Techniques for Flexible Image/Video Resolution Conversion with Heterogeneous Terminals

Ming-Sui Lee, Mei-Yin Shen, and C.-C. Jay Kuo, University of Southern California
Akio Yoneyama, KDDI R&D Laboratories, Japan

The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement no. EEC-9529152. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

ABSTRACT

Multimedia capturing and display devices of different resolutions and aspect ratios can be easily connected by networks and, thus, there is a great need to develop techniques that facilitate flexible image/video format conversion and content adaptation among these heterogeneous terminals. Quality degradation due to downsampling, up-sampling, coding/decoding, and some content adaptation mechanism (say, image mosaicking) in the transmission process is inevitable. It is desirable that multimedia contents can be easily captured, displayed, and seamlessly composed. Challenges and techniques to achieve this goal are reviewed first. Then, two specific topics, i.e., image/video mosaicking and super resolution (SR) conversion, are highlighted. As compared with previous work developed for these problems, the challenge under the current context is to strike a balance between low computational complexity and high quality of resultant image/video. Several new developments along this line are discussed.

INTRODUCTION

Multimedia terminals such as digital cameras, cellular phones, PDA's (personal digital assistants), computers, and HDTV's are prevalent these days. These personal and/or home electronic devices provide accessibility and convenience for users in many applications, such as communicating with other people, learning the latest information, acquiring multimedia content, and being personal secretary anytime and anywhere. Besides expediency as described above, the demand from consumers has arisen to an even higher level. That is, people would like to capture, manipulate, and display multimedia content more flexibly with heterogeneous terminals over the network so that this content can be shared with friends and peers effortlessly.

For this purpose, format/rate conversion and content adaptation techniques are needed to compensate for the differences in these devices. This scenario is illustrated in Fig. 1, where we have desired multimedia outputs from certain media acquisition devices. This input and output content is related by some media processing operations such as resolution conversion, quality enhancement, aspect ratio alternation, image/ video stitching, color depth adjustment, chromaticity tuning, etc. One such application occurs in the digital home environment, where various electronic devices are connected by wirelines or wirelessly. Digital TV programs can be transmitted from the set-top-box to cellular phones or PDAs so that users can view them conveniently any place at home. Since the display resolutions of portable devices are much lower than that of TV, content has to be converted to fit the requirements of low resolution (LR) terminals without extensive degradation of visual quality to end users. This involves the downsampling of video content in spatial and, maybe, temporal resolutions. This imposes some challenges if we would like to perform downsampling in the compressed video domain to save the computation of decoding and re-encoding.

On the other hand, one may want to view multimedia contents captured by portable devices on higher resolution terminals such as PC or TV screens at home. As compared with high resolution video capturing devices, the price of high resolution display panels has significantly dropped in recent years. Generally speaking, a high definition (HD) video capturing device is more expensive than an HD display device, and it is challenging to get interesting HD content for HD display devices. Three different scenarios for this application are summarized here. The first one is that there is only one single LR input to generate a single HR output. SR techniques are required to upsample the low

Wireless and mobile communications over wide areas become more and more popular due to the emerging wireless IP networks and services. However, video transmission and streaming may suffer from unreliable Internet connection and heterogeneous bandwidth of different receivers.

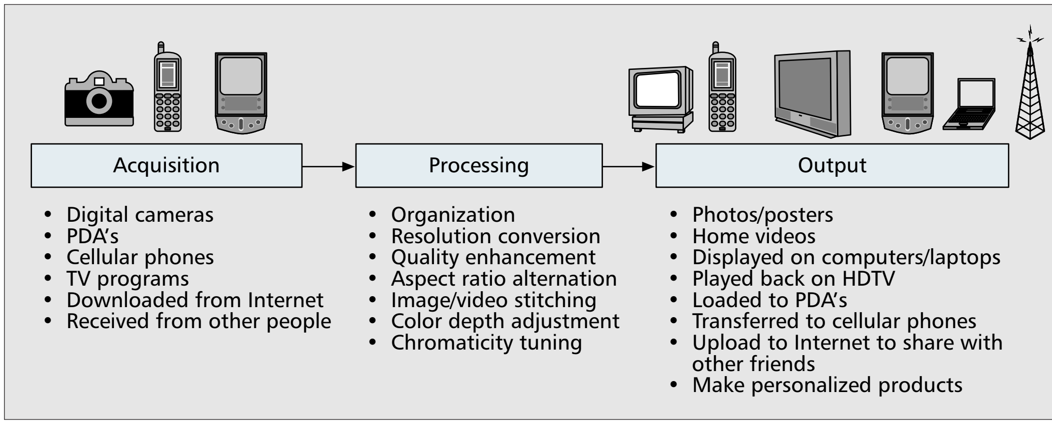


Figure 1. *Media content acquisition, processing, and display on heterogeneous terminals.*

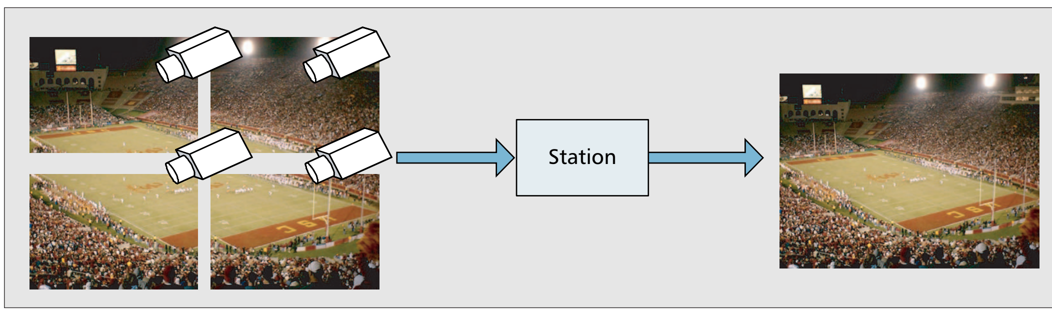
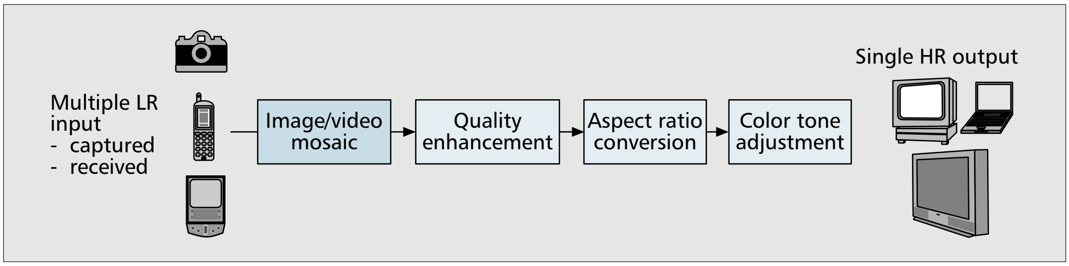


Figure 2. *Video mosaic generation using multiple cameras.*

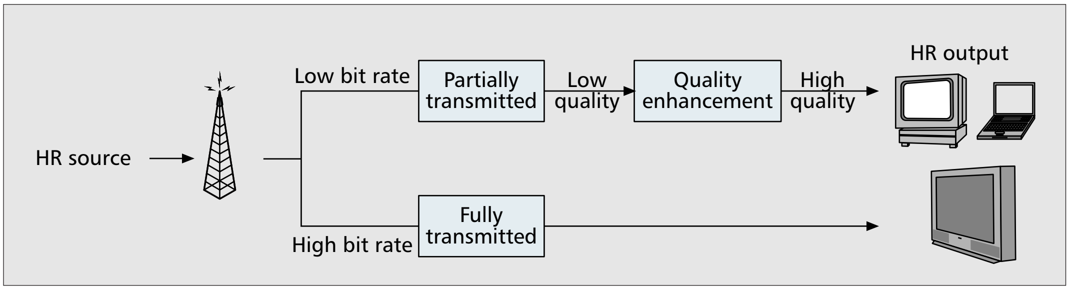
resolution content to enlarge the size and post-process it to enhance the resultant video quality. However, the single input video cannot provide extra information so that the improvement of image/ video resolution is limited. The second scenario, which is more practical nowadays, is that multiple captured images from the same scene are taken as the input. Those input images are taken by the same capturing device at different time instances with a limited range of camera motion. The application of SR techniques to LR images generates images with the same viewing angle but a higher resolution. The aspect ratio conversion is needed since the aspect ratios of digital cameras, cellular phones, PDAs, computers, and HDTVs are usually different. Besides, in order to show vivid colors on the target display panel, chromaticity tuning has to be performed so that the color gamut of different devices can be optimally mapped. The third scenario to obtain HD content is through image/video stitching, which is usually known as image/video mosaicking. An example is shown in Fig. 2, where one takes multiple low resolution (LR) image/video content as input to generate a panorama view to fit the HD display. Since multiple capturing devices may not be synchronized well in space and time, image/video mosaicking is in general a challenging task; they only work well under many constraints. One possible application is video conferencing of multiple people in two rooms located remotely, where multiple cameras can be calibrated well in advance. Once the panorama view is formed, post-processing techniques such as quality enhancement, aspect ratio conversion, and color tone adjustment may still be needed, as shown in Fig. 3. This system will benefit users to generate an HR output at a lower cost.

Wireless and mobile communications over wide areas become more and more popular due to the emerging wireless IP networks and services. However, video transmission and streaming may suffer from unreliable Internet connection and heterogeneous bandwidth of different receivers. To address this scenario, the layered representation of video contents and their layered transmission and broadcasting (or multicasting) are of great value, as shown in Fig. 4. Techniques for layered video transmission could be a topic worth full treatment by themselves. Here we would like to emphasize the potential need of a quality enhancement module to convert the base video layer back to some high resolution (HR) content.

In summary, techniques for flexible image/ video resolution conversion and content adaptation are in great demand to facilitate image/video data exchange over the network consisting of heterogeneous capture and display terminals. Technical challenges to achieve this goal will be reviewed first. Then, two specific topics, i.e., image/video mosaicking and super resolution conversion, are highlighted, respectively. As compared with previous work developed for these problems, the challenge under the current context is to strike a balance between low computational complexity and high quality of resultant image/video. Several new developments along this line are discussed. Finally, concluding remarks are drawn and future work directions are pointed out.



■ **Figure 3.** *Combining multiple LR inputs to generate a single HR output display via image/video mosaicking followed by post-processing techniques.*



■ **Figure 4.** *An example of layered video broadcasting followed by video quality enhancement for received LR video.*

A more complicated processing technique is required to convert a lower resolution image to a higher one since some image pixel is not present in the input and has to be interpolated based on values of its local neighbors. This is essentially an ill-posed problem.

Research Issues

Several challenging problems in achieving flexible image/video format conversion and content adaptation among heterogeneous terminals are examined in this section.

Frame Resolution Conversion and Aspect Ratio Alternation

Resolution is used to describe the spatial dimension of digital image/video or the total number of pixels of a capture and/or display device. The resolutions of commonly used displays are shown in Table 1. On one hand, although HR images can contain 16 mega-pixels or higher, most images are downsampled to around five megapixels or even smaller to avoid a huge file size which makes transmitting, editing, and storage difficult. On the other hand, if a low resolution image captured by a cellular phone has a resolution 128 × 128, and we would like to display it on a higher resolution screen, say, 1024×768, some SR processing techniques are needed to add more details to the image so that the blur visual effect can be reduced to improve visual quality.

Resolution conversion includes downsampling and upsampling, which shrink and enlarge the image size, respectively. It is easier to convert an image from a higher resolution to a lower one. Downsampling can be implemented by either averaging or eliminating some rows and columns without severe degradation. In contrast, a more complicated processing technique is required to convert a lower resolution image to a higher one since some image pixel is not present in the input and has to be interpolated based on values of its local neighbors. This is essentially an ill-posed problem. The simplest method to achieve upsampling is to perform linear interpolation, which is equivalent to the convolution of the input with a triangular-shaped function. For the 2D case, a bilinear interpolation which estimates the inserted pixel value from its four nearest neighbors can be easily applied. A general higher-order interpolation framework [1] to smooth an image by considering the average of a larger area of pixels can be derived. Examples include the famous cubic-spline interpolation. Unfortunately, interpolation via linear filtering blurs sharp edges. To address this problem, interpolation techniques with nonlinear filtering have been developed in [2]. More sophisticated SR techniques will be reviewed later.

The aspect ratio of an image is defined as its width divided by height. Several commonly used aspect ratios for various applications are shown in Table 1. If an image is displayed on a device with an aspect ratio different from that of an image, a proper modification is required. Typically, the image is stretched to fit the screen or some area is cropped out or left blank in a straightforward manner. Finally, frame rate conversion is another interesting topic. For a thorough treatment on this topic, we refer to [3].

Color Depth Adjustment and Chromaticity Tuning

Color depth, also known as pixel depth or bit depth, indicates how many bits are used for representing the color of each pixel. Color depth may vary from 1 bit to 24 bits. The Truecolor system that uses 24 bits for each pixel by assigning 8 bits for each color component is popular today. The system has $2^8 = 256$ levels for each component, and it can create 16.7 million distinct colors. The more bits, the more colors can be represented, which implies better visual quality at the expense of a larger file size and longer processing time. Multimedia devices usually have different color depth depending on their applications. The trade-off would be processing com-

Given multiple compressed video inputs, it is desirable to conduct the preprocessing in the DCT domain to generate the corresponding compressed image/video mosaic, since Motion JPEG, MPEG, and H26x coding standards all adopt the DCT representation during the coding process.

	Resolution	Aspect ratio
SDTV	640 × 480	4:3
HDTV	1920 × 1080	16:9
Computer (VGA)	640 × 480	4:3
Cellular phone	128 × 128	1:1
PDA	320×200/480×320	8:5/3:2
PAL DV	720 × 576	5:4

Table 1. *Frame resolution and aspect ratio comparison of different devices.*

plexity and visual quality. For graphics intensive applications such as computer games, better output visual quality requires more color depth, which results in a longer rendering time. In this case, visual quality could be sacrificed by lowering the color depth for fast display. For other applications where visual quality is critical, such as digital cinemas, more bits are assigned for the best possible performance. Another factor that would affect the selection of color depth is the memory space of the device. For portable devices such as cellular phones, their memory space and computing power are usually restricted. Consequently, visual quality could be sacrificed so that lower color depth is allowed. When we exchange multimedia content over all heterogeneous terminals via the Internet, the color depth should be adjusted so as to fit in a broad range of capture and display devices of different color depth.

The CIE chromaticity diagram shows color composition as a function of red and green components. Its value depends on dominant wavelength and saturation. Each point within the diagram represents the mixture of spectrum colors. The point with equal energy represents the white light and any point located on the boundary is fully saturated. A triangle within the diagram forms a color gamut, a range of colors, which can be generated on a specific display screen. Different display devices have a color gamut of different area and position. This explains why the color tone varies from one screen to another even for the same content. To exchange multimedia data freely across different terminals, it is important to reduce the color difference between display panels or screens, which is the goal of chromaticity tuning. Some work forms a mapping function by taking the nearest rule, while others stretch one color gamut to approximate the other as close as possible. However, how to achieve this goal automatically under a generic setting is still an open question.

Image/Video Mosaicking

The image/video mosaicking technique, which combines several image/video inputs into a panorama output, has been widely used in image processing, computer graphics, computer vision, and remotely sensed data processing. We may consider multiple video sources captured by an arbitrary number of cameras with different parameter settings. There arise many challenging problems in creating the video mosaic, including temporal synchronization, focal length readjustment, image registration, and color difference compensation. The discrepancies between smaller video tiles have to be resolved for seamless composition. This topic has been studied for decades, and high quality output image/video can be obtained if there are enough computational resources.

One fundamental tool for successful image/video mosaicking is image/video registration. Generally speaking, the image registration process consists of two major steps: feature detection and feature matching [4]. Feature detection can be done either manually or automatically. Since human eyes are sensitive to geometric patterns, it is straightforward for people to choose the matched patterns. However, an automatic feature selection process based on the particular application context is demanded for computer processing. Feature detection techniques can be classified into two categories: the feature-based and the area-based approaches. The main task of the feature-based approach is to extract salient points such as corners, line intersections, line ends, and centroids of closed-boundary regions. For example, the wavelet transform was used to extract local maxima. The partial derivatives of image pixel values were used for corner detection. However, this process is sensitive to noise and time-consuming. The area-based approach uses the correlation function to determine the degree of closeness. To be specific, it computes the cross-correlation of intensities of certain regions of input images to find the best match. This approach is more suitable for images that do not have many details. Once the feature information is available, the next step is to find the optimal correspondence between image tiles, which is called feature matching. This can be achieved by finding spatial relations among the extracted features.

Although the above methods lead to good results, they were primarily developed in the pixel (or spatial) domain, which is computationally expensive. Given multiple compressed video inputs, it is desirable to conduct the preprocessing in the DCT domain to generate the corresponding compressed image/video mosaic, since Motion JPEG, MPEG, and H26x coding standards all adopt the DCT representation during the coding process. Some recent developments along this line will be described later.

Recent Developments in Super Resolution Technique

Rather than dealing with resolution enhancement of a single video input via linear or nonlinear interpolation, the super resolution technique attempts to synthesize an HR output from several LR inputs. This topic becomes hot nowadays due to the availability of low-cost HR display devices. The effect of warping, blurring, decimating, and noise will all affect the quality of the output video. The goal of the SR technique is to optimize the HR output based on the information of LR inputs. It is an ill-posed problem since the optimal HR video is unknown.

Most SR algorithms can be classified into two types based on the underlying principles, namely, the projection onto convex sets (POCS) and the maximum a posteriori (MAP) estimation. POCS is a set theoretic approach that can incorporate the prior information in the system model easily by forming a convex set for each constraint. The constraints may include the smoothness assumption, validity of the amplitude range, bounded energy and error, etc. A feasible solution should lie in the intersection of all convex sets. By setting up those constraints properly, the final solution is guaranteed not only theoretically but also perceptually satisfied. POCS can be implemented by projecting the calculated solution onto all constraint-based convex sets iteratively, $\underline{X}_{k+1} = P_N P_{N-1} \ldots P_2 P_1 X_k$. However, this method suffers from slow convergence and it may not have a unique solution. Furthermore, if the constraints are over addressed, the resultant video frames lose some details such as sharpness, which results in degraded blurry output.

The MAP estimation is a stochastic approach that finds the estimate of SR video to maximize the conditional probability density function of the output video given multiple degraded input video content. Mathematically, we have,

$$\hat{x} = \arg\max_{x} P(x \mid y_1 y_2, \ldots, y_p),$$

where x is the estimated SR video frame and y_1, y_2, …, y_p are observed LR video frames. A higher conditional probability means that the estimated HR frame is more accurate as compared to the original HR frame. Each observed LR frame is supposed to be the degraded version of an HR frame. The degradation process may contain downsampling, warping, optical distortion, motion blur, noise, etc. One key issue here is the estimation of the degradation model based on the prior information. If the degradation process, H, is appropriately estimated, the SR problem becomes a minimization optimization problem, which can be formulated as

$$\hat{x} = \arg\min_{x} \left[\sum_{k=1}^{p} \| H_k x - y_k \|^2 \right].$$

In other words, the degradation process is performed on the initially guessed HR frame, x, and then subtracted from the observed LR frames so that the error between the degraded LR frame and the observed LR frames can be computed. The goal is to minimize the error as much as possible. The convexity of the optimization function ensures the existence and uniqueness of the solution. The solution process such as the gradient descent method involves matrix inversion and multiplications so that the dimension of the frame is critical to the SR problem in terms of computation complexity.

To combine the advantages of MAP and POCS, a hybrid MAP-POCS reconstruction approach was proposed in [5], where the optimization problem was modified to capture the characteristics of both MAP and POCS. The linear space/time invariant blur and additive Gaussian noise were taken into consideration in this system. Also, input LR frames may not necessarily have the same resolution as the output. By using the Gauss-Seidel conjugate gradient and the steepest descent methods, the unique solution is ensured while the prior information can be successively incorporated simultaneously. Later, Patti and Altunbasak [6] proposed a POCS algorithm to estimate the HR output from an image sequence. They provided a more accurate description of the observations by adopting higher-order interpolants. Besides, they reduced the visibility of artifacts around boundaries, and modified the blur function by a weighting function to make the blur function more like an impulse in the edge gradient direction so that the artifacts around boundaries can be eliminated. Thus, the method proposed in [6] improves both accuracy and robustness of POCS with little extra computation complexity. Instead of reconstructing SR video from an image sequence, Gunturk *et al*. [7] proposed a Bayesian approach which takes multiple video clips as inputs and considers both additive noise and quantization errors since quantization errors contribute to quality degradation in most popular applications.

The SR technique has been developed for color video frames. Since color frames have three channels, the acquired color frames suffer severe degradation due to the low spatial resolution and color filtering so that a demosaicking process is needed to go along with the SR technique in order to enhance the visual quality of the output HR color frame. Farsiu *et al*. [8] proposed a MAP-based algorithm that integrates SR and demosaicking techniques into a single process, where the cost function consists of four terms. The first term, which is the same as the traditional cost function, is the data fidelity penalty term which measures the similarity between the estimated HR frame and LR frames. The second term is the spatial luminance penalty term. It is added since human eyes are more sensitive to the luminance component. The third term plays a regularization role for the chrominance component. The last term, called the intercolor dependencies penalty term, penalizes the mismatch between locations or orientations of edges across color channels. By introducing the three extra terms into the cost function, the resultant color frame has sharper edges with less color artifact. Most of the methods described above do generate good quality of output video. However, their computational complexity is relatively high for real-time applications.

A new research direction for the SR technique is to employ a block-based SR approach. That is, we can decompose one whole frame into small blocks and apply content-adaptive SR techniques to each block. For example, some blocks may have a lower noise level and simple content (say, a smooth plane). It is possible to generate an HR output from this block using low complexity SR algorithms. Sometimes, it may be sufficient to use the simple bilinear interpolation method to reduce the computational complexity. On the other hand, for blocks with blurring degradation, a high noise level and

The SR technique has been developed for color video frames. Since color frames have three channels, the acquired color frames suffer severe degradation due to the low spatial resolution and color filtering so that a demosaicking process is needed to go along with the SR technique

Users will not be satisfied by rich functionalities of an isolated device only, but will also expect compatibilities between different terminals so that they can get the best output based on the platform available. The difference between terminals has to be compensated by software algorithms.

complicated content (say, edges and textures), it tends to demand more sophisticated SR algorithms of higher complexity to improve the visual quality of the output HR frame.

Even though the idea described above appears to be straightforward, there are three major challenges to be solved. First, we need to establish the proper correspondence between multiple LR input frames and the target HR output frame. This demands an effective and robust image registration technique. Second, we need an effective approach for block classification so that we can choose the proper SR technique. Third, there may be artificial block boundaries generated as a result of block partitioning taken by this approach. We will need some post-processing techniques to stitch the blocks of HR output frame seamlessly. As compared with the previous work, where the same SR technique is applied to the whole frame, this new approach can save computation in those areas without critical features. Furthermore, the computational and memory requirements for block processing are much lower than for frame processing due to a great reduction in the dimension of variables. Thus, this new approach is expected to be significantly faster if the above three obstacles can be properly addressed.

Recent Developments in the Image/Video Mosaicking Technique

Image/video mosaicking is the process of stitching two or more images/videos taken by different cameras from different viewpoints. Applications of image/video mosaic techniques can be found in computer vision, pattern recognition, and remotely sensed data processing. When input image/video content is taken from different viewpoints, sampling times, and sensors, image registration is needed to integrate these image/video tiles together. Over the past few decades, much research has been done to obtain an image/video mosaic. For an extensive survey of previous work, we refer to [4]. Although previous work may yield good results, some of them require specific instruments and they were primarily developed in the pixel (or spatial) domain, which is computationally expensive in general. Since Motion JPEG, MPEG, and H26x coding standards all adopt the DCT representation during the coding process, given multiple compressed inputs of image/video, it is desirable to conduct the registration process directly in the DCT domain to generate the corresponding compressed image/video mosaic.

Several color matching algorithms that compensate color differences from two input image sequences captured by different cameras were studied before. Some of them are conducted in the pixel domain, while others can be carried out in the DCT domain. Two techniques, histogram matching and polynomial contrast stretching, in [9] can eliminate the seam lines due to color difference successfully and save more than 80 percent of computation cost as compared with the pixel domain techniques while maintaining a high quality output. Moreover, when comparing these two approaches, the polynomial-based contrast stretching method outperforms the histogram matching method in terms of the processing time and memory requirement.

The DCT-domain registration techniques for MPEG video were applied to the context of video mosaic with both indoor and outdoor scenes in [10]. Both of them can achieve certain quality of composition while the computational cost can be reduced significantly in comparison with the pixel-domain based techniques. A post-processing technique called hybrid block/pixel level alignment, which is partially conducted in the pixel domain and partially in the DCT domain, is introduced to enhance alignment accuracy. In the hybrid block/pixel level alignment methods, an algorithm is proposed to detect corner blocks based on the DCT coefficients. Thus, instead of performing inverse DCT transform on the whole image, only corner blocks which are detected in the DCT domain are converted back to the spatial domain for alignment fine-tuning. This hybrid technique can achieve perfect alignment at the cost of slightly increased complexity. Furthermore, the DCT-domain registration technique is more robust than the pixel-domain registration in the presence of noise since the DC coefficients of a DCT block are actually obtained by pixel averaging over an 8×8 block, which corresponds to a lowpass filtering operation.

It is worthwhile to emphasize that a complete treatment of image/video mosaicking in the DCT domain is not an easy job. There are many open problems such as the handling of motion vectors and residual images of P and B frames, the synchronization of image frames in the temporal domain, etc. It appears that the DCT-domain video mosaicking technique will be more likely to be applied to the motion JPEG video format rather than the MPEG video format in the near future.

Conclusion and Future Work

The demand on flexible media content conversion across heterogeneous capture and display terminals will continue to grow when more and more terminals are linked by IP networks. Users will not be satisfied by rich functionalities of an isolated device only, but will also expect compatibilities between different terminals so that they can get the best output based on the platform available. The difference between terminals has to be compensated by software algorithms to facilitate multimedia data migration from one machine to the other with minimal degradation.

In this article, two related active research topics, i.e., super resolution and image/video mosaicking, have been highlighted and some new research directions have been pointed out. The emphasis will be on the balance of computational complexity and resultant image/video quality. Unlike traditional SR methods, we considered a content adaptive SR approach that segments the whole image into blocks and processes each of them differently according to their content property. A new image/video mosaicking technique conducted in the DCT domain was also discussed. Algorithms for color matching and image registration were developed in the DCT domain to lower computation complexity. However, this approach has its limitation in applicability. More research efforts toward an integrated networking system that offers flexibility and compatibility among heterogeneous terminals are expected in the near future.

REFERENCES

[1] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[2] M.-Y. Shen and C.-C. Jay Kuo, "A Robust Nonlinear Filtering Approach to Inverse Halftoning," *J. Visual Commun. and Image Representation*, vol. 12, no. 1, Mar. 2001, pp. 84–95.

[3] A. Telkap, *Digital Video Processing*, Prentice Hall, Upper Saddle River, NJ, 1995.

[4] B. Zitova and Jan Flusser, "Image Registration Methods: a Survey," *Image and Vision Computing 21*, 2003, pp. 977–1000.

[5] M. Elad and A. Feuer, "Restoration of a Single Super-Resolution Image from Several Blurred, Noisy, and Undersampled Measured Images," *IEEE Trans. Image Proc.*, vol. 6, no. 12, Dec. 1997.

[6] A. J. Patti, and Y. Altubasak, "Artifact Reduction for Set Theoretic Super Resolution Image Reconstruction with Edge Adaptive Constraints and Higher-Order Interpolants," *IEEE Trans. Image Proc.*, vol. 10, no. 1, Jan. 2001.

[7] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-Resolution Reconstruction of Compressed Video using Transform-Domain Statistics," *IEEE Trans. Image Proc.*, vol. 13, no. 1, Jan. 2004.

[8] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe Demosaicing and Super-Resolution of Color Images," *IEEE Trans. Image Proc.*, vol. 15, no. 1, Jan. 2006.

[9] M.-S. Lee, M.-Y. Shen, and C.-C. Jay Kuo, "Color Matching Techniques for Video Mosaic Applications," *IEEE Int'l. Conf. Multimedia and Expo (ICME)*, Taipei, Taiwan, June 27–30, 2004.

[10] M.-S. Lee, M.-Y. Shen, and C.-C. Jay Kuo, "A DCT-Domain Video Alignment Techniques for MPEG Sequences," *IEEE Int'l. Wksp. Multimedia Sig. Proc. (MMSP)*, Shanghai, China, Oct. 30–Nov. 2, 2005.

BIOGRAPHIES

MING-SUI LEE (mingsuil@usc.edu) received her B.S. degree in mathematical sciences from National Cheng-Chi University, Taiwan, in 1999. Then she received the M.S. degree in electrical engineering from the University of California, Los Angeles (UCLA) and the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 2002 and 2006, respectively. Her research interests include image/video mosaicking, compressed-domain image/video processing, and super resolution techniques.

MEIYIN SHEN (meiyinsh@sipi.usc.edu) received her B.S. degree in electrical engineering from National Cheng-Kung University, Taiwan, in 1990. She received the M.S. and Ph.D. degrees from the University of Southern California (USC), Los Angeles, in 1993 and 1999, respectively, also in electrical engineering. Since 2002, she has been a postdoctoral researcher at USC. Her research interests cover a broad range of multimedia technologies, including image/video compression and processing, rate control, image/video mosaicking, and 3D mesh compression.

AKIO YONEYAMA (yoneyama@kddilabs.jp) received the B.E. and the M.E. degrees in electrical engineering from Keio University, Japan, in 1992 and 1994, respectively. He joined KDD in 1994. Since 1996, he has been with KDDI R&D Laboratories Inc. From 2002 to 2003 he was a visiting researcher at the University of Southern California. His current research interests include video compression, image processing.

C.-C. JAY KUO (cckuo@sipi.usc.edu) is Director of the Signal and Image Processing Institute and Professor of electrical engineering, computer science and mathematics at the University of Southern California. He is a Fellow of the IEEE and SPIE. He has published more than 800 papers in international conferences and journals and seven books. He received the National Science Foundation Young Investigator Award and Presidential Faculty Fellow Award in 1992 and 1993, respectively.

Unlike traditional SR methods, we considered a content adaptive SR approach that segments the whole image into blocks and processes each of them differently according to their content property.