

Feature-Based Facial Image Coding Method Using Wavelet Transform

Chuxi Yang¹, Yan Zhao^{1(✉)}, Shigang Wang¹, and Yuanyuan Liu²

¹ Communication Engineering, Jilin University, Changchun 130012, China
zhao_y@jlu.edu.cn

² Information Technology, Jilin Agricultural University, Changchun 130118,
Jilin, China

Abstract. This paper proposed a novel image coding scheme for facial images in the transform domain based on Scale Invariant Feature Transform (SIFT). With the characteristic of rotation scale invariability, SIFT descriptors are perfect to describe the details in an image. Our proposed coding scheme combine wavelet transform with SIFT algorithm. High frequency sub band images of the input image are encoded according to SIFT key points rather than pixel value on the encoder side. On the decoder side, we can get restoring patch from an image database according to key point matching. After clustering and transformation, the matching patches are used to restore the high frequency sub band images. Final reconstructed image was produced by wavelet inverse transform from low frequency sub band image and restored high frequency sub band images. Experimental results show that compared with HEVC and JPEG2000, the proposed algorithm can reach a higher compression ratio and a satisfactory reconstruction performance.

Keywords: Image coding · Wavelet transform · SIFT · Local feature Database

1 Introduction

Now Internet is an inseparable part of our daily life. Almost every day, we get news from the news site, search information through the search engine and watch some pictures and videos on the Internet. In recent years, Internet, as a large scale information carrier, has a rapid-growing image database. Facial images become an important part of all kinds of images and are widely used in various applications. Therefore effective facial image coding is vital for storage and transmission of facial images.

From the earliest standard JPEG to recently proposed HEVC/H.265, the development of mainstream image coding method took more than two decades. Some classical coding algorithms emerged, including JPEG2000 [1], AVC/H.264 and HEVC (High Efficiency Video Coding) [2]. H.264 uses intra-frame prediction in the spatial domain to remove redundancy between adjacent image blocks. JPEG2000 is a representative transform domain oriented coding method, which transforms the image into wavelet domain and then compresses it by quantizing the wavelet coefficients. Since cloud database is huge, it is very likely to find some similar images from the cloud database. For facial image, even

for different people, the face similarity can reach 70%. If we can use the database resource to help for image coding, we may get a higher compression ratio.

Currently, there are many visual feature extraction algorithms. For a facial image, we can extract some local features in the sense organs and hair areas. These local features can well reflect the portrait characteristics and can be invariant in some cases. Therefore we consider to encode facial images according to the local features, and try to keep the feature area and neglect the non-feature area appropriately.

In this paper, we propose a feature-based coding method for facial image in the transform domain. First of all, the input image is transformed by multi-layer wavelet transform. After transformation, we can get one low frequency sub band (LF sub band) and some high frequency sub bands (HF sub bands). We use local feature to describe the HF sub bands. The encoded bit stream is consists of encoded feature information and encoded LF sub band. On the decoded side, we use the decoded feature information to match features in a database and find similar patches to restore the HF sub bands, Final reconstructed image is obtained using wavelet inverse transform from decoded LF sub band and restored HF sub bands.

The rest of this paper is organized as follows: Sect. 2 introduces related works of proposed algorithm; Sect. 3 provides our coding method in detail; Sect. 4 shows our experimental results; Final conclusions are given in Sect. 5.

2 Related Works

In 2004, Lowe proposed SIFT algorithm [3]. SIFT features have the advantage of scale and rotation invariance. A scale space is established firstly, and extreme values are found in the scale space as feature key points in SIFT algorithm. There are three basic parameters for each key point: position, scale and orientation. Scale expresses the layer of the key points, and orientation gives the gradient orientation value of the neighborhood of the key point. The scale invariance of the SIFT feature can be used for matches between images at different scale, and rotation invariance can be used to adapt to the rotation transformation.

In this paper, we use SIFT algorithm to extract features of facial images. Each key point can be expressed by:

$$K_i = (Pt_i, s_i, o_i) \quad (1)$$

where Pt_i is the coordinate of the key point, s_i and o_i are the scale and orientation of the key point respectively.

After the feature extraction, the feature key points need to be described. In [3], Lowe used a 128-dimension descriptor to described each key point. According to the location of the key point, a neighborhood image block is selected, whose size is related to the scale of the key point. To realize the rotation invariance, the image block is rotated to its main orientation o_i . Then the block is divided into 4×4 subblocks, and a 8-dimension vector is calculated for each subblock. Finally, a vector with $4 \times 4 \times 8 = 128$ dimensions is obtained. We can get the SIFT descriptor of one each point after normalization, which is expressed by D_i .

SIFT algorithm had immediately attracted a wide spread attention when proposed. It was widely used in image reconstruction [4], image classification [5], image detection and recognition [6, 7] and image search [8, 9]. The method of image reconstruction based on feature was first proposed in [4]. Weinzaepfel et al. extracted image features by SIFT algorithm, then did the sift-match with a big image database. However, the reconstructed image had a serious distortion. Based on this idea, Yue et al. proposed a cloud-based image coding algorithm in [10]. A down sampled image was transferred to the decoder side as a guide of image reconstruction. At the decoder side, matching patches could be found from the cloud according to the feature match. Patches was pasted to the up sampled image after processing. This method can achieve high compression ratio, as well as good restoration quality. However, as it said in [10], when the high similarity image was removed from the database, the reconstructed image becomes blurred. In 2014, Song et al. proposed an image reconstruction method based on thumbnail [11], which got better performance in compression ratio and PSNR comparing with JPEG.

3 The Proposed Coding Method

3.1 General Analysis

On the encoder side, the image is multi-level wavelet transformed first. After wavelet transformation, *LF* sub band and *HF* sub bands are produced. *LF* sub band is transmitted to the decoder side after being encoded. For *HF* sub bands, we just remain some features of *HF* sub bands, and transmit the feature key point information to the decoder side. The key points of *HF* sub bands are got from the difference set between key points of input image and key points of HF-missed inverse wavelet transform image. The coded bit stream is consist of coded *LF* sub band L_{en} and coded key point information K_{en} .

In the decoder, L_{en} and K_{en} are first decoded. Then HF-missed wavelet reverse transformation is done to get an inverse wavelet transform image which is as large as the input image. Then the decoded key point information is used to calculate the feature descriptor in the inverse wavelet transform image. After we get the descriptors, we can do SIFT match to a large image database. Each matching key points in the database corresponds a small image patch. These small patches are used to restore the *HF* sub bands. When all the *HF* sub bands are restored, we can do HF-restored inverse wavelet transform to get the reconstructed image (Fig. 1).

3.2 Encoder

STEP 1 Wavelet transform

After the image wavelet transform, the input image can be decomposed into 4 sub band images: low frequency *LL*, and high frequency sub bands *HL*, *LH*, *HH*. The structure and color information are included in the *LF* sub band, while the *HF* sub bands include the texture and detail information. In order to obtain a higher compression ratio, low frequency sub band can continue to be decomposed. For instance, LL_{n-1} can be divided into LL_n , LH_n , HL_n and HH_n . In our image coding scheme, we use 2-level wavelet

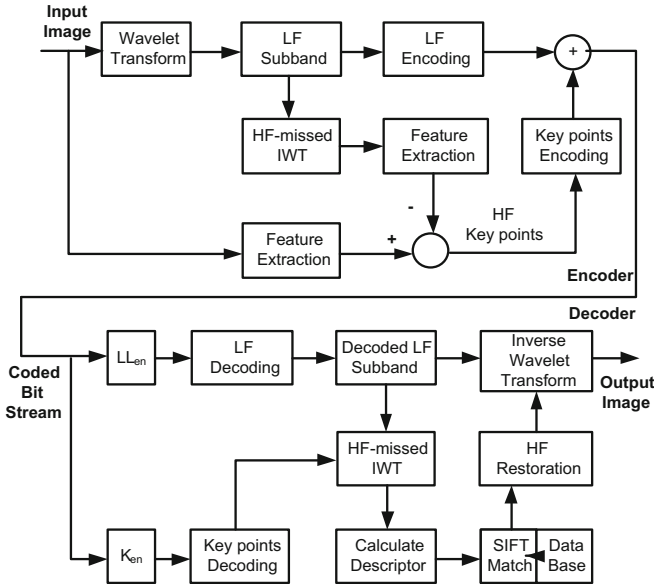


Fig. 1. System framework of proposed image coding scheme

transform, and totally produce 1 low frequency sub band LL_2 , and 6 high frequency sub bands ($HL_2, HH_2, HH_2, LH_1, HL_1, HH_1$). We use L and H to represent low and high frequency parts for short respectively.

STEP 2 HF key point extraction

The goal of this step is to find effective information in the HF sub bands. SIFT algorithm can well detect the edges, corners and some texture. So the SIFT key points can well describe the effective information in an image. Therefore, effective information in HF sub bands can be described by HF key points. To get all the HF key points information, we first extract SIFT key point from the input image, denoted by K_I . Meanwhile, using LL_2 to do a 2-level inverse wavelet transform with HF sub bands are all zero, Each inverse transform level can be expressed by:

$$I_{iwt} = \frac{1}{2} \sum_{j,k} WT_f(j,k) \cdot \psi_{j,k}(t) \tag{2}$$

where, $\psi_{j,k}(t)$ is the wavelet basis. The inverse transform image is denoted by I_L . Then key point from I_L is extracted and denoted by K_L . It is obviously that K_L contains only low frequency information. The difference set of K_I and K_L is the HF key points. The information of HF feature key points can be described as:

$$K_H = \{(pt_1, s_1, o_1), (pt_2, s_2, o_2), \dots, (pt_n, s_n, o_n)\} \tag{3}$$

where n is the number of key points.

STEP 3 LF sub band encoding

Since low frequency image includes most of the color and intensity information of the input image, we can use coding algorithm with low distortion to encode the *LF* sub band such as JPEG2000 and HEVC. In our coding scheme, we use HEVC to encode the low frequency image LL_2 . The encoded *LF* sub band is expressed by LL_{en} .

STEP 4 Feature encoding

The locations of the key points are represented by a binary matrix M , which has the same size as the input image. If $k(x, y)$ is a feature key point, M_{xy} is set to 1, otherwise M_{xy} is set to 0. The binary matrix M is encoded using binary run-length coding. We use 14 bits to encode each sift key point, 7 bits for the scale and others for the orientation. $K_{en} = \{Pt_{en}, S_{en}, O_{en}\}$ is used to express the encoded feature points. The encoded bit stream is consist of LL_{en} and K_{en} .

3.3 Decoder

STEP 1 Bit stream decoding

The coded bit stream is first split into two parts— LL_{en} and K_{en} . HEVC decoder is used for low frequency decoding to get LL_{de} . And Pt_{de}, S_{de} and O_{de} are got by the corresponding key point decoding.

STEP 2 Feature descriptor calculation

After we get the location, scale and orientation of all the key points, we need to use these data to calculate the SIFT descriptor of each key point to do the match with image database. We take 2-level HF-missed inverse wavelet transform with LL_{de} and use Pt_{de}, S_{de} and O_{de} to calculate the SIFT descriptors in the IWT image I_{Lde} . Then we can get the descriptor $D_{de}\{d_1, d_2, \dots, d_n\}$.

STEP 3 SIFT Key point matching

There is a large image database in the decoder. Images in the database are processed as the same way as for the input images. There are lot of features in the database images, making a key point set K^D , whose corresponding descriptor set is $D^D\{d_1^D, d_2^D, \dots, d_m^D\}$. Assume that we have a certain *HF* key point k_i , we use its descriptor d_i to do the matching. If the Euclidean distance between d_i and d_j^D is the minimum value of all the database key point descriptors, k_i and k_j^D are a pair of matching points, which is denoted by M_i .

STEP 4 Optimizing and Clustering

In the encoder, the *HF* sub band key point K_H is the difference set of K_I and K_L . In the decoder, the SIFT descriptors are calculated in the I_{Lde} , which may have errors when calculating the descriptors, so as the matching process. Therefore we do not use all the matching points to restore the *HF* sub bands. A threshold value

$$Rs = 0.3 \times \max\{dist(k_i, M_i)\} \tag{4}$$

is used to reduce mismatching points.

Where $\{dist(k_i, M_i)\}$ is the set of all the Euclidean distance between descriptors of k_i and M_i . For a certain key point k_i , if the Euclidean distance between k_i and its

matching point M_i is smaller than R_s , We regard M_i as a good matching, and this matching pair will be reserved. Otherwise, we regard M_i as a mismatching, the matching pair will be ignored.

After optimization, we need to take features into cluster. One reason is that clustering can effectively avoid image patches overlapped closely to each other when pasting to the HF sub bands. The other reason is to calculate the transformation matrix more precisely—we can calculate different transformation matrix for different cluster. We use K-means algorithm to do the clustering, which can be expressed by:

$$Cluster_{mk} = \left\{ d_{kj} = \min \sum_{j=1}^K \sum_{i=1}^J \| pt_i - \mu_j \|^2 \right\} \quad (5)$$

where m is the index number of a database image, μ is the coordinate value of the cluster center.

STEP 5 HF sub bands restoration

When the cluster process is completed, for each cluster, we use RANSAC algorithm [12] to wipe off some mismatching points, and calculate the transformation matrix according to the rest of matching key points. Then we use the transformation matrix to all points in the cluster, make perspective transformation of the patches and paste them to the HF restoration sub bands to get the restored HF sub band images H_r .

STEP 6 Wavelet inverse transform

2-level inverse wavelet transform is used at this step. In the first level, LL_{1de} is got from LL_{de} and H_{r2} . Similarly, the final reconstructed image is produced by the second transform level.

4 Experimental Results

The facial database Utrecht ECVP [13] is used as our experimental image database. There are 131 images taken from 49 men and 20 women whose expression is usually natural or smile. The image resolutions is 900×1200 for all images. Each person has 1–3 pictures. The sense organs and hair cuts are different between different persons. We pick 6 images as the input image and the input image need to be removed from the database. After we get the reconstructed image, we compared it with HEVC and JPEG2000. The input images are shown in Fig. 2.

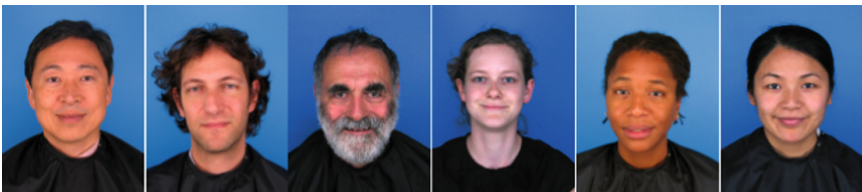


Fig. 2. Input images. From left to right marked “a” to “f”.

Basic information of the input images are shown in Table 1.

Table 1. Basic information of input images

Image index	Size (KB)	LF size (KB)	No. of K_I	No. of K_L	No. of K_H
a	3164.06	198	301	102	258
b			766	292	648
c			638	252	536
d			220	128	163
e			175	107	122
f			306	125	256

Take image *b* as an example, the reconstructed image is shown in Fig. 3. The comparison of compress ratio and image quality is shown in Tables 2 and 3.



(a)



(b)

Fig. 3. The reconstructed results of image *b*. In the row (a), images from left to right: input image, proposed algorithm, HEVC, JPEG2000. The second row is the corresponding details. In the row (b), details from left to right: proposed algorithm, HEVC, JPEG2000.

Table 2. Comparison of compression ratio

Image	L_{en} (KB)	K_{en} (KB)	Compression ratio		
			Proposed	HEVC	JPEG2000
a	1.68	1.45	1011	1010	500
b	2.63	3.64	505	509	500
c	2.40	3.01	585	591	501
d	1.72	0.92	1198	1055	498
e	1.81	0.69	1265	1044	498
f	1.79	1.44	980	1014	508

Table 3. Comparison of image quality

Image	Proposed		HEVC		JPEG2000	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
a	36.31	0.93	35.68	0.93	38.2	0.93
b	34.01	0.91	33.86	0.91	33.08	0.90
c	33.13	0.91	34.48	0.91	33.62	0.90
d	36.67	0.93	36.18	0.92	36.37	0.92
e	36.19	0.93	36.36	0.92	34.59	0.91
f	35.55	0.94	36.21	0.93	36.15	0.93

From Fig. 3 we can see our coding scheme has a satisfactory subjective visual effect. Under high compression ratio, details of HEVC-coded images will become blurred, and JPEG2000 would appear blocking artifact.

Table 2 shows the comparison of compression ratio between proposed algorithm, HEVC and JPEG2000. In our experiment, when doing the LF sub band encoding, we set the VQ to 31, the compression ratio of proposed algorithm is shown on the first line of compression ratio table. When compared to HEVC, making compression ratio of HEVC close to that of proposed algorithm, the corresponding PSNR and SSIM is shown in Table 3. We can see that compared with HEVC, when the compression ratio is close to or higher than HEVC, the proposed algorithm has a higher PSNR and SSIM. When compared to JPEG2000, we set the JPEG2000 to the highest compression ratio it can reach. We can see that when the compression ratio of proposed algorithm is much higher than that of JPEG2000, the PSNR and SSIM of proposed algorithm can be equal or better than JPEG2000.

From the experiment we can see that the proposed coding method has a good performance not only in compression ratio but also in the image quality of reconstructive image. A further optimization will be used to the feature detection step to improve the coding performance.

5 Conclusions

This paper proposed a novel facial image coding scheme in the transform domain base on SIFT. The input image is first decomposed into low frequency sub band image and high frequency sub band images by wavelet transform. While the high frequency part is encoded according to SIFT descriptors rather than pixel value at the encoder side. SIFT-match is used at the decoder side to restore the high frequency images. Final reconstructed image is obtained by inverse wavelet transform. Experimental results show high compression ratio of our coding scheme. However, it still needs to be improved for processing edge information.

Acknowledgment. This work was supported by:

1. The State Key Program of National Natural Science Foundation of China (61631009);
2. The Jilin province science and technology development project of China (20150204006GX);
3. Graduate Innovation Fund of Jilin University.

References

1. Skodras, A., Christopoulos, C., Ebrahimi, T.: The Jpeg 2000 still image compression standard. *IEEE Sig. Process. Mag.* **18**(5), 36–58 (2001)
2. Sullivan, G.J., Ohm, J., Han, W.J., et al.: Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012)
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **26**, 91–110 (2004)
4. Weinzaepfel, P., Jegou, H., Perez, P.: Reconstructing an image from its local descriptors. In: *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 337–344. IEEE Computer Society (2011)
5. Xiao, X., Xu, C., Wang, J., et al.: Enhanced 3-D modeling for landmark image classification. *Multimedia IEEE Trans.* **14**(4), 1246–1258 (2012)
6. Kang, L.W., Hsu, C.Y., Chen, H.W., et al.: Secure SIFT-based sparse representation for image copy detection and recognition. In: *2010 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1248–1253 (2010)
7. Hsieh, S.L., Chen, Y.W., Chen, C.C., et al.: A geometry-distortion resistant image detection system based on log-polar transform and scale invariant feature transform. In: *2011 IEEE 13th International Conference on High Performance Computing and Communications (HPCC)*, pp. 893–897 (2011)
8. Wengang, Z., Houqiang, L., Richang, H., et al.: BSIFT: towards data-independent codebook for large scale image search. *IEEE Trans. Image Process.* **24**(3), 967–979 (2015)
9. Chimlek, S., Piamsa-Nga, P.: Landmark image searching with inattentive salient regions. In: *2014 International Conference on Information Science and Applications (ICISA)*, pp. 48–52. IEEE (2014)
10. Yue, H., Sun, X., Yang, J., Wu, F.: Cloud-based image coding for mobile devices—toward thousands to one compression. *IEEE Trans. Multimedia* **15**(4), 845–857 (2013)

11. Song, X., Peng, X., Xu, J., Wu, F.: Cloud-based distributed image coding. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 4802–4806 (2014)
12. Bolles, R.C., Fischler, M.A.: Ransac-based approach to model fitting and its application to finding cylinders in range data. In: International Joint Conference on Artificial Intelligence (1981)
13. Utrecht ECVF database. http://pics.psych.stir.ac.uk/2D_face_sets.htm