

# DCT-Domain Image Registration Techniques for Compressed Video

Ming-Sui Lee, Meiyin Shen, Akio Yoneyama\* and C. -C. Jay Kuo  
Integrated Media Systems Center and Department of Electrical Engineering  
University of Southern California, Los Angeles, CA90089-2564  
E-mails: mingsuil@usc.edu and {meiyinsh, cckuo}@sipi.usc.edu  
\* KDDI R&D Laboratories Inc., Japan  
\* E-mail: yoneyama@kddilabs.jp

*Abstract* — An image registration technique for compressed video such as motion JPEG or the I-picture of MPEG is investigated in this paper. The proposed technique is based on the DCT (Discrete Cosine Transform) coefficient matching. First, the coarse edge features are extracted by applying several edge detectors to luminance DC coefficients. Each detector generates one difference map for a single input image. A threshold is set up for each difference map to produce a binary map. Then, the alignment parameters are determined based on the binary maps of both input images generated by the same detector. Finally, the actual displacement in the pixel domain is calculated by averaging parameters from all detectors. It is shown by experimental results that the proposed method reduces the computational cost of image registration dramatically as compared with the pixel domain and edge-based DCT domain registration techniques while achieving certain quality of composition.

## I. INTRODUCTION

Image registration is the process of aligning two or more images taken by different cameras. Applications of image registration techniques can be found in computer vision, pattern recognition, and remotely sensed data processing. Although this topic has been studied for several decades, most techniques were primarily developed based on the pixel domain information. Here, we study image registration techniques for multiple captured video clips compressed by motion JPEG and MPEG. To speed up the registration process, it is desirable to perform image registration directly in the DCT domain to shorten the processing time.

The main advantage of image registration in the DCT domain is that the computational complexity can be significantly reduced. In our proposed method, the coarse edge features are extracted by applying several edge detectors to luminance DC coefficients. Then, a threshold is set up to generate a binary image that contains some specific properties. Finally, the displacement parameters for image alignment are determined based on the match of the binary images.

The rest of this paper is organized as follows. The

background of this research is presented in Section 2. The proposed registration algorithm in the DCT domain is described in Section 3. Experimental results are given in Section 4 to demonstrate that the DCT-domain algorithm can achieve good composition quality while dramatically reducing the computational complexity. Finally, concluding remarks are drawn in Section 5.

## II. BACKGROUND REVIEW

Image/video mosaic, 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. Recently, the advance of digital camera techniques and the affordable pricing of the corresponding commercial products make it possible for consumers to generate their own multimedia contents easily. When the input image/video contents are taken from different viewpoints, sampling times and sensors, image registration is demanded to integrate these image/video tiles together. Over the past few decades, a lot of work has been conducted to obtain image/video mosaic. For an extensive survey of previous work, we refer to [1][2].

Generally speaking, the image registration technique consists of two major steps: feature detection and feature matching. Feature detection can be done either manually or automatically. Since human eyes are sensitive to geometrical patterns, it is straightforward for people to choose the matched patterns. However, it is desirable to develop an automatic feature selection process based on the particular application context 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 in [3] to extract the local maxima. 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 region of input images to find the best match. This approach

is more suitable for images that do not have many details. However, its computational complexity is high. Once the feature information is available, the next step is to find the optimal correspondence between image tiles. Feature matching is a process to determine the relationship between similar objects contained by different images. This can be achieved by finding the spatial relations among the extracted features.

Although the above methods lead to good results, they are primarily developed in the pixel (or spatial) domain, which is computationally expensive. Given multiple compressed video inputs, it is desirable to conduct the registration process in the DCT domain to generate the corresponding compressed image/video mosaic, since most of coding standards adopt the DCT representation during the coding process. The panorama can be either stored for later use or decoded for display. This is however by no means a straightforward task since image registration is inherently a spatial domain job.

For a more generic scenario, we may consider multiple video sources captured by an arbitrary number of cameras with different parameter settings. There arise many challenging problems in creating video mosaic, including temporal synchronization, focal length readjustment, image registration, and color difference compensation. The discrepancies among smaller video tiles have to be resolved for seamless composition. This work reports our effort towards this general goal. In this paper, we focus on the registration of two arbitrarily translated images in the DCT domain under several simplifying assumptions. For example, temporal synchronization and focal length distortion problems have been well solved. Besides, since this is a fairly complex problem, we are only concerned with the registration of the intra-coded frames (i.e. the I-picture) for MPEG video or image frames from the Motion JPEG format in this work. The registration of P- or B-pictures will be addressed separately in the future.

### III. PROPOSED DCT-DOMAIN IMAGE REGISTRATION TECHNIQUE

Here, the registration problem is addressed under the assumption that all other factors are well compensated in advance. That is, all image frames in the input video sources are temporally synchronized, and the focal lengths of their lens are adjusted to be the same. Moreover, the rotational and zoom-in zoom-out distortions are compensated during the video capturing process. Thus, the registration of two image frames that contain only translation displacement in the horizontal and vertical directions is studied in this paper. We deal with the problem in the DCT domain by processing luminance DC coefficients only. Generally speaking, the proposed approach consists of three major steps; namely, edge detection, thresholding and displacement parameter estimation. These three steps are shown in Fig. 1 and will be detailed in the following subsections.

#### 3.1 Edge detection on DC map

For each input image, a DC map of each input image is formed by luminance DC coefficients from all DCT blocks. Note that the size of the DC map is 1/64 of that of the original one. Several edge detectors ( $H1\sim H4$ ) are applied to the DC map. The edge detector adopted here are as follows:

$$H1 = \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix} \quad H2 = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

$$H3 = \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix} \quad H4 = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

$H1$ ,  $H2$ ,  $H3$ , and  $H4$  measure the variation of the image in vertical, horizontal, 45 degree, and 135 degree directions, respectively. Each detector can produce a difference map. There are four difference maps of each input image. That is,  $D_{11}$ ,  $D_{12}$ ,  $D_{13}$  and  $D_{14}$  are the difference maps corresponding to detectors  $H1$  to  $H4$  of image1 while  $D_{21}$ ,  $D_{22}$ ,  $D_{23}$  and  $D_{24}$  are the difference maps corresponding to detectors  $H1$  to  $H4$  of image2, respectively. Note that the values of all above difference maps are normalized to fall between 0 and 1.

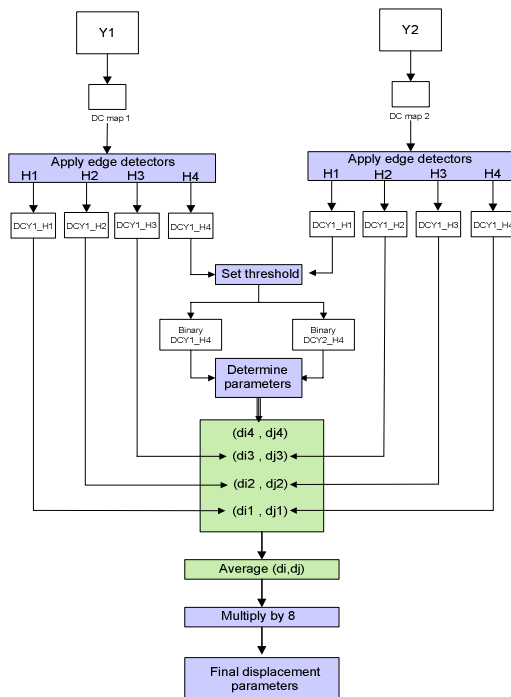


Figure 1. An overview of the proposed approach.

#### 3.2 Thresholding

In this step, a content adaptive threshold is masked on each pair of difference maps. The thresholds are set with the purpose of filtering out some minor changes to speed up the following alignment step. Fig. 2(a) shows the difference maps resulting from applying detector  $H1$  of two images that we intend to align. After setting up the threshold on both maps, their corresponding binary activity maps are shown in Fig. 2(b). As we can see from the binary images,

only the vertical difference is preserved for the estimation of displacement parameters so that the processing time can be reduced.

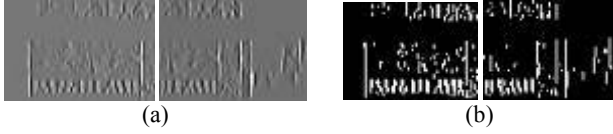


Figure 2. Conversion from a difference map (after applying H1) to a binary activity map: (a) the difference map and (b) the binary activity map

### 3.3 Displacement parameter estimation

Let the sizes of two original input images and their DC maps be  $P_i \times P_j$  and  $N_i \times N_j$ , respectively, where  $N_i = P_i / 8$  and  $N_j = P_j / 8$ . Based on the results obtained from Section 3.2, our goal is to determine the displacement parameters to align the left and right two images. One simple way to achieve the alignment is to compute the cross-correlation between two binary maps,  $B_{11}$  and  $B_{21}$ . The vector,  $(d_{i1}, d_{j1})$ , which leads to the maximal correlation value, gives the optimal displacement in the vertical and the horizontal directions. Similarly,  $(d_{i2}, d_{j2})$ ,  $(d_{i3}, d_{j3})$  and  $(d_{i4}, d_{j4})$  can be derived by following the same procedures. Once those four candidate displacement vectors are available, the optimal parameters  $(d_i, d_j)$  can be computed by checking the best result in the DC coefficient map. Since the horizontal or the vertical size of the binary map is 1/8 of that of the original one, the actual amount of displacement should be scaled up by a factor of 8 with respect to the coordinates of the original input images.

## IV. EXPERIMENTAL RESULT

In this section, we present some experimental results with eight test image pairs as shown in Fig. 3, where (a) and (b) are indoor scenes and (c) to (f) are outdoor scenes (600x448). Each of them has a different complexity, a different type and a different amount of displacement. Please note that the color mismatch does not affect the registration results since the proposed approach does not rely on the color information.



(a) The 1<sup>st</sup> test image pair



(b) The 2<sup>nd</sup> test image pair



(c) The 3<sup>rd</sup> test image pair



(d) The 4<sup>th</sup> test image pair



(e) The 5<sup>th</sup> test image pair



(f) The 6<sup>th</sup> test image pair

Figure 3. The original test sets

### 4.1 Performance Comparison in Processing Time

The comparison of the execution time of the traditional pixel domain approach and the proposed process is shown in Table 1 and Fig. 4. The computational saving comes from two different parts. First, the DCT domain approach avoids the inverse DCT and the forward DCT processes required by the pixel domain approach. Second, the resolution of the pixel domain image pair is much finer than that of the DCT domain image pair, i.e. 64 versus 1. The search for the displacement vector demands much more time. We can see the time saving is around 97% as compared with the traditional pixel domain approach.

TABLE 1. Comparison between the proposed and the traditional one in processing time (sec) – (a) traditional (b) proposed method. (c) and (d) are the savings in terms of seconds and percentages

	1	2	3	4	5	6
(a)	35.22	35.28	39.13	35.45	36.11	37.64
(b)	1.407	1.406	1.437	1.516	1.407	1.422
(c)	34.19	34.25	38.13	34.43	35.08	36.58
(d)	97.07	97.07	97.44	97.13	97.14	97.18

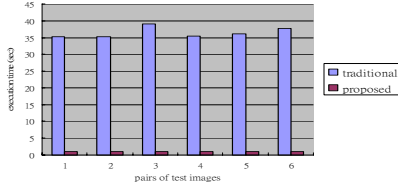


Figure 4. Performance comparison in processing time

#### 4.2 Comparison of Output Image Quality

The registered output image results are compared in Table 2 and shown in Fig. 5. Here we assume that the exact displacement parameters of each pair of test images are known so that we can compute the difference between the actual ones and the ones that are determined by the proposed approach.

Table 2. Comparison between the displacement parameters (di,dj) derived based on the proposed approach and the actual ones (di',dj')

	1	2	3	4	5	6
8*dj1	296	304	304	200	400	504
8*di1	448	600	304	448	520	496
8*dj2	296	304	304	200	408	504
8*di2	448	600	296	448	520	496
8*dj3	304	296	296	200	400	504
8*di3	448	600	296	448	528	496
8*dj4	296	296	296	200	400	504
8*di4	448	600	296	448	528	496
8*dj	298	300	300	200	402	504
8*di	448	600	298	448	524	496
actual dj	300	300	300	200	400	503
actual di	448	600	300	448	524	495
8*(dj-dj')	-2	0	0	0	+2	+1
8*(di-di')	0	0	-2	0	0	+1

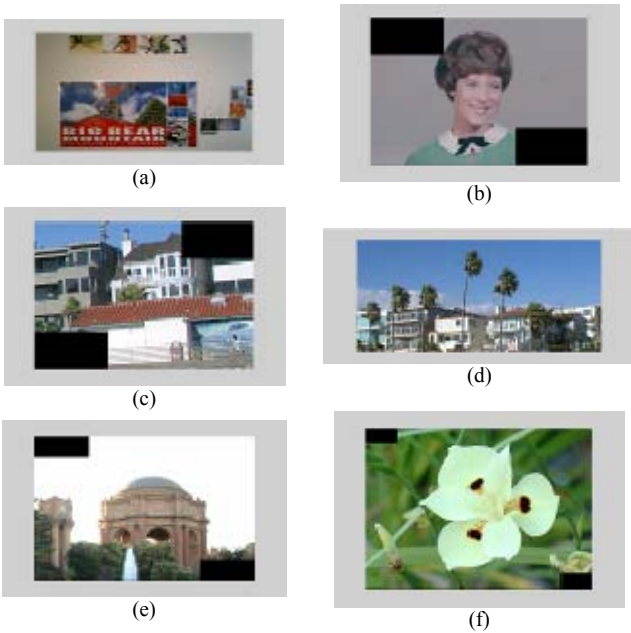


Figure 5. The experimental results based on the proposed method

Table 2 shows that the parameters determined by the proposed method only have +2 or -2 pixels difference compared to the actual displacements. This means that the precision can be reached to the sub-block. This is due to considering several sets of displacement parameters generated by different edge detectors. Each detector indicates the changes of the image regarding to a specific direction. The quality of the output is improved step by step after applying the detectors one by one. Averaging all the parameters would give us a better result. Therefore, if more appropriate detectors are applied to the images, the result would be enhanced even more so that the pixel-accuracy or the sub-pixel accuracy is possible. However, applying more detectors requires more processing time. With this consideration, we have to find a balance between the processing time and the accuracy of the alignment.

#### V. CONCLUSION

A DCT-domain technique for registering two input images with an arbitrary displacement was proposed. It was demonstrated by experimental results that the proposed method saves more than 95% of the computational cost as compared to the pixel domain techniques while the visual quality of the stitched image remains about the same. The performance of the proposed method is consistent regardless of indoor or outdoor scenes. Although it is a block-based processing technique, the quality of the alignment has been enhanced to quarter-block (2-pixel) accuracy.

#### ACKNOWLEDGEMENT

The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, under the Cooperative Agreement No. EEC-9529152, and in part by KDDI Laboratories, Inc. 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 and KDDI Laboratories, Inc.

#### REFERENCES

- [1] Lisa G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, 24(4): 325-376, December 1992.
- [2] Barbara Zitova and Jan Flusser, "Image registration methods: a survey," *Image and Vision Computing* 21, pp. 977-1000, 2003
- [3] J.- W. Hsieh., H. -Y. M. Liao, K.- C. Fan, M. -T. Ko, and Y. -P. Hung, "Image registration using a new edge-based approach," *Computer Vision and Image Understanding*, vol. 67, No.2, pp. 112-130, Aug. 1997.
- [4] Richard Szeliski, "Video mosaics for virtual environments," *IEEE trans. Computer Graphics and Applications*, Vol. 16, Issue 2, pp. 22-33, 1996
- [5] Richard Szeliski and Heung-Yeung Shum, "Creating full view panoramic image mosaic and environment maps," *Proceedings of the SIGGRAPH 1997*, pp. 251-258, 1997
- [6] Aditi Majumder, Gopi Meenakshisundaram, W. Brent Seales and Henry Fuchs, "Immersive teleconferencing: a new algorithm to generate seamless panoramic video imagery," *Proceedings of the Seventh ACM International Conference on Multimedia*, October 30 - November 5, 1999.