

CARTOON BASED IMAGE RETRIEVAL : AN INDEXING APPROACH

Suman Muralidhar¹ and Sharath Kumar Y H²

¹Department of Computer Engineering, Maharaja Institute of Technology, Mysuru

² Department of Computer Engineering, Maharaja Institute of Technology, Mysuru

ABSTRACT

This paper proposes a methodology for the content based image retrieval which is implemented on the cartoon images. The similarities between a query cartoon character image and the images in database are computed by the feature extraction using the fusion descriptors of SIFT (Scale Invariant Feature Transforms) and HOG (Histogram of Gradient). Based on the similarities, the cartoon images same or similar to query images are identified and retrieved. This method makes use of indexing technique for more efficient and scalable retrieval of the cartoon character. The experiment results demonstrate that the proposed method is efficient in retrieving the cartoon images from the large database.

Keywords

Content based image retrieval, SIFT, HOG, Indexing

1. INTRODUCTION

Cartoon is one of the most admired artificial arts in the history. Cartoon images play essential role in our everyday life especially in the field of entertainment, education and advertisement due to its increasing intensive research in the field of multimedia and computer graphics. The rapid growth in the volume of cartoon characters in the database the necessity for retrieving the cartoon images has been increased. Cartoon images without proper annotation make them incapable of retrieving the exact image of interest; hence content based cartoon image retrieval system retrieves the cartoon images with content based instead text annotation. The cartoon based image retrieval systems can be defined as: For a given query cartoon image, finding similar cartoon images stored in database. The search process relies on how faster the images can be retrieved from the large database. To achieve a fast retrieval speed and to make the retrieval system truly scalable for the large size of the image collections, an effective indexing structure is a paramount part of the whole system. The process of cartoon based image retrieval system involves many stages. To retrieve the images, the shape information for each image is extracted and the special relationship between the images is considered this is used for the close similarity between the query image and the database image. The image contains the edge information, texture information and region information. The similarity between the relationship among the query image and the database image can be used to retrieve the images based on the indexing mechanism.

Cartoon based image retrieval has its remarkable application in the field of advertising, education and entertainment. It has its vast application in the field of computer graphics and multimedia applications. Retrieved Cartoon images can be used to produce cartoon movie clips and key-frames drawing and hence it reduces the human efforts in producing new cartoon clips. It can be useful poster making and designing the operating system themes. Retrieved cartoon images are useful for animators and cartoonists to effectively create new animations by reusing and

synthesizing the retrieved images without much of effort. Content based cartoon image retrieval-when drawing new cartoon frames the cartoonist can retrieve similar images for the reference. It helps non professional user to synthesize new cartoon clips by reusing the existing cartoon resources.

Challenges in the cartoon based image retrieval includes: The representation and appearance of the cartoon images may vary according to the context. Cartoon images have different colour or drastically changed shapes which is not fixed [1]. There may be inter-class variation among the each class of the images. Another challenge in cartoon image retrieval is that the gesture of the cartoon images may vary. Due to the striking growth in cartoon image database needs speedy and efficient cartoon image retrieval system for this purpose, so in this work we use an approach called indexing.

2. RELATED WORK

Haseyama and Matsumura [2] have proposed a cartoon image retrieval system; they defined Regions and Aspects to compute cartoon image similarities. However, the partial features were retrieved as results which are not suitable for cartoon character gesture recognition. It can only be used to retrieve face image of cartoon characters.

In Yu et al. [3] developed Graph based cartoon clip synthesis (GCCS) that combines similar characters into user-directed sequence based on dissimilarity measure in edges. It performs well for simple cartoon characters. Since the synthesis result is fixed by initial specification, for a complex pattern GCCS fails to generate smooth clips because edges can encode neither the color information nor the gesture pattern.

Yang et al. [4] recently proposed Retrieval based Cartoon Clip Synthesis (RCCS) that combines similar characters based on dissimilarity measure in cartoon characters motion and edges. Based on these two features, RCCS uses an unsupervised bi distance metric learning (UBDML) algorithm to measure dissimilarity. Drawback of RCCS is that, they fail to retrieve dissimilarity in color and gesture.

In [5], an unsupervised distance metric learning algorithm is proposed for cartoon image retrieval. The algorithm is formulated as a trace ratio maximization problem, which is optimized by the iterative approach. The iterative approach usually converges after several iterations. It requires performing eigen decomposition of a matrix during each iteration and there is no closed solution for the optimization.

3. METHODOLOGY

In this section, we present details on proposed method to build an efficient cartoon image retrieval system which can retrieve the similar cartoon image for a given query image with some target which includes: Collection of dataset, Pre-Processing of cartoon images, Efficient feature extraction and dimensionality reduction technique for cartoon representation, Effective indexing method for efficient retrieval. Architecture of the proposed approach is shown in the Figure 1.

3.1. Pre-Processing

The pre-processing task involves enhancing the image to improve the comprehensiveness information in the image and Edge detection is used reduce the unnecessary information in the image while preserving the structural part of the image.

3.2. Feature Extraction

A well defined feature extraction will help in precise retrieval of images. It performs two important tasks: Transforms input parameter vector into feature vector and reduces the dimension of feature vector. The pre-processed image is used for the feature extraction. The features for the input query is extracted using SIFT (Scale Invariant Feature Transformation), HOG (Histogram of Oriented Gradients) and fusion of this features. The dimensions of the extracted features are reduced using dimension reduction technique called Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The reduced features are indexed for the retrieval process.

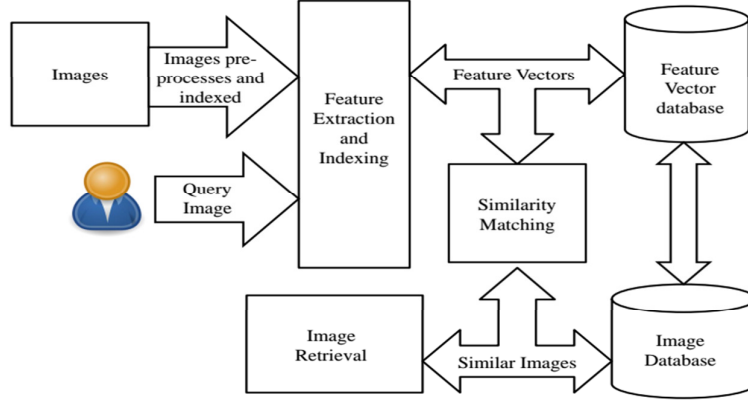


Figure 1. Block diagram of proposed work

3.2.1 Scale Invariant Feature Transform (SIFT)

Scale Invariant Feature Transform (SIFT) [6] algorithm has four important phases, Extrema detection: The images that are not similar to their surroundings are inspected into separate points based on scales and octaves. These points become potential candidates for image features called as extrema. Key point Detection: The algorithm first finds the αN^2 extrema then further refines them to $\alpha\beta N^2$ key points, which is key point of the image. To detect the extrema on edges, a 2×2 matrix is generated and some simple calculations are performed on it to generate a ratio of principle of curvatures and this quantity is compared with a threshold value to decide whether an extrema is to be rejected or not and low contrast points on the edges are rejected. Orientation Assignment: $\alpha\beta N^2$ points are converted to vector points by computing the magnitude and direction. The magnitude $m_i^j(x, y)$ and orientation $\theta_i^j(x, y)$ for each point $L_i^j(x, y)$ can be computed by:

$$m_i^j(x, y) = \sqrt{(L_i^j(x+1, y) - L_i^j(x-1, y))^2 + (L_i^j(x, y+1) - L_i^j(x, y-1))^2} \quad (1)$$

$$\theta_i^j(x, y) = \tan^{-1} \frac{(L_i^j(x, y+1) - L_i^j(x, y-1))}{(L_i^j(x+1, y) - L_i^j(x-1, y))} \quad (2)$$

Key point descriptor: The vector points obtained in the previous phase is consolidated to a set of 8 vector points called descriptor. Each descriptor is converted into feature by normalizing the sum of vector.

3.2.2 Histogram of Oriented Gradients (HOG)

In HOG feature extraction process [7], the feature extraction is done from a local region with 16×16 pixels. The 8 orientations Histogram of Gradients are calculated from of 4×4 cells. The total number of HOG features become $128 = 8 \times (4 \times 4)$. The HOG feature extraction is done from 16×16 local regions. The cells are the local regions that are divided into small spatial area. 4×4 pixels is the size of each cell. 8 orientations Histogram of gradients are calculated from each local cell. Hence, the total number of HOG features will become $128 = 8 \times (4 \times 4)$ and it builds a HOG feature vector.

3.2.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) [8] is the global feature extraction method. It is a powerful technique for extracting the global structure from the high dimensional dataset and that has been broadly used to reduce the dimensionality and also to extract the abstract features of the cartoon images.

Algorithm 1: Principal Component Analysis

Input: Input feature set where features, $f_1, f_2, f_3, \dots, f_n, f_i \in F$

Output: Reduced feature vector $f_p, f_{p+1}, f_{p+2}, \dots, f_q$ where $p \geq 1$ and $q \leq n$

1. Subtract the mean value from each data dimension, New data set with mean is zero is created.
 2. Calculate the covariance matrix: $C^{m \times n} = (C_{i,j}, C_{i,j} = Cov(Dim_i, Dim_j))$ where, $C^{m \times n}$ is a covariance matrix between two separate dimensions.
 3. Calculate the eigenvectors and eigen values of the covariance matrix.
 4. Construct a feature vector from the list of eigenvectors:
FeatureVector = (eig_1, eig_2, ..., eig_n)
 5. Acquire the new dataset by transposing the FeatureVector and multiply it to the original data set. FinalData = RowFeatureVector x RowDataAdjusted.
-

3.2.4 Linear Discriminant Analysis (LDA)

Linear discriminant analysis [9] [10] reduces the number of cartoon image features to a more manageable number before classification. Every new dimension is a linear combination of pixel values. LDA finds the feature representation where within-class distance is minimized and the between-class distance is maximised. Linear Discriminant Analysis easily handles the case in which the within-class frequencies are unequal and their performance has been observed on randomly generated test data. This ratio of the between-class variance to the within-class variance of any particular data set is maximized in this method and this guarantees maximum separability (Refer algorithm2).

Algorithm 2: Linear Discriminant Analysis

Input: Input feature set where features, $f_1, f_2, f_3, \dots, f_n, f_i \in F$

Output: Reduced feature vector $f_p, f_{p+1}, f_{p+2}, \dots, f_q$ where $p \geq 1$ and $q \leq n$.

1. Compute the d -dimensional mean vectors for the different class from the dataset.
 2. Compute the between-class and within-class scatter matrices.
 3. Compute Eigen vectors (e_1, e_2, \dots, e_d) and eigen values ($\lambda_1, \lambda_2, \dots, \lambda_d$) for scatter matrix.
 4. Eigen vectors are sorted in decreasing eigen values and then select k eigen vectors which has the eigen values to form a $d \times k$ dimensional matrix.
 5. Transform the samples onto the new subspace.
-

3.3. Indexing

The remarkable growth in cartoon image database needs speedy and efficient retrieval. Since the search time for Kd-tree indexing is optimal, we use Kd-tree indexing approach in this work for the retrieving the images from the database.

3.3.1 Kd-Tree Indexing

To find the nearest neighbour of a query point, a top-down searching is done from the root to leaf nodes [11]. At each internal node, it is required to check where the partition hyper plane the query point lies, and then the associated child node can be accessed accordingly. To reach the leaf node, the bottom up process requires the comparisons of \log_2^n . The data points with the first leaf node becomes the first element for the nearest neighbour. A backtracking procedure or iterative search is followed to other leaf nodes for searching the better elements. Priority search is the most widely used scheme with high chance to find the true nearest neighbour. All the cells are searched in order of distance from the query point. The search process terminates when there are no cells within the distance which is defined as the best point.

4. EXPERIMENTS

The database consists of thirty classes. Each class has twenty cartoon images in it. All the cartoon images are collected from the different cartoon series and internet. The images are present in a different background, position or angle. The features of these cartoon images can be classified into two groups, Intra-features is useful for measuring the similarity between data of the same class and Inter features are used to measure the dissimilarity among the data from different classes. The large intra-class variability and the small inter-class variability make this dataset very demanding. Training dataset of Mickey Mouse class is shown in the Figure 2. During experimentation, we conducted three different sets of experiments. For the first set of experiments, we used 30% of the images for training and remaining 70% of images for testing. In the second set of experiments, the number of training and testing ratio is 50:50. In the third set of experiments, we used 70% for training and 30% for testing. The experiments were conducted for using the feature extraction of SIFT, HOG and Fusion descriptors for different dimensionalities using PCA and LDA. Experimental results are shown in Figure 3 and 4. In the plot X-axis represents the top samples and Y-axis represents the accuracy. The results for the fusion descriptor are of higher accuracy when compared to SIFT and HOG. The experimental results prove that the proposed method efficiently retrieves the cartoon images from a large database.



Figure 2. Training Dataset of Mickey Mouse Class

5. CONCLUSIONS

This paper proposes a method to retrieve the cartoon images for the user query. To obtain the efficiency of the proposed method we pre-processed the image and extracted the shape feature by SIFT, HOG and Fusion of both from the query image and the pre-processed image dimensionality reduction techniques such as PCA and LDA for different dimensions. Later similarity matching is done and image is retrieved from the large database. The experiment has been conducted on database of having training samples of 30, 50 and 70 percent of database. For efficient retrieval Kd-Tree indexing is used. The result of fusion descriptors has relatively higher accuracy when compared to the HOG and SIFT descriptors. The experimental results prove that our proposed approach is more effective in cartoon based image retrieval from a large database. We have observed that the cartoon based image retrieval is still an open problem [12]; so far we have worked on thirty classes, in future the database can be extended to more classes and manually draw the images in an interface and to retrieve the image from large database.

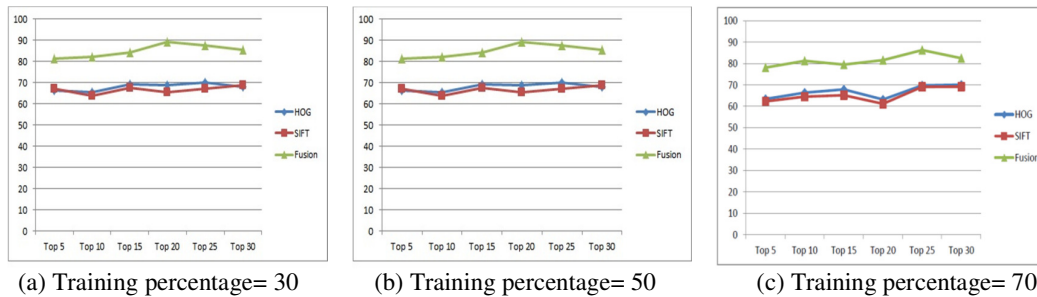


Figure 3. Experimental results of Principal Component Analysis with dimension =50

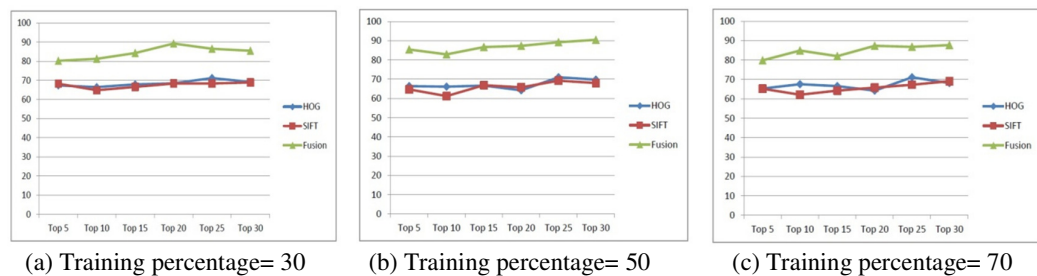


Figure 4. Experimental results of Linear Discriminant Analysis with dimension =50

REFERENCES

- [1] T. Zhang, Q. Han, H. Hou, and X. Niu, "Local invariant shape feature for cartoon image retrieval," in Robot, Vision and Signal Processing(RVSP), 2013 Second International Conference on. IEEE, 2013, pp. 107–110.
- [2] M. Haseyama and A. Matsumura, "A trainable retrieval system for cartoon character images," in Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on, vol. 3. IEEE, 2003, pp. III–673.
- [3] J. Yu, D. Liu, D. Tao, and H. S. Seah, "On combining multiple features for cartoon character retrieval and clip synthesis," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 42, no. 5, pp. 1413–1427, 2012.
- [4] Y. Yang, Y. Zhuang, D. Xu, Y. Pan, D. Tao, and S. Maybank, "Retrieval based interactive cartoon synthesis via unsupervised bi-distance metric learning," in Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009, pp. 311–320.
- [5] Y. Yang, Y. Zhuang, D. Tao, D. Xu, J. Yu, and J. Luo, "Recognizing cartoon image gestures for retrieval and interactive cartoon clip synthesis," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 20, no. 12, pp. 1745–1756, 2010.
- [6] T. Lindeberg, "Scale invariant feature transform," Scholarpedia, vol. 7, no. 5, p. 10491, 2012.
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, 2005, pp. 886–893.
- [8] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, no. 4, pp. 433–459, 2010.
- [9] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," Institute for Signal and information Processing, 1998.
- [10] A. J. Izenman, "Linear discriminant analysis," in Modern Multivariate Statistical Techniques. Springer, 2008, pp. 237–280.
- [11] U. Jayaraman, S. Prakash, and P. Gupta, "Indexing multimodal biometric databases using kd-tree with feature level fusion," in Information Systems Security. Springer, 2008, pp. 221–234.
- [12] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 12, pp. 1349–1380, 2000.