Video Coding With Key Frames Guided Super-Resolution

Qiang Zhou, Li Song, and Wenjun Zhang

Institute of Image Communication and Information Processing
Shanghai Jiao Tong University, Shanghai, 200240, China
{sjtu08zq,song_li,zhangwenjun}@sjtu.edu.cn

Abstract. In this paper a video coding scheme with Layered Block Matching Super-resolution (LBM-SR) is presented. At the encoder side, it divides the video frames into key and non-key frames, which are encoded at original resolution and reduced resolution respectively. During the resolution reduction process, most of the high frequency information in non-key frames is dropped to save the bit-rate. At the decoder side, LBM-SR utilizes a Layered Block Matching method in wavelet domain to restore the lost high frequency parts of the non-key frame, with the nearby key frames as a reference. Due to the similarity between key frames and non-key frames, the experimental result is remarkable and the whole scheme is demonstrated to be a promising one.

Keywords: Video coding, Super-resolution, Layered Block Matching.

1 Introduction

Video coding has been thriving for decades. The latest video coding standard is H.264 [1], which achieves the state-of-art performance. Besides, the still-in-lab H.265 [2] appears to be the next-generation coding standard. In general, the classical video coding schemes and improvements made to them are devoted to cut down the bit-rate of the coded video while keeping its quality. But they sticks to a fixed pattern, trying to compress a video without doing anything in advance.

Generally speaking, a pre-processing stage and the corresponding post-processing stage can be incorporated into a video coding framework, on the premise that the pre-processing stage makes the encoding process more efficient and the post-processing stage is able to restore the information lost due to the pre-processing operation, as shown in Fig. 1.

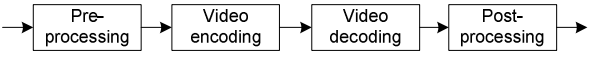

Fig. 1. Video coding scheme with pre- and post-processing

In order to reduce the bit-rate of the coded video, a down-sampling process can be used in the pre-processing stage. Then, at the post-processing stage, an up-sampling process is required, which can be attributed to a super-resolution problem.

G. Qiu et al. (Eds.): PCM 2010, Part II, LNCS 6298, pp. 309–318, 2010.

Super-resolution usually obtains one high resolution (HR) image from several low resolution (LR) images with sub-pixel shifts from each other [3]. But in video coding scenario, there exists little sub-pixel shifts in consecutive frames. As a result, only single frame super-resolution is possible to obtain a full size video with the help of limited available information from nearby frames, if all frames are down-sampled. But more prior information can be exploited to achieve better results [4] [5].

Our proposed Layered Block Matching Super-resolution (LBM-SR) algorithm evolves from semi super-resolution in [5], where a video is divided into key frames with full size and non-key frames with a quartered size. The novelty of LBM-SR is its carrying out block matching in wavelet domain, coupled with a layering notion and an overcomplete manner.

When LBM-SR is applied to a video coding scheme, all the non-key frames should be down-sampled, which contributes to the reduction of bit-rate. Then at the decoder side, these non-key frames with reduced resolution will be restored via LBM-SR. This novel coding scheme with super-resolution will outperform the classical one if the bit-rate is lower than a certain threshold [7].

This paper is structured as follows. We introduce the Layered Block Matching Super-resolution method step by step in section 2. And a video coding scheme using this algorithm can be found in section 3. Section 4 presents the experimental results. And we conclude this paper in section 5.

2 Layered Block Matching Super-Resolution

The semi super-resolution method [5] divides an image into two explicit components, the low frequency one and the high frequency one. The low frequency part is a down-sampled image, while the high frequency one is obtained through subtracting the interpolated image from the original one.

Such a subtraction manner is not enough when handling an image. According to the wavelet decomposition theory, a wavelet filter decomposes an image I into four matrices, CA, CH, CV and CD, as shown in Fig. 2. CA represents the low frequency part while the other three matrices denote the high frequency parts. Better results can be expected if the wavelet decomposition theory is introduced.

A wavelet filter can be used to generate a down-sampled image I_d, which inherits the low frequency information of I. When super-revolving I_d to I', we need to estimate the CH, CV and CD parts. The restored HR I' can be evaluated via equation (1), where CH', CV' and CD' are the estimated matrices, and W performs the wavelet synthesis.

$$I' = W(I_d, CH', CV', CD') \tag{1}$$

We may denote Simple Inverse [6] as the process of setting CH', CD' and CV' to zero then performing equation (1). Simple Inverse produces a full size image whose most high frequency information is lost. Obviously, simply performing Simple Inverse isn't enough. The high frequency parts lost in the down-sampling process should be restored as much as possible.

Following the notion in semi super-resolution, some frames are selected as the key frames, which aren't down-sampled and contain useful high frequency information.

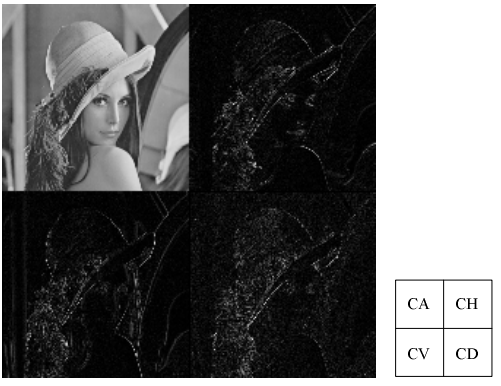

Fig. 2. Image decomposition using a wavelet filter

And non-key frames refer to the frames going through down-sampling. There are a predefined number of non-key frames between two key frames. So how much high frequency we can make use of from the key frames nearby determines the overall performance of the super-resolution.

It should be noted that the four parts generated from wavelet decomposition, which characterize an image from different respects, are relevant to each other, as illustrated in Fig. 2. Based on that prior relevance, a Block Matching Super-resolution method is depicted in Fig. 3.

Let I'_{NK} denotes a down-sampled non-key frame, and the two key frames nearby are I_{K1} and I_{K2}, as shown in Fig. 3. The full size frame corresponding to I'_{NK} is I_{NK} and the unknown matrices to be estimated are I'_{NK-CH}, I'_{NK-CV} and I'_{NK-CD}. Besides, we decompose I_{K1} and I_{K2} via a wavelet filter, and I'_{K1} and I'_{K2} are their corresponding low-frequency parts. After partitioning I'_{NK} into N blocks with size $M \times M$, we denote the n[th] block in I'_{NK} as B^n_{NK}, and $B^{(i,j)}_{Mat}$ stands for a $M \times M$ block located at (i, j) in matrix Mat. According to Fig. 3, we have

$$[i_n, j_n, idx] = \arg\min_{(i,j)\in I'_{K1}} \{MAD(B^n_{NK}, B^{(i,j)}_{I'_{K1}}), MAD(B^n_{NK}, B^{(i,j)}_{I'_{K2}})\}, \quad idx \in \{1,2\} \tag{2}$$

where MAD is the Minimum-Absolute-Difference criteria, and i_n, j_n tells us the location of the most matching block in I'_{K-idx}.

Then, we can figure out I'_{NK-CH} as

$$I'_{NK-CH} = \{n \in [1, N] \mid B^n_{NK-CH}\}, B^n_{NK-CH} = B^{(i_n, j_n)}_{K_{idx}-CH} \ . \tag{3}$$

I'_{NK-CV}, I'_{NK-CD} can be obtained in the same way.

The overall performance of this super-resolution method relies largely on the matching degree of the searched block. To improve it, a Layered Block Matching (LBM) method is introduced, as shown in Fig. 4. That is, partitioning I'_{NK} in several steps using block B_{Li} with decreasing size. In each step the block matching result $posMat_i$ is used as the initial position parameter for matching in the next step. And $posMat_0$ is initialized to be the original position of B_{L1} in I'_{NK}. Because larger block contain more global information, which helps to approximate the position of the most

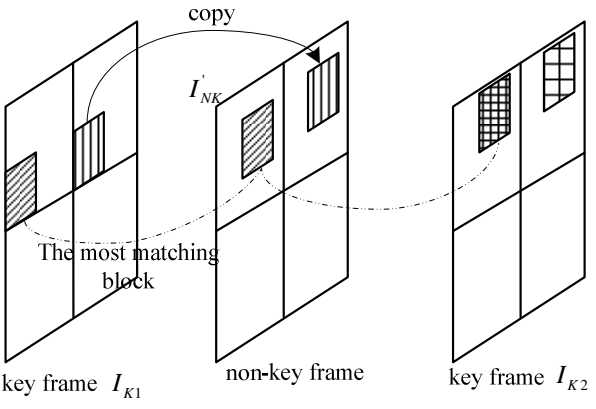

Fig. 3. Block Matching Super-resolution

matching block, LBM is expected to perform better than classical block matching methods. Summary of this iterative process is shown as

$$\begin{aligned} &init \quad posMat_0 \\ &for\ i = 1\ to\ \mathrm{m} \\ &\quad posMat_i = \mathrm{BM}(B_{Li}, posMat_{i-1}) \\ &end \end{aligned} \tag{4}$$

Then $\hat{I}_{NK}$, the reconstructed frame, can be synthesized via

$$\hat{I}_{NK} = W(I'_{NK}, I'_{NK-CH}, I'_{NK-CV}, I'_{NK-CD}) . \tag{5}$$

To improve the overall robustness, we can perform block matching in the final step of LBM with an overcomplete manner. In that case, a frame is divided into overlapped blocks when performing block matching and the finally matched blocks are averaged over the overlapping factor. If a pixel is included in n blocks, then the overlapping factor is n for this pixel. The overlapped blocks can be obtained by shifting them half of the block size in horizontal and vertical directions respectively. Though this over-complete block matching method may significantly increase the computational cost, the gain of image quality is proved to be remarkable.

3 Video Coding With LBM-SR

In this Section, we integrate the proposed LBM-SR algorithm into a classical video coding framework, where H.264 is utilized as a video coding tool for example. The encoder of the proposed scheme is responsible for encoding a normal video to a hybrid stream, in which there are key frames and non-key frames with different size. And the decoder does the opposite thing, where LBM-SR works.

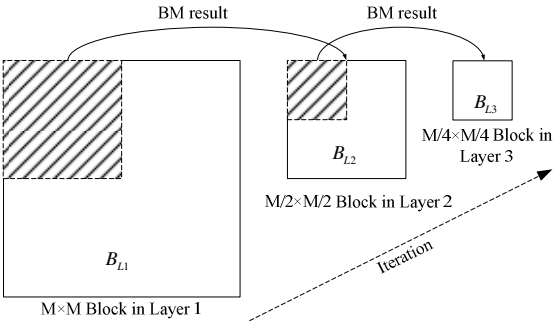

Fig. 4. Illustration of Layered Block Matching

A. Scheme of the Encoder

At the encoder side, the video frames are divided into key and non-key frames. Then key frames are encoded in intra mode of H.264, and non-keys frames are encoded as P/B frames, with the down-sampled key frames as a reference. A detailed diagram of this encoding procedure is illustrated in Fig. 5, where key frames experience intra coding and decoding to avoid drifting errors.

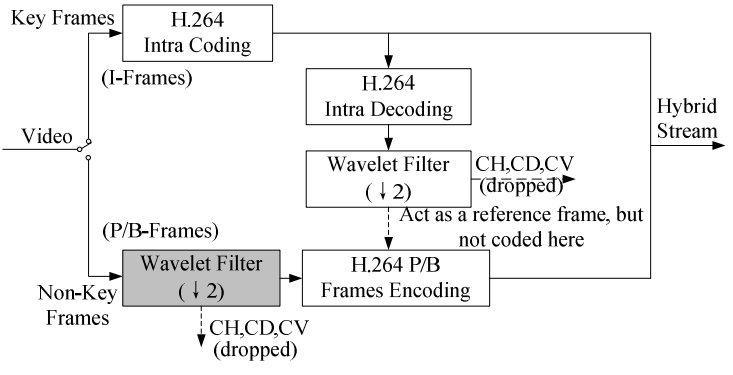

Fig. 5. Scheme of the encoder

The wavelet filters in Fig. 5. are responsible for down-sampling an image, as mentioned in section 2. After down-sampling, the high frequency components of a non-key frame are dropped and an image with a quarter of its original size is generated, which will be encoded as P/B frames in the next step.

In general, the hybrid stream produced by the encoder contains intra-coded key frames and non-key frames going through H.264 P/B frames encoding process.

B. Scheme of the Decoder

The decoder side is where we decode the hybrid stream and restore the missing high frequency information in non-key frames. Fig. 6. gives a detailed description.

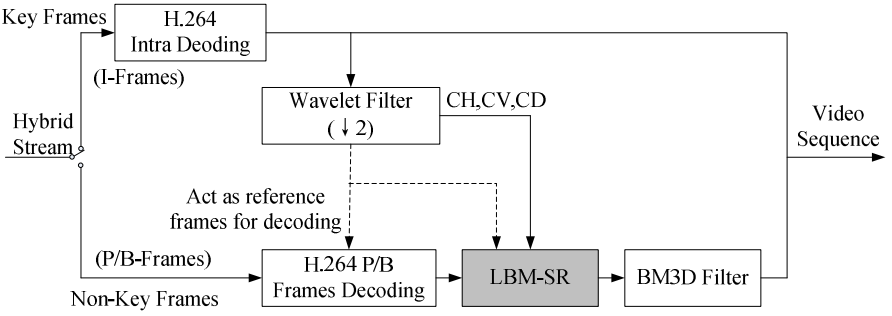

Fig. 6. Scheme of the decoder

In Fig. 6, intra decoding is used to decode the key frames. And the down-sampled key frames after intra decoding play the role of reference frames for decoding non-key frames. Then LBM-SR super-resolves the non-key frames to restore the dropped high frequency components due to the down-sampling process at the encoder side.

The non-key frames reconstructed by LBM-SR may contain some high frequency noise, due to the possible mismatch when performing block matching. Hence, one extra step after LBM-SR for denoising is necessary. Here, a BM3D filter [8] is utilized because it can achieve the state-of-art performance in removal of high frequency noise.

At the end of the decoding part, the restored non-key frames and the decoded key frames are reordered to a video sequence.

4 Experimental Results

Because LBM-SR is the essential part of the whole video coding scheme, the proposed scheme will be evaluated from two respects, the performance of LBM-SR and the performance of proposed video coding scheme, which will be compared with Pure H.264. Here, Pure H.264 stands for the classical H.264 coding scheme without extra pre- and post-processing stages.

4.1 The Performance of LBM-SR

Experiments about super-resolving video with LBM-SR are conducted in this part. Firstly, we down-sample Foreman (CIF format, 91 frames) using a wavelet filter, with the key frames excluded from this procedure. And we denote GOP (Group of Pictures) as a group of pictures containing a key frame and the consecutive non-key frames before the next key frame. The size of a GOP is set to be 15 in the following experiment. Then the down-sampled non-key frames can be super-resolved using the proposed LBM-SR algorithm.

The graph in Fig. 7. illustrates the PSNR values of each restored non-key frame, where the red points stand for non-key frames which are immediately before or next to a key frame. It's interesting to note that the PSNR values of each GOP form a "valley", where non-key frames farther away from the key frames have smaller PSNR values compared with those closer to the key frames. This "valley" effect is mostly determined

by the fact that the similarity between two frames decreases as the distance between them increases. But this won't do much harm to the overall visual quality, because a video is an active set of images, and this property will attenuate this minor difference between frames.

Furthermore, our LBM-SR algorithm is compared with the Simple Inverse [6] method, so that we can see how much high frequency information LBM-SR can restore during super-resolution. The results are shown in Fig. 8, where the 2nd frame, the 54th frame and the 75th frame represent the frames next to a key frame, between two key frames and before a key frame respectively. Obviously, the frames reconstructed by LBM-SR look much smoother, and the edges of the wall are never jagged.

The computational cost for LBM-SR is low due to its simplicity. Moreover, since the overcomplete block matching process can be divided into several independent steps, parallel computation can be utilized to reduce the reconstruction time with a large degree.

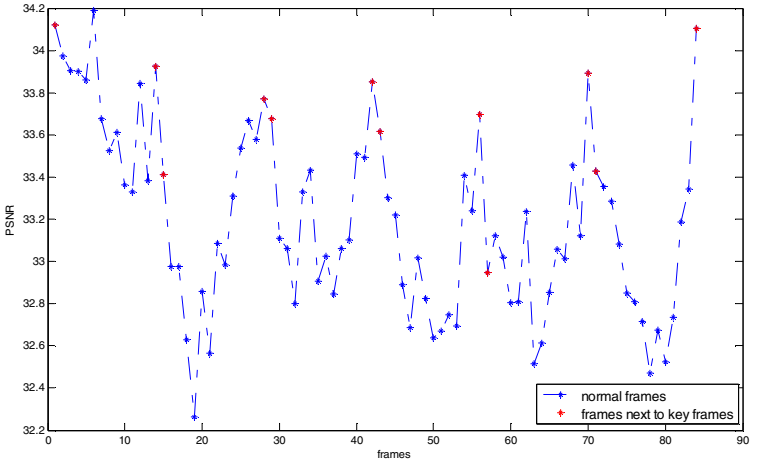

Fig. 7. PSNR values of the non-key frames in Foreman restored by LBM-SR

A. The Performance of LBM-SR for Video Coding

The proposed video coding scheme is compared with Pure H.264 in this part, on the premise that coded video of the two schemes are of about the same bit-rate. QP parameter of H.264 encoder controls the bit-rate, and GOP size of these two schemes is set to 15. By the way, x264 encoder is used to perform H.264 encoding [9].

Table 1. gives the SSIM [10] values, where Pure H.264 sets the global QP for all frames, whereas our proposed scheme sets the QP of key frames (KF) and non-key frames (NKF) respectively. It is proved by experiments that the QP value of the key frame is more important to the final visual quality, compared with the QP value of the non-key frame. So in our experiment, we pay more attention to the QP value of the key frames.

Fig. 8. Foreman example of super-resolution using Simple Inverse and LBM-SR respectively

From Table 1, our proposed scheme is comparable to Pure H.264. Part of the video frames generated by these two schemes is illustrated in Fig. 9, where the visual quality of the frames from our proposed scheme is also comparable to Pure H.264. There may be some defects in these frames, mainly due to the blur effect caused by the incomplete high frequency part. Moreover, during LBM-SR, the quantization noise is augmented and this noise may contribute to the mis-match when the high frequency components are restored.

Table 1. SSIM values of proposed scheme compared with pure H.264

Video sequences	Pure H.264			Proposed Scheme			
	SSIM	Bit-rate	QP	SSIM	Bit-rate	QP of KF	QP of NKF
Foreman	0.9427	479kbps	28	0.8977	437kbps	22	22
Carphone	0.9505	397kbps	28	0.9278	407kbps	22	24
City	0.9886	3167kbps	28	0.9794	3075kbps	22	26
Crew	0.9828	2633kbps	28	0.9776	2445kbps	22	22

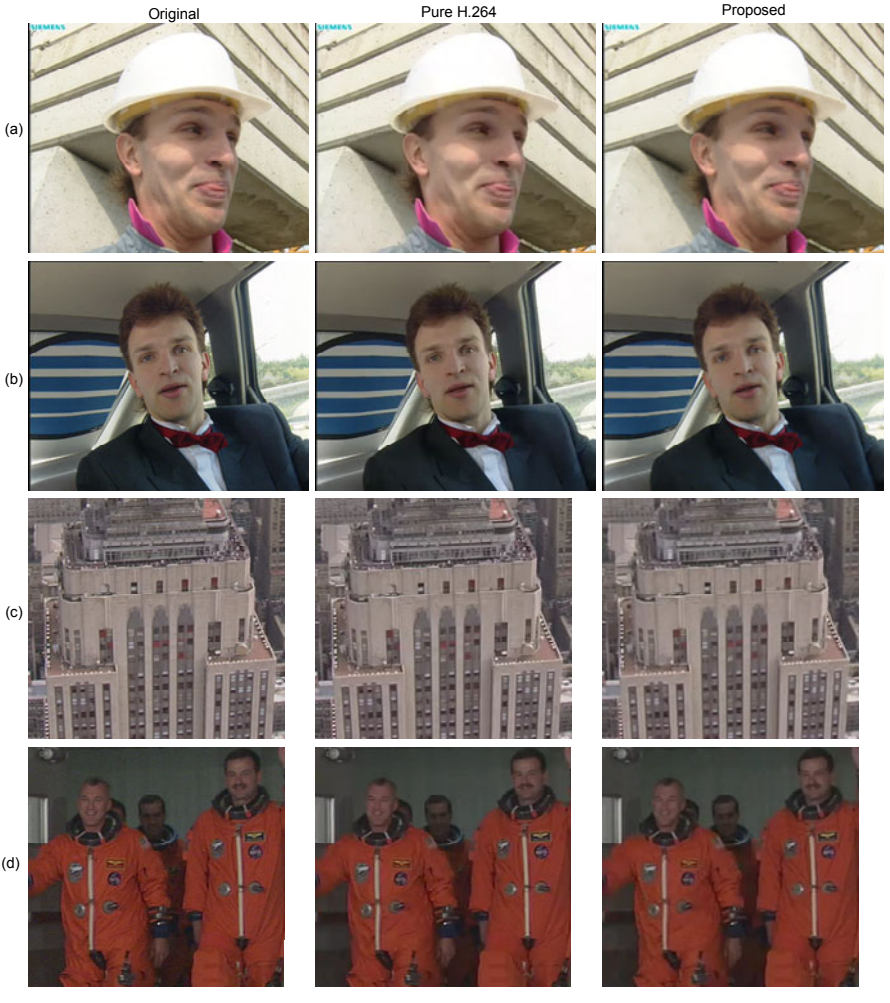

Fig. 9. Subjective video coding results of the 7th frame from four video sequences, using Pure H.264 and proposed coding scheme respectively. (a) Foreman (CIF), (b) Carphone (CIF), (c) a 320x300 block at (450,420) in City (720p), (d) a 320x300 block at (600,120) in Crew(720p).

When you stare at the isolated frames in Fig. 9, the difference between them will be obvious. It will be another story if you are watching a video which may be displayed at 25 frames per second. In that scenario the difference between frames will be attenuated due to the motion of pictures and the duration of vision effect of our eyes.

Let's come back to Fig. 9. Objects with simple shapes like the wall in Foreman are reconstructed with little distortion because they can be characterized by size-fixed blocks. However intricate objects, like human faces in Foreman and Crew, require more exact models to characterize them. So this may be the focus of our future work.

5 Conclusion

In this paper, LBM-SR and its application in video coding are presented. This LBM-SR method is demonstrated to be powerful when super-resolving a video. Then a video coding scheme with LBM-SR is described, which achieves remarkable performance with low computational cost. Thus the proposed video coding scheme is demonstrated to be a promising one, and may indicate a new video coding methodology. Our future work will focus on how to make the scheme more robust, through introducing more regularization to improve the accuracy of the restored high frequency components, and making the whole scheme adaptive to the content of the video.

Acknowledgments

This work was supported in part by NSFC(60702044, 60625103, 60632040), MIIT of China(2010ZX03004-003) and 973(2010CB731401, 2010CB731406).

References

1. Wiegand, T., Sullivan, G.J., Bjøntegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. 13(7), 560–576 (2003)
2. The H.265 website, http://www.h265.net/
3. Park, S., Park, M., Kang, M.: Super-resolution image reconstruction: A technical overview. IEEE Signal Process. Mag. 20(3), 21–36 (2003)
4. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super resolution. IEEE Comput. Graph 22(2), 56–65 (2002)
5. Brandi, F., de Queiroz, R.L., Mukherjee, D.: Super resolution of video using key frames. In: Proc. of International Symposium on Circuits and Systems, Seattle, WA, USA (May 2008)
6. Boon, C.S., Guleryuz, O.G., Kawahara, T., Suzuki, Y.: Sparse super-resolution reconstructions of video from mobile devices in digital TV broadcast applications. In: Proc. SPIE Conf. on Applications of Digital Image Processing XXIX, in Algorithms, Architectures, and Devices, San Diego (August 2006)
7. Molina, R., Katsaggelos, A., Alvarez, L., Mateos, J.: Towards a new video compression scheme using super-resolution. In: Proceedings of the SPIE Conference on Visual Communications and Image Processing, San Jose, CA, USA, vol. 6077, pp. 607706/1–607706/13 (2006)
8. Dabov, K., Foi, A., Katkovnic, V., Egiazarian, K.: Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Trans. Image Process. 16(8), 2080–2095 (2007)
9. The x264 software, http://www.videolan.org/developers/x264.html
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing 13, 600–612 (2004)