Chapter 2
Taxonomy

Abstract In Chap. 1, we introduced the concept of feature coding as well as its role in the Bag-of-Features model. In this chapter, we will discuss the taxonomy of feature coding. First intuitively describe several classic feature coding algorithms, and then categorize them in two ways, namely **Taxonomy I** according to their final representations and **Taxonomy II** according to their original motivations. Taxonomy I involves the number of codewords for describing a feature and the dimensions for the coding response on a codeword. This way is easy for readers to quickly know about the algorithms, especially their main steps. The second way is for an in-depth understanding of feature coding. According to their motivations, most feature coding algorithms belong to one of the five main categories: (1) Voting based coding; (2) Fisher coding; (3) Reconstruction based coding; (4) Local tangent based coding; and (5) Saliency based coding. This chapter is closely related to the content in the following chapters, e.g. the formulation, motivations and relationships of various feature coding algorithms, as well as how they evolve.

2.1 Taxonomy Based on Representation

There are two main rules to decide the final representation of feature coding. One is how many codewords are used in encoding a feature. The other is how many dimensions of the response are maintained on each codeword in encoding a feature. The above two rules are illustrated in Fig. 2.1, based on which we can group most feature coding algorithms into four categories as shown in Fig. 2.2.

In this chapter, we only give a simple and intuitive description for each feature coding algorithm appearing in Fig. 2.2, and leave their mathematic formulation in Chap. 3. This manner of explanation, we believe, is an better way for readers to gradually understand feature coding.

- In **hard voting** (HV) [1], a feature is encoded by its nearest codeword with response of 1 and other codewords with responses of 0.
- In **salient coding** [2], a feature is encoded by its nearest codeword with response of the saliency degree, which can be defined as the ratio between the *distance* of the feature to its nearest codeword and the *average distance* of the feature to other

Y. Huang and T. Tan, *Feature Coding for Image Representation and Recognition*,
SpringerBriefs in Computer Science, DOI 10.1007/978-3-662-45000-0_2

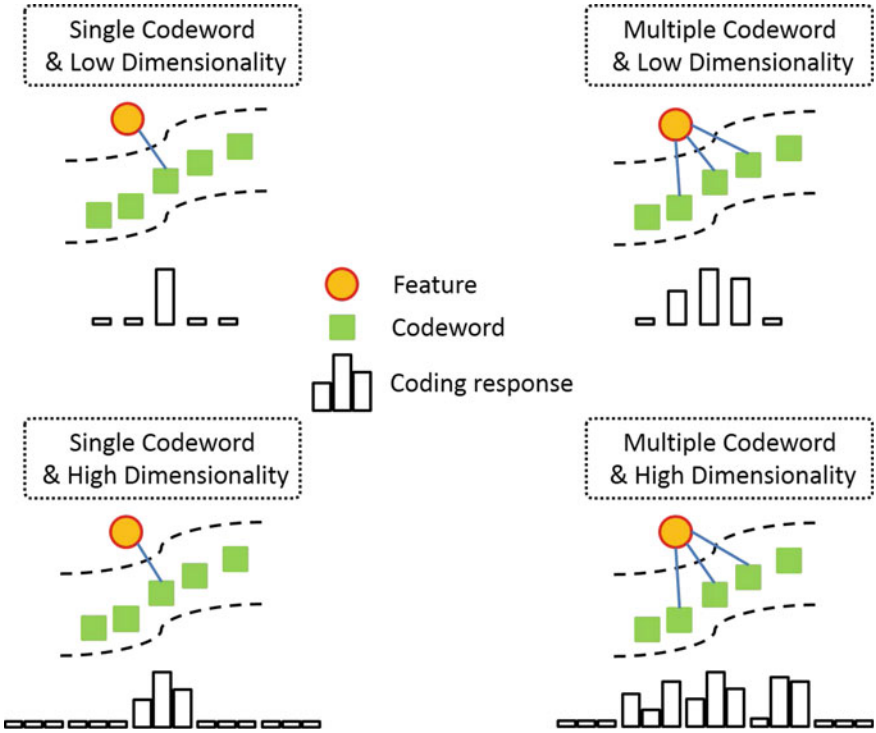

Fig. 2.1 Two rules to classify different feature coding algorithms

codewords. Here, 'other codewords' means a set of near codewords except the nearest one.

- In **soft voting** (SV) [3], a feature is encoded by its several nearest codewords. The response on each codeword is a function of the distance between the feature and its codeword. Although the function is not fixed (depending on practical applications), it is basically inversely proportional to the distance, which aims to enhance the importance of near codewords and weaken the influence of far codewords. The number of codewords is usually a predefined parameter.
- In **sparse coding** (SC) [4], a feature is encoded by a set of codewords that best reconstruct the feature with a least-square optimization plus a sparse constraint on the codewords used to encode the feature. The responses on the codewords are the coefficients produced in solving the least-square optimization problem.
- **Local coordinate coding** (LCC) [5] is similar to sparse coding. The difference lies in the constraint wherein the distance between the feature and the codewords is used as a penalty whose motivation is similar to that of soft voting. Thus, compared with sparse coding, it adds a distance constraint.
- **Local-constraint linear coding** (LLC) [6] is derived from local coordinate coding. The distance and the sparse constraint are smartly replaced by using the several

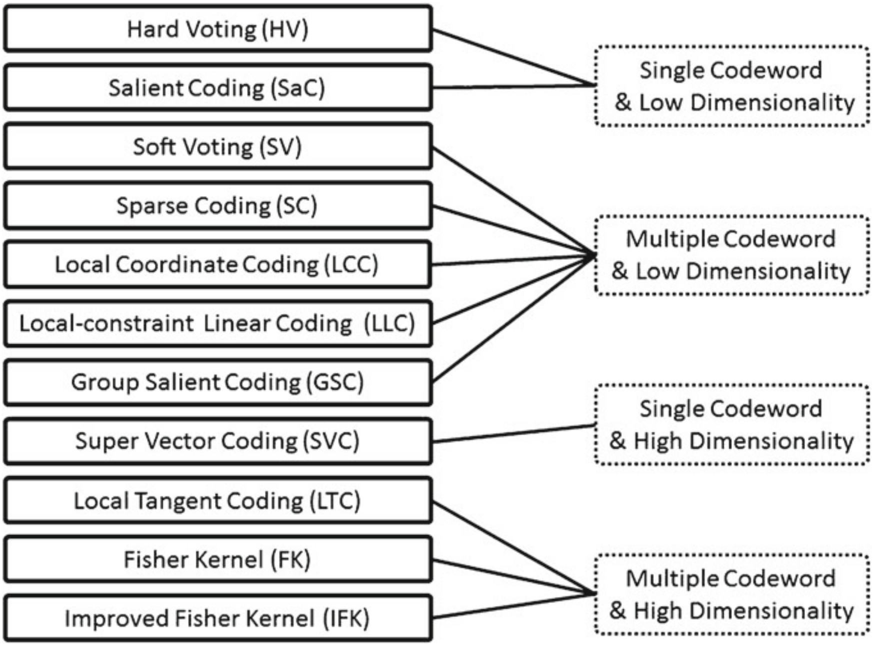

Fig. 2.2 Taxonomy of feature coding algorithms based on their final representation

nearest codewords to reconstruct the feature. In particular, the sparse constraint is implemented by choosing a small number of codewords. In this manner, the distance constraint is also satisfied due to choosing the nearest codewords.

- **Group salient coding** (GSC) [7] is the extension of salient coding in the sense that it chooses a number of codewords for feature coding. The salient degree is calculated for a group of codewords, which are then fed back to all the codewords in the group. The final coding response of the feature on a codeword is the maximum of all responses calculated according to different group sizes of codewords.
- In **Super vector coding** (SVC) [8], a feature is encoded by its nearest codeword with the response of the distance between the feature and this codeword, preserving the distance in all dimensions of the feature space. That's why we describe this coding method as "single codeword and high dimensionality".
- **Local tangent coding** (LTC) [9] is a little complicated to explain without mathematic equations, and we leave it to Chap. 3.
- In **Fisher kernel** (FK) [10], each codeword is the center of Gaussian mixture models (GMM) which is constructed by all features. A feature is encoded by several codewords, and the response on each codeword consists of three functions. One is the derivation from the log likelihood of features to the mean vector of the Gaussian distribution of the codeword. The second one is the derivation from the log likelihood of features to the covariance matrix of the Gaussian distribution of

the codeword. The third one is the derivation from the log likelihood of features to the weight of the Gaussian distribution of the codeword. In improved Fisher kernel, the third function is removed because it is found that this term makes little contribution to the performance.

With the explanation above, readers may not be clear enough about some feature coding algorithms due to its mathematical complexity, which will be detailed in Chap. 3. This chapter is only to give a rough impression of feature coding.

2.2 Taxonomy Based on Motivation

According to their motivations, the existing coding strategies can be divided into five major categories, as shown in Fig. 2.3.

Global coding is generally designed to estimate the PDF of features. It focuses on the global description of all features rather than each individual feature. There are mainly two kinds of strategies in global coding:

- **Voting-based methods** [1, 3] describe the distribution of features with a histogram, which carries the occurrence information of codewords. Such a histogram is usually constructed by hard quantization or soft quantization.
- **Fisher coding-based methods** [10, 12] estimate the distribution of features with the Gaussian Mixture Models, consisting of the weights, the means and the covariance matrix of multiple Gaussian distributions, each of which reflects one pattern of features.

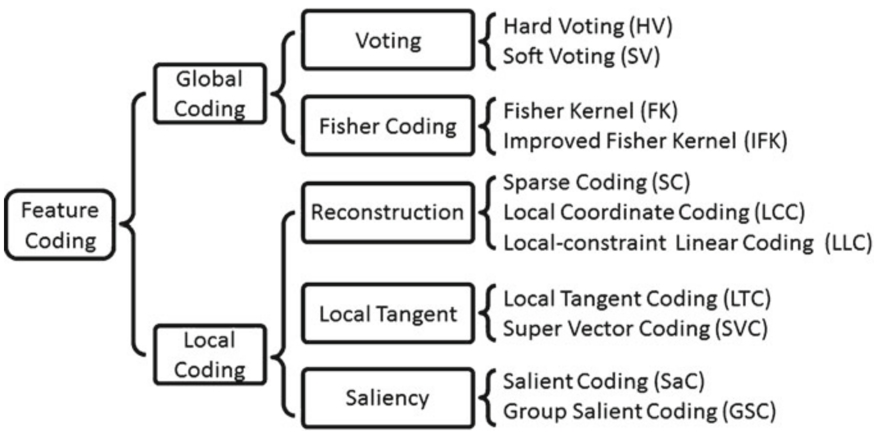

Fig. 2.3 Taxonomy of feature coding methods according to their motivations. Several representative algorithms are listed for each category. ©[2014] IEEE. Reprinted, with permission, from [11]

Local coding is proposed to describe each individual feature. Three kinds of local coding methods have been developed:

- **Reconstruction-based methods** [4–6] use a small part of codewords to describe each feature via solving a least-square-based optimization problem with constraints on codewords.
- **Local tangent-based methods** [8, 9] derive an exact description for each feature through approximating the Lipschitz smooth manifold where features are located.
- **Saliency-based methods** [2, 7] encode each feature by the saliency degree, which is calculated using the ratio or the difference of the distances from a feature to its nearby codewords.

It should be noted that the use of the concepts "global" and "local" here is to keep consistent with the motivations presented in the original papers. Discussion on the formulations and relations of these five kinds of feature coding methods will be detailed in Chaps. 3 and 4, respectively.

References

1. G. Csurka, C. Bray, C. Dance, L. Fan, Visual categorization with bags of keypoints, in *European Conference on Computer Vision* (2004)
2. Y. Huang, K. Huang, Y. Yu, T. Tan, Salient coding for image classification, in *IEEE Conference on Computer Vision and Pattern Recognition* (2011)
3. J. Gemert, J. Geusebroek, C. Veenman, A. Smeulders, Kernel codebooks for scene categorization, in *European Conference on Computer Vision* (2008)
4. J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in *IEEE Conference on Computer Vision and Pattern Recognition* (2009)
5. K. Yu, T. Wang, Y. Gong, Nonlinear learning using local coordinate coding. *Advances in Neural Information Processing Systems* (2009)
6. J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in *IEEE Conference on Computer Vision and Pattern Recognition* (2010)
7. Z. Wu, Y. Huang, L. Wang, T. Tan, Group encoding of local features in image classification, in *International Conference on Pattern Recognition* (2012)
8. Xi. Zhou, K. Yu, T. Zhang, T. Huang, Image classification using super-vector coding of local image descriptors, in *European Conference on Computer Vision* (2010)
9. K. Yu, T. Zhang, Improved local coordinate coding using local tangents, in *International Conference on Machine Learning* (2010)
10. F. Perronnin, C. Dance, Fisher Kernels on visual vocabularies for image categorization, in *IEEE Conference on Computer Vision and Pattern Recognition* (2007)
11. Y. Huang, Z. Wu, L. Wang, T. Tan, Feature coding in image classification: a comprehensive study. IEEE Trans. Pattern Anal. Mach. Intell. **36**(3), 493–506 (2014)
12. F. Perronnin, J. Sanchez, T. Mensink, Improving the Fisher Kernel for large-scale image classification, in *European Conference on Computer Vision* (2010)

http://www.springer.com/978-3-662-44999-8

Feature Coding for Image Representation and Recognition
Huang, Y.; Tan, T.
2014, XIII, 74 p. 36 illus., 32 illus. in color., Softcover
ISBN: 978-3-662-44999-8