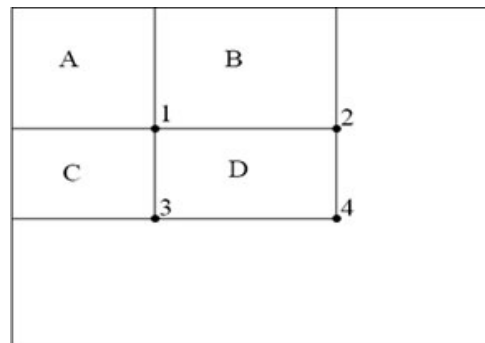


Fig 02 – Displaying pixel areas

In Fig 02 the sum of points in area D can be summed up using four references. Point 1 holds the sum of area A whereas, point 2 holds the sum of area A+B. Point 3 holds the sum of A+C. Point 4 holds the sum of A+B+C+D. Therefore, by this we can get the sum of area D by computing all the areas as  $4+1-(2+3)$  (Viola and Jones, 2001).

## 2.1 Limitations of Integral Images

As shown in the previous section, integral images have the advantage that with only 3 summation/additions the summation over a square region in the image can be computed. However, this comes at the cost of computing the integral image. Since only one integral images need to be computed, the algorithm will be more efficient when many features are computed in the image. This leads to our first research question:

*How many features per image need to be computed for integral images to be more efficient than the classical approach of summing regions?*

A drawback of integral images is that they only compute the summation of squared regions in the image. For many applications circular features might be more desirable. Secondly, it is known that assigning more importance to center pixels than to pixels far away from the center often improves results. The most used weighting filter is the Gaussian filter. These limitations of traditional integral images lead to our second research question:

*Do circular and Gaussian weighted features outperform squared features in an object recognition task?*

A positive outcome of the above question would justify further research into the development of circular and Gaussian integral images. A sketch of such features is given in the next section.

## 2.2 Circular and Gaussian Integral Images

The integral of an image can be taken in a crossed manner (Fig 03). The purpose of taking integral in this manner is to be able to compute features detected using circular features. Circular features method has not yet been fully developed but, it is expected that using integral image approach for computing circular features will be as efficient as it has been seen in this research for rectangular features.

In Fig 01, the integral image is computed by taking the integral in the horizontal and vertical direction. However, theoretically it is also possible to take integrals in other directions, for example the diagonal direction. In this case, each point in the image would contain the integral (sum) of the pixels of a triangle, as illustrated in Fig 03. Combining these skewed integral images would allow us to approximate the sum of a circular region. A combination of circular integrals at different scales could be used to approximate Gaussian weighted features.

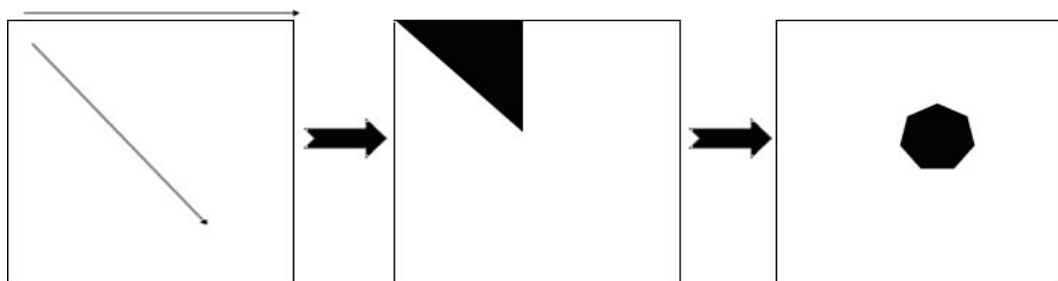


Fig 03 – Illustration of circular integral Images

### 3. Experimental Setup

The performance of the integral images for centered features has been tested on the dataset of Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce (2004). The details of the procedure of testing and the methods being used for this purpose are being summarized in the following sections.

#### 3.1 Object Recognition Framework

For the purpose of object recognition, Bag of Visual Words approach is known to be the most thriving approach. There are five major steps involved in this approach detection, description, visual vocabulary, classification and assignment (Weijer, and Schmid, 2007).

In the First step, it detects the regions in an image and its related scales using some detection method such as Harris-Laplace Detector (Mikolajczyk and Schmid, 2004) and Difference of Gaussian Method (DoG). The second step in this framework is to normalize all the patches of images to a standard size and compute the descriptor for all the detected regions. A well known and common method for feature detection is SIFT (D. Lowe, 1999, 2004).

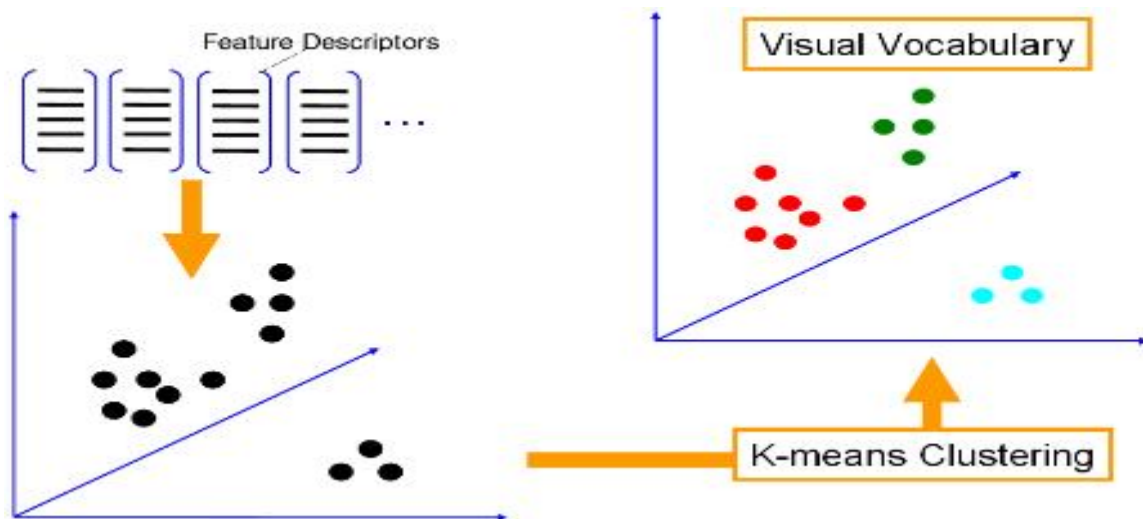


Fig 04 – Displaying how information receives and clustered

Third step is to make a visual vocabulary (vocabulary of visual words). Like other steps, this process also requires some method for arranging the visual words. K-means is a renowned and widely used partition clustering method (Xiong, Wu and Chen, 2009). Fig 04 illustrates how this method works in Bag-of-Visual Words framework. Fourth step is to represent all the images through a frequency histogram of visual images. After this, a classifier is trained according to the histograms of images for every class. Last and final step in this whole process is to assign all the images to the classes accordingly (Weijer and Schmid, 2007).

#### 3.2 Feature Detection

Features detection is the first step in image processing. The purpose of feature detection is to find points and regions of interest in an image. There are two different types of feature detector exists, region detector and corner detector. Region detector is used to detect some region in an image whereas corner detector detects the corners in an image (Jing and Allinson, 2008).

In this research, random feature detection method has been used. This method detects features randomly from an image and stores it into a file with x-axis features and y-axis features format.

### **3.3 Feature Descriptor**

Features description is the second step in image processing. The main purpose of feature description is to retrieve the vector of pixel values of an image. Features can further be divided into two types one is local features and second is global features.

Local features are generally described as such features which are a part of an object in an image. Global features are such features that are in the surroundings of an object in an image. In this research, we have used three different types of feature descriptors Hue Descriptor (Weijer and Schmid, 2006), Color Name Descriptor (Weijer and Schmid, 2007) and Dense Derivative Histogram.

### **3.4 Visual Vocabulary**

Creating a visual vocabulary is a challenging task. Visual vocabulary (Bag of Visual Words) contains all the local descriptors extracted from the images. In this project, the vocabulary is constructed by applying the K-means algorithm to the set of local descriptors extracted from training images. Euclidean distance was used in clustering. The vocabulary size is optimized on the performance score.

### **3.5 Classification and Assignment**

Classification and assignment is an important step of this framework. The classifier is used to classify the images. A classifier is trained for all the classes (categories) of images based on the information obtained from visual vocabulary. This process is done using training images. After this, all the images are assigned to the appropriate classes.

In this project the dataset was divided into train set and test set. The train set contains 200 images and 200 images for test set. For the purpose of experiment two, multiclass SVM (Support Vector Machines) has been used for classification. SVM is a set of learning methods used for classification. SVM is often used for less object categories. To check the classification performance and accuracy, classification score has been used. The classification score gives the percentage of correctly classified images in the test set.

### **3.6 Dataset**

In image processing field, dataset is known as set of different types and kinds of images. These images later are divided into two halves. Half of the images are used for training purpose of Bag of Visual Words framework and other half is used for the purpose of testing the results.

In this research, we had used dataset of Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce (Lazebnik, Shmid, and Ponce, 2004). This dataset contains 619 images in seven different classes of different types of butterflies in jpeg image format. But, we had used 400 images with 5 classes.

## 4 Experiments

This section explains the details of experiments conducted. The first experiment shows the results obtained by applying integral images approach on rectangular features. The first experiment leads to our first research question stated in section 2.1. The second experiment shows the results obtained by comparing three feature descriptor methods. This experiment leads to our second research question stated in section 2.1.

The experiments were tested on an IBM T60 Laptop machine having Dual Core processor 1.66 Ghz and 1.0 Gigabyte of RAM.

### 4.1 Experiment 1: Integral Image versus Classical Computation

The first experiment is about the integral image approach being used with rectangular features. Integral image approach has been compared with the classical computational method. For this experiment, features (points) ranging from 50 to 300 in the area (radius) ranging from 10 to 60 has been used for testing.

In Fig 05 it can be seen that the integral images approach starts performing better than the classical approach after 180 image features approximately. The results in fig 05 also show that the speed of feature computation using integral image approach remains almost linear even if the size of features grows.

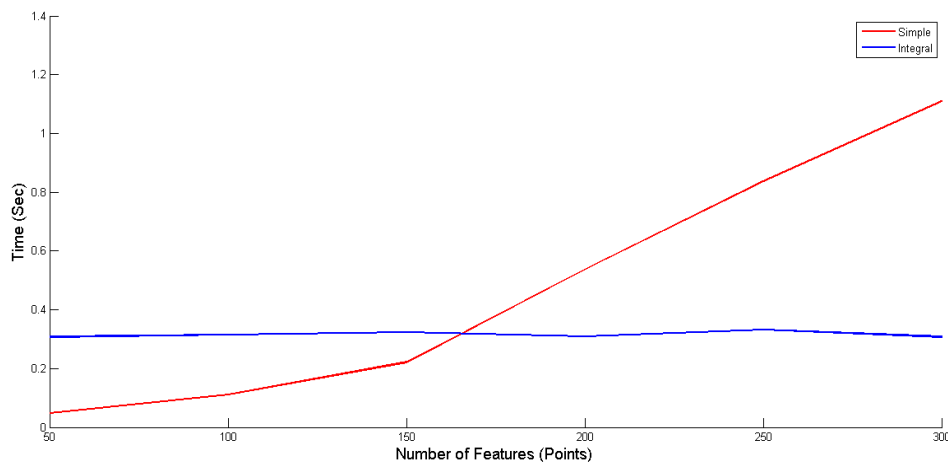


Fig 05 – Results of Integral vs Classical Computation



## Experiment 2: Comparison of Feature Description Methods

The second experiment is about the comparison between three different features description methods Hue, Color Name and Derivative Histogram. The results are presented in the tabular form showing the classification accuracy rate for object recognition. Three different methods were used and have been tested against three types of image features methods i.e. Rectangular, Circular and Gaussian.

	<b>Rectangular</b>	<b>Circular</b>	<b>Gaussian</b>
Hue Descriptor	35.1 %	33.4 %	32.9 %
Color Name	39.2 %	37.4%	33.2 %
Derivative Histogram	40.5 %	41.0 %	41.5 %

The results from the above table shows that the circular image features and Gaussian image features were not able to perform so well for the purpose of object recognition. But, both the image features methods (Circular and Gaussian) are underdeveloped methods so these results may require retesting once these methods are finalized.

## 5 Conclusion and Future Work

This research has been more focused towards computing features using integral images approach. The result clearly shows in the experiment one that the integral images approach can compute features more quickly as compare to the classical computation. The approach has been tested with rectangular features and the results clearly outperformed the classical method.

The results from the experiment shows that the integral image approach is more efficient if there are more than 180 image features in an image. But, the speed of computation remains almost linear even if the numbers of points grow more. From this research, it is also concluded that the fact integral images are rectangular does not limit their performance.

This research further motivates to develop circular and Gaussian feature method as well as circular and Gaussian integral image method for computing circular and Gaussian features.

## **ACKNOWLEDGEMENTS**

I would like to specially thank Joost van de Weijer (Phd) for being my supervisor. It was an honor working with him. His assistance and support throughout this research helped me to finalize this research. I would also like to thank Maria Vanrell (project leader of color and texture group) for giving me a chance to work in her group at Computer Vision Center, UAB, Barcelona.

I'm also great full to Fahad Shahbaz for helping me in understanding technical issues of Bag of Features Framework. Thanks to Faramarz Agahi, my supervisor at IT University of Gothenburg, Sweden and my opponnet Che fu Christopher for their useful comments on this research document.

The love and support of family needed to complete this long educational journey can never be exaggerated. I am so much thankful to my mama for her prayers and abbu (maternal uncle) for their never ending support. Without both, I may not be able to what I am today. Last but not least, Sehrish Rafi for proof reading my document.

## References

- CROW, F. C. 1984. Summed-area tables for texture mapping. In SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques, ACM Press, New York, NY, USA, 207–212.
- D. Lowe. Distinctive image features from scale invariant, keypoints. International Journal of Computer Vision (IJCV), 60:91-110, 2004
- G. Dorko and C. Schmid. “Selection of scale-invariant parts for object class recognition. In Proc. ICCV, 2003”.
- Hui Xiong; Junjie Wu; Jian Chen; “K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective Systems”, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on Volume 39, Issue 2, April 2009 Page(s):318 – 331
- J. van de Weijer, Cordelia Schmid, “Applying Color Names to Image Description “, Proc. ICIP, San Antonio, USA, 2007.
- J. van de Weijer, C. Schmid, Coloring Local Feature Extraction , Proc. ECCV, Part II, 334-348, Graz, Austria, 2006.
- Jianguo Zhang, Marcin Marszalek, Svetlana Lazebnik, and Cordelia Schmid. Local features and kernels for classification of texture and object categories: a comprehensive study. International Journal of Computer Vision, 73(2):213{238, 2007.
- Jing Li and Nigel M. Allinson “A comprehensive review of current local features for computer vision“, 2008
- K. Mikolajczyk and Cordelia.Schmid, “scale and affine invariant interest point detectors” International Journal of Computer Vision, vol 60, no.1, pp. 62-86, 2004
- L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In Proc. CVPR, 2005.
- Lowe, D.G.; "Object recognition from local scale-invariant features", IEEE International Conference on computer vision, Volume 2, 20-27 Sept. 1999 pg :1150 - 1157 vol.2
- P. Viola and M. Jones, “Robust real time object detection,” In Proc. of IEEE ICCV Workshop on Statistical and Computational Theories of Vision, July 2001.
- Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Computer Vision and Pattern Recognition CVPR 2001
- R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999.

Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Semi-Local Affine Parts for Object Recognition. Proceedings of the British Machine Vision Conference, September 2004, vol. 2, pp. 959-968.

W. Grosky and R. Mehrotra. Index-based object recognition in pictorial data management. Computer Vision, Graphics and Image Processing, 52(3):416{436, December 1990.

Y. Chen and J. Z. Wang. Region-based fuzzy feature matching approach to content-based image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(9):1252{ 1267, 2002.