A Review on Image Feature Extraction and Representation Techniques

Dong ping Tian[1,2]

[1] *Institute of Computer Software, Baoji University of Arts and Sciences, Baoji, Shaanxi, 721007, China*
[2] *Institute of Computational Information Science, Baoji University of Arts and Sciences, Baoji, Shaanxi, 721007, China*

{tdp211}@163.com

Abstract

Feature extraction and representation is a crucial step for multimedia processing. How to extract ideal features that can reflect the intrinsic content of the images as complete as possible is still a challenging problem in computer vision. However, very little research has paid attention to this problem in the last decades. So in this paper, we focus our review on the latest development in image feature extraction and provide a comprehensive survey on image feature representation techniques. In particular, we analyze the effectiveness of the fusion of global and local features in automatic image annotation and content based image retrieval community, including some classic models and their illustrations in the literature. Finally, we summarize this paper with some important conclusions and point out the future potential research directions.

Keywords: *Feature Extraction, Feature Representation, Global Feature, Local Feature, Bag-of-Visual-Words*

1. Introduction

"A picture is worth a thousand words." As human beings, we are able to tell a story from a picture based on what we see and our background knowledge. Can a computer program discover semantic concepts from images? The short answer is yes. The first step for a computer program in semantic understanding, however, is to extract efficient and effective visual features and build models from them rather than human background knowledge. So we can see that how to extract image low-level visual features and what kind of features will be extracted play a crucial role in various tasks of image processing. As we known, the most common visual features include color, texture and shape, *etc.* [1-9], and most image annotation and retrieval systems have been constructed based on these features. However, their performance is heavily dependent on the use of image features. In general, there are three feature representation methods, which are global, block-based, and region-based features. Chow *et al.*, [10] present an image classification approach through a tree-structured feature set, in which the root node denotes the whole image features while the child nodes represent the local region-based features. Tsai and Lin [11] compare various combinations of image feature representation involving the global, local block-based and region-based features for image database categorization. In addition, a block-based image feature representation is proposed by Lu [12] in order to reflect the spatial features for a specific concept. However, little atten- tion has been paid to image feature extraction compared to a significant amount of research on annotation /retrieval model itself construction. Therefore, in this paper, we focus our review on the latest development in image feature extraction, especially the way for image feature extraction techniques so as to complement the existing surveys in literature.

The rest of the paper is organized as follows. Section 2 elaborates the most common image visual features, including their characteristics and some classic applications in the literature. In Section 3, the methods of image feature representation are summarized. In particular, fusion of global and local features in image processing community is elaborated. Finally, some important conclusions and future potential research directions are proposed in Section 4.

2. Image Feature Extraction

2.1. Color features

Color is one of the most important features of images. Color features are defined subject to a particular color space or model. A number of color spaces have been used in literature, such as RGB, LUV, HSV and HMMD [2]. Once the color space is specified, color feature can be extracted from images or regions. A number of important color features have been proposed in the literatures, including color histogram [13], color moments(CM) [14], color coherence vector (CCV) [15] and color correlogram [16], *etc.* Among them, CM is one of the simplest yet very effective features. The common moments are mean, standard deviation and skewness, the corresponding calculation can be defined as follows:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \tag{1}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij}-\mu_i\right)^2\right)^{\frac{1}{2}} \tag{2}$$

$$\gamma_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij}-\mu_i\right)^3\right)^{\frac{1}{3}} \tag{3}$$

where f_{ij} is the color value of the i-*th* color component of the j-*th* image pixel and N is the total number of pixels in the image. μ_i, σ_i, γ_i (i=1,2,3) denote the mean, standard deviation and skewness of each channel of an image respectively.

Table 1 provides a summary of different color methods excerpted from the literature [17], including their strengths and weaknesses. Note that DCD, CSD and SCD denote the dominant color descriptor, color structure descriptor and scalable color descriptor respectively. For more details of them, please refer to reference [17] and the corresponding original papers.

Table 1. Contrast of different color descriptors

Color method	Pros.	Cons.
Histogram	Simple to compute, intuitive	High dimension, no spatial info, sensitive to noise
CM	Compact, robust	Not enough to describe all colors, no spatial info
CCV	Spatial info	High dimension, high computation cost
Correlogram	Spatial info	Very high computation cost, sensitive to noise, rotation and scale
DCD	Compact, robust, perceptual meaning	Need post-processing for spatial info
CSD	Spatial info	Sensitive to noise, rotation and scale
SCD	Compact on need, scalability	No spatial info, less accurate if compact

2.2. Texture features

Texture is a very useful characterization for a wide range of image. It is generally believed that human visual systems use texture for recognition and interpretation. In general, color is usually a pixel property while texture can only be measured from a group of pixels. A large number of techniques have been proposed to extract texture features. Based on the domain from which the texture feature is extracted, they can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods. For the former approach, texture features are extracted by computing the pixel statistics or finding the local pixel structures in original image domain, whereas the latter transforms an image into frequency domain and then calculates feature from the transformed image. Both spatial and spectral features have advantage and disadvantages. Table 2 summarizes their pros. and cons.

Table 2. Contrast of texture features

Texture method	Pros.	Cons.
Spatial texture	Meaningful, easy to understand, can be extracted from any shape without losing info.	Sensitive to noise and distortions
Spectral texture	Robust, need less computation	No semantic meaning, need square image regions with sufficient size

As the most common method for texture feature extraction, Gabor filter [18] has been widely used in image texture feature extraction. To be specific, Gabor filter is designed to sample the entire frequency domain of an image by characterizing the center frequency and orientation parameters. The image is filtered with a bank of Gabor filters or Gabor wavelets of different preferred spatial frequencies and orientations. Each wavelet captures energy at a specific frequency and direction which provide a localized frequency as a feature vector. Thus, texture features can be extracted from this group of energy distributions [19]. Given an input image $I(x,y)$, Gabor wavelet transform convolves $I(x,y)$ with a set of Gabor filters of different spatial frequencies and orientations. A two-dimensional Gabor function $g(x,y)$ can be defined as follows.

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jW_x\right] \tag{4}$$

where σ_x and σ_y are the scaling parameters of the filter (the standard deviations of the Gaussian envelopes), W is the center frequency, and θ determines the orientation of the filter. Figure 1 shows the Gabor function in the spatial domain.

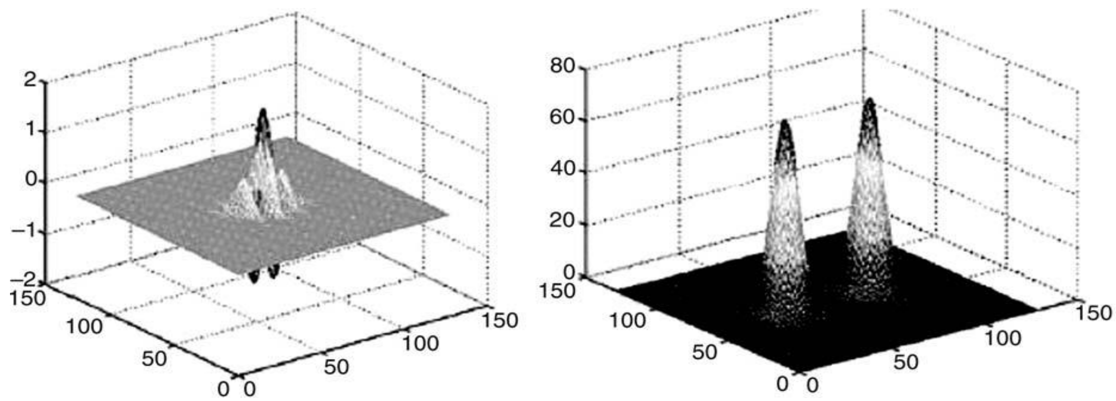

Figure 1. Gabor function in the spatial domain

2.3. Shape features

Shape is known as an important cue for human beings to identify and recognize the real-world objects, whose purpose is to encode simple geometrical forms such as straight lines in different directions. Shape feature extraction techniques can be broadly classified into two groups [20], viz., contour based and region based methods. The former calculates shape features only from the boundary of the shape, while the latter method extracts features from the entire region. For more details of image shape feature extraction and representation, please refer to the literature [20].

In addition, spatial relationship is also considered in image processing, which can tell object location within an image or the relationships between objects. It mainly includes two cases: absolute spatial location of regions [21] and relative locations of regions [22, 23]. Figure 2 shows an example of a 2D string representation. The image in Figure 2(a) is decomposed into regions (blocks). For simplicity, the block identifiers are used as object symbols. Two relationship symbols '<' and '=' are used in this case. In horizontal and vertical directions, the symbol '<' denotes 'left-right' and 'below-above' relationships respectively. The symbol '=' means the spatial relationship 'at the same spatial location as'. A 2D string takes the form (u,v), where u and v are the relationships of objects in horizontal and vertical directions respectively. Figure 2(d) shows the 2D string for the image of Figure 2(a).

d		
	b	c
a	a	

a

Object Symbols	a, b, c, d

b

Relationship symbols	Meaning
<	left-right or below-above
=	at the same spatial location as

c

$(a = d < a = b < c \ , \ a = a < b = c < d)$ d

Figure 2. Illustration of a 2D string:(a) an image decomposed into blocks,(b) object symbols as block names,(c) definitions of relationship symbols, and (d) a 2D string for (a)

Alternatively, as a very good review literature for shape feature extraction, Yang *et al.*, [24] present a survey of the existing approaches of shape-based feature extraction. The following Figure 3 shows the hierarchy of the classification of shape feature extraction approaches excerp- ted from the corresponding literature.

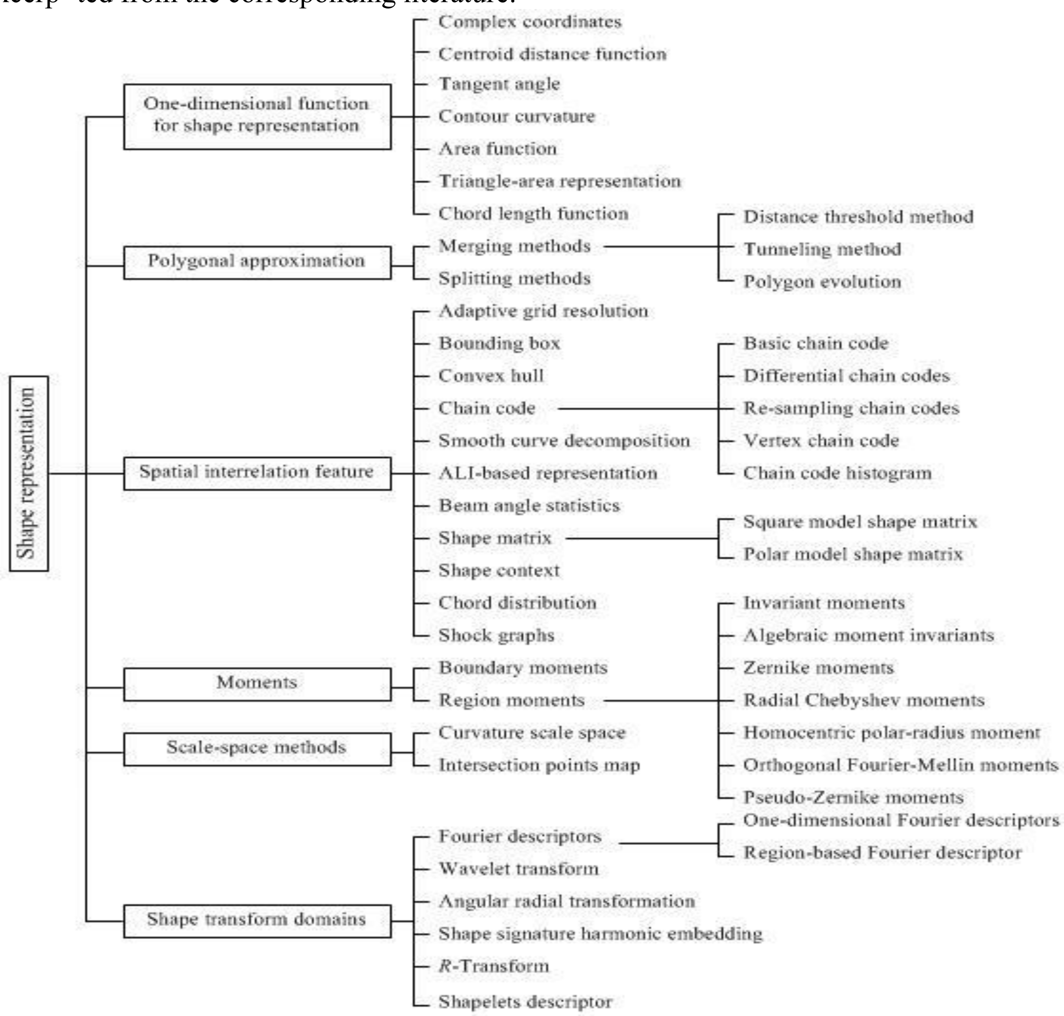

Figure 3. An overview of shape description techniques

3. Image Feature Representation

Besides the image features briefly reviewed above, how to partition an image and how to organize the image features are also challenging problems. In general, there are mainly three methods to transform an image into a set of regions: regular grid approach, unsupervised image segmentation and interest point detectors. Figure 4 illustrates the regions obtained by these three methods, where (a) shows the original image, (b) gives the image segmented by a regular grid, (c) provides the image segmented by the JSEG [25], and (d) outlines the salient regions detected by the Difference of Gaussian (DoG) [26] detector .

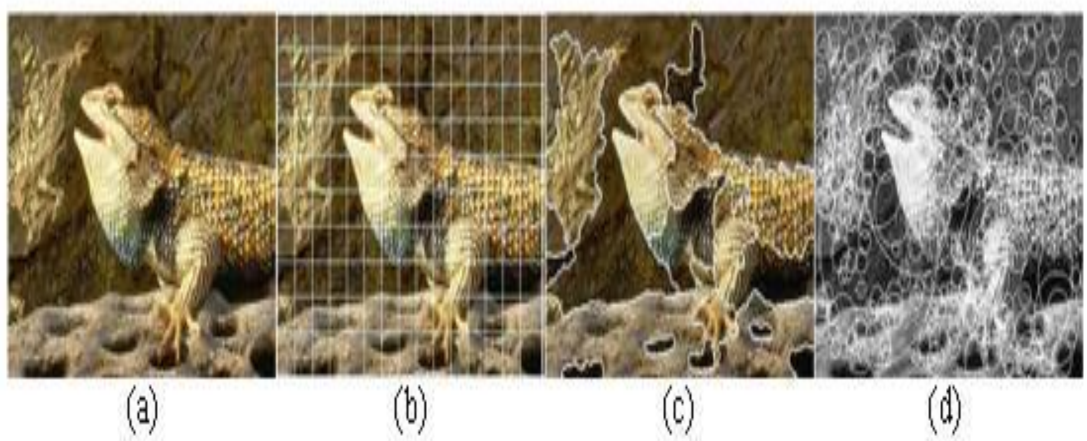

Figure 4. Three different approaches to transform an image into a set of regions: (a) original image, (b) image segmented by a regular grid, (c) the image segmented by the JSEG[25],and (d) salient regions detected by the DoG [26] detector

As the representative work of using both global and local image features, Chow *et al.*, [10] utilize a two-level tree to integrate both global and local image features for image classification, in which the child nodes of the tree contain the region-based local features, while the root node contains the global features. The following Figure 5 shows the representation of image contents by integrating global features and local region-based features, where (a) shows a whole image, whose color histogram is extracted and served as the global feature at the root node, (b) illustrates six segmented regions of image (a), and color, texture, shape and size features are extracted from all the regions and acted as the region features at the child nodes, and (c) depicts the tree representation of the image.

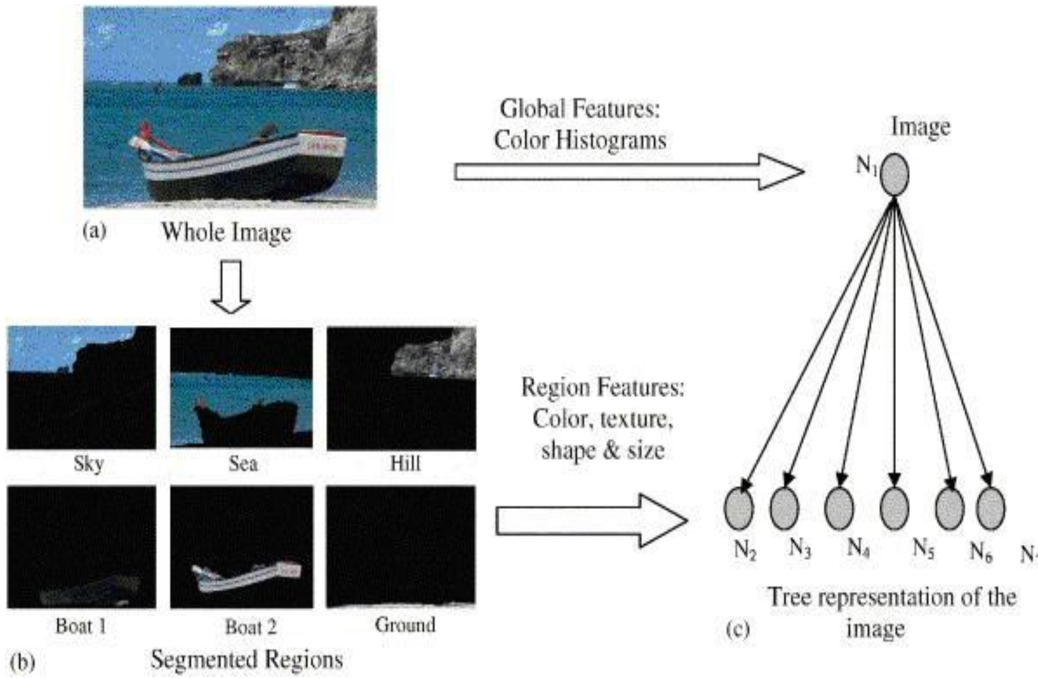

Figure 5. Representation of image contents by integrating global features and local region-based features (a) whole image (b) segmented regions, and (c) tree representation of the image

In the literature [11], Tsai and Lin compare various combinations of feature representation methods including the global and local block-based and region-based features for image database categorization. Then the significant conclusion, i.e. the combined global and block-based feature representation performs the best, is drawn in the end. Zhu *et al.*, [27] believe that an appropriate fusion of global and local features will compensate their shortcomings, and therefore improve the overall effectiveness and efficiency. Thus they consider grid color moment, LBP, Gabor wavelets texture and edge orientation histogram as image global features, while SURF descriptor is employed to extract image local features. More recently, Tian *et al.*, [28] present a combined global and local block-based image feature representation method so as to reflect the intrinsic content of images as complete as possible, in which the color histogram in HSV space is extracted to represent the global feature of images, and color moments, Gabor wavelets texture and Sobel shape detector are used to extract local features. Here, shape feature can be extracted by the convolution of 3×3 masks with the image in 4 different directions (horizontal, 45°, vertical and 135°). Finally, they combine the global feature and local features, *i.e.*, features of the blocks connected by left-to-right and top-to-down orders together, which results in a so-called block-line feature structure. Table 3 summarizes the global and local features employed in these references as below.

Table 3. Contrast of global and local feature extraction

Sources	Global features adopted	Local features adopted
Chow et al. [10]	Color histogram in HSV space	Color moments, Gabor texture, shape and size
Tsai et al.[11]	Color moment in HSV space, four levels of Daubechies-4 wavelet decomposition	Color moment in HSV space, four levels of Daubechies-4 wavelet decomposition
Zhu et al.[27]	Grid color moment, LBP, Gabor wavelets texture and edge orientation histogram	SURF
Tian et al.[28]	Color histogram in HSV space	Color moments in HSV space, Gabor wavelets texture and Sobel shape
Lisin et al. [29]	LBP and shape index	SIFT
Zhao et al.[30]	Pseudo Zernike moments	SIFT

In addition, Zhou *et al.*, [31] propose a joint appearance and locality image representation called hierarchical Gaussianization(HG), which adopts a Gaussian mixture model (GMM) for appearance information and a Gaussian map for locality information. The basic procedure of HG can be succinctly described as follows:

- Extract patch feature, *e.g.*, SIFT descriptor from overlapping patches in the images.
- From the images of interest, generate a universal background model (UBM) that is a Gaussian mixture model (GMM) describing the patches from this set of images.
- For each image, adapt the UBM to obtain another GMM to describe patch feature distribution within the image.
- Characterize the GMM using its component means and variances as well as a Gaussian map which contains certain patch location information.
- Perform a supervised dimension reduction, named discriminant attribute projection (DAP), to eliminate within-class feature variation.

Figure 6 illustrates the procedure for generating a HG representation of an image.

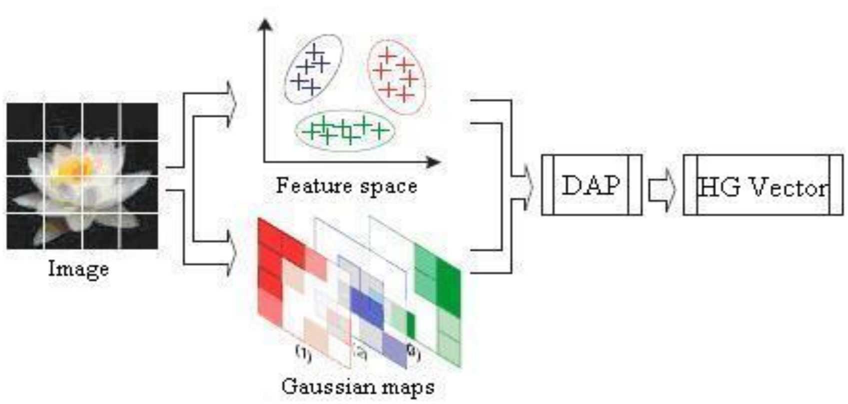

Figure 6. Procedure for generating a HG representation of an image

Last but not the least, bag of visual words representation has been widely used in image annotation and retrieval [32, 33]. This visual-word image representation is analogous to the bag-of-words representation of text documents in terms of form and semantics. The procedure of generating bag-of-visual-words can be succinctly described as follows. First, region features are extracted by partitioning an image into blocks or segmenting an image into regions. Second, clustering and discretizing these features into visual word that represents a specific local pattern shared by the patches in that cluster. Third, mapping the patches to visual words and then we can represent each image as a bag-of-visual-words. Compared to previous work, Yang *et al.*, [34] have thoroughly studied the bag-of-visual-words from the choice of dimension, selection, and weighting of visual words in this representation. For more detailed information, please refer to the corresponding literature. Figure 7 illustrates the basic procedure of generating visual-word image representation based on vector-quantized region features.

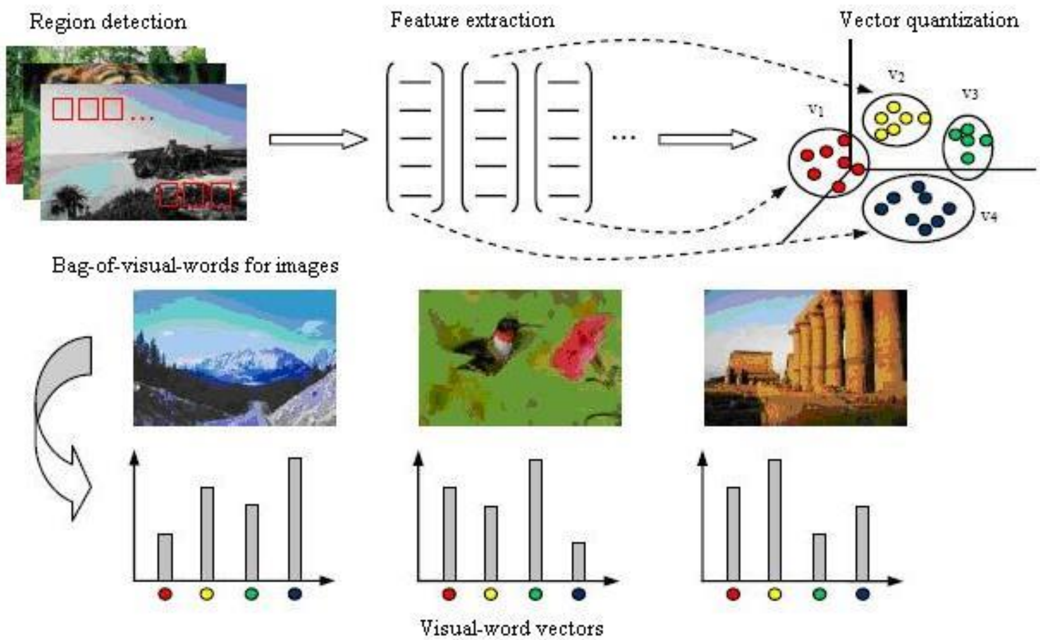

Figure 7. Procedure of generating visual-word image representation based on vector-quantized region features

4. Conclusion and Future Work

As few previous studies review both image feature extraction and image feature representation, which play a crucial role in multimedia processing community. So in this paper, we provide a comprehensive survey on the latest development in image feature extraction and image feature representation. Particularly, we analyze the effectiveness of the fusion of global and local features in image processing, including some classic models and their illustrations in the literature. Followed by another type of feature representation, i.e., the bag-of-visual-word is elaborated.

In conclusion, there are a number of interesting issues which should be considered as future work. First, it is worth exploring the relationship between features' number and the final performance. Intuitively, it is not possible that the more the features' number, the better the final performance. Second, to explore the relationship between features' representation and the final performance is also a very interesting and challenging topic. It involves the feature representation methods (global, block-based and region-based features). Specifically, in the case of block-based and region-based features, the final performance partially depends on the size of the partition or segmentation. Third, it is also interesting to explore the relationship between their appropriate combination and the final performance to see whether the combination can further improve the performance.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.61035003, No.61072085, No.60933004, No.60903141) and the National Program on Key Basic Research Project (973 Program) (No.2013CB329502).

References

[1] T. K. Shih, J. Y. Huang, C. S. Wang, *et al.*, “An intelligent content-based image retrieval system based on color, shape and spatial relations”, In Proc. National Science Council, R. O.C., Part A: Physical Science and Engineering, vol. 25, no. 4, **(2001)**, pp. 232-243.
[2] P. L. Stanchev, D. Green Jr. and B. Dimitrov. “High level colour similarity retrieval”, International Journal of Information Theories and Applications, vol. 10, no. 3, **(2003)**, pp. 363-369.
[3] N. C. Yang, W. H. Chang, C. M. Kuo, *et al.*, “A fast MPEG-7 dominant colour extraction with new similarity measure for image retrieval”, Journal of Visual Comm. and Image Retrieval, vol. 19, **(2008)**, pp. 92-105.
[4] M. M. Islam, D. Zhang and G. Lu. “A geometric method to compute directionality features for texture images”, In Proc. ICME, **(2008)**, pp. 1521-1524.
[5] S. Arivazhagan and L. Ganesan, “Texture classification using wavelet transform”, Pattern Recognition Letters, vol. 24, **(2003)**, pp. 1513-1521.
[6] S. Li and S. Shawe-Taylor, “Comparison and fusion of multi-resolution features for texture classification”, Pattern Recognition Letters, vol. 26, no. 5, **(2005)**, pp. 633-638.
[7] W. H. Leung and T. Chen, “Trademark retrieval using contour-skeleton stroke classification”, In Proc. ICME, **(2002)**, pp. 517-520.
[8] Y. Liu, J. Zhang, D. Tjondronegoro, *et al.*, “A shape ontology framework for bird classification”, In Proc. DICTA, (2007), pp. 478-484.
[9] C. F. Tsai, “Image mining by spectral features: A case study of scenery image classification”, Expert Systems with Applications, vol. 32, no. 1, **(2007)**, pp. 135-142.
[10] T. W. S. Chow and M. K. M. Rahman, “A new image classification technique using tree-structured regional features”, Neurocomputing, vol. 70, no. 4-6, **(2007)**, pp. 1040-1050.
[11] C. F. Tsai and W. C. Lin, “A comparative study of global and local feature representations in image database categorization”, In Proc. 5th International Joint Conference on INC, IMS & IDC, **(2009)**, pp. 1563-1566.
[12] H. Lu, Y. B. Zheng, X. Xue, *et al.*, “Content and context-based multi-label image annotation”, In Proc. Workshop of CVPR, **(2009)**, pp. 61-68.
[13] A. K. Jain and A. Vailaya, “Image retrieval using colour and shape”, Pattern Recognition, vol. 29, no. 8, **(1996)**, pp. 1233-1244.
[14] M. Flickner, H. Sawhney, W. Niblack, *et al.*, “Query by image and video content: the QBIC system”, IEEE Computer, vol. 28, no. 9, **(1995)**, pp. 23-32.
[15] G. Pass and R. Zabith, “Histogram refinement for content-based image retrieval”, In Proc. Workshop on Applications of Computer Vision, **(1996)**, pp. 96-102.
[16] J. Huang, S. Kuamr, M. Mitra, *et al.*, “Image indexing using colour correlogram”, In Proc. CVPR, **(1997)**, pp. 762-765.
[17] D. S. Zhang, Md. M. Islam and G. J. Lu, “A review on automatic image annotation techniques”, Pattern Recognition, vol. 45, no. 1, **(2012)**, pp. 346-362.
[18] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of large image data”, IEEE PAMI, vol. 18, no. 8, **(1996)**, pp. 837-842.
[19] S. E. Grigorescu, N. Petkov and P. Kruizinga, “Comparison of texture features based on Gabor filters”, IEEE TIP, vol. 11, no. 10, **(2002)**, pp. 1160-1167.
[20] D. Zhang and G. Lu, “Review of shape representation and description techniques”, Pattern Recognition, vol. 37, no. 1, **(2004)**, pp. 1-19.
[21] C. Yang, M. Dong and F. Fotouhi, “Image content annotation using Bayesian framework and complement components analysis”, In Proc. ICIP, **(2005)**.

[22] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "An ontology approach to object-based image retrieval", In Proc. ICIP, **(2003)**, pp. 511-514.
[23] D. Zhang, M. M. Islam, G. Lu, *et al.*, "Semantic image retrieval using region based inverted file", In Proc. DICTA, **(2009)**, pp. 242-249.
[24] M. Yang, K. Kpalma and J. Ronsin. "A survey of shape feature extraction techniques", Pattern Recognition, **(2008)**, pp. 43-90.
[25] Y. N. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video", IEEE PAMI, vol. 23, no. 8, **(2001)**, pp. 800-810.
[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, **(2004)**, pp. 91-110.
[27] J. Zhu, S. Hoi, M. Lyu, *et al.*, "Near-duplicate keyframe retrieval by nonrigid image matching", In Proc. ACM MM, **(2008)**, pp. 41-50.
[28] D. Tian, X. F. Zhao and Z. Shi, "Support vector machine with mixture of kernels for image classification", In Proc. ICIIP, **(2012)**, pp. 67-75.
[29] D. A. Lisin, M. A. Mattar, M. B. Blaschko, *et al.*, "Combining local and global image features for object class recognition", In Proc. Workshop on CVPR, **(2005)**.
[30] J. Zhao, Y. Fan and W. Fan, "Fusion of global and local feature using KCCA for automatic target recognition", In Proc. ICIG, **(2009)**, pp. 958-962.
[31] X. Zhou, N. Cui, Z. Li, *et al.*, "Hierarchical gaussianization for image classification", In Proc. ICCV, **(2009)**, pp. 1971-1977.
[32] F. Monay and D. Gatica-Perez, "Modeling semantic aspects for cross-media image indexing", IEEE PAMI, vol. 29, no. 10, **(2007)**, pp. 1802-1817.
[33] D. Tian, X. Zhao and Z. Shi, "Refining image annotation by integrating PLSA with random walk model", In Proc. MMM, Part I, LNCS 7732, **(2013)**, pp. 13-23.
[34] J. Yang, Y.Jiang, A. G. Hauptmann and C. Ngo, "Evaluating bag-of-visual-words representations in scene classification", In Proc. Workshop on MIR, **(2007)**, pp. 197-206.

Authors

Dong ping Tian

He received his M.Sc. and Ph.D. degrees from Shanghai Normal University and Institute of Computing Technology (ICT), Chinese Academy of Sciences in 2007 and 2013 respectively. Now he is an associate professor in Baoji University of Arts and Sciences. His main research interests include computer vision, machine learning and evolutionary computation.