Digital Image Stabilization Based on Circular Block Matching

Lidong Xu, Xinggang Lin, *Senior Member,* IEEE

Abstract —*In this paper, a novel digital image stabilization algorithm based on circular block matching is proposed to estimate the global scaling, rotational and translational motion parameters between each two consecutive frames in jitter videos. In the algorithm, the rotation-invariant-features based circular block matching is proposed for the first time to estimate the local true motion vectors for the centers of all considered circular blocks. With these local motion vectors and their corresponding block positions, a linear system is constructed and solved by repeated least-squares to generate the global motion parameters (GMP) which are then refined by a gradient descent based iteration scheme. The experimental results demonstrate that the proposed DIS technique can generate more precise GMP and deal with larger motions compared with the ever presented optical flow based digital image stabilization technique.*[1]

Index Terms — Circular Block Matching, Digital Image Stabilization, Repeated Least-squares, Gradient Descent.

I. Introduction

Image stabilization techniques have been studied for decades to improve visual qualities of image sequences captured by compact and light weight digital video cameras. When such cameras are hand held or mounted on unstable platforms, the captured video generally looks shaky because of undesired camera motions. Unwanted video vibrations would lead to degraded view experience and also greatly affect the performances of applications like video coding [1]-[4], video surveillance [5], etc. Many image stabilization schemes have been proposed. Among these schemes, digital image stabilization (DIS) is the most popular and in several aspects outperforms traditional electronic image stabilizer (EIS) and optical image stabilizer (OIS), both of which are hardware dependent. DIS can be implemented on hardware that is smaller than mechanical devices used for EIS and OIS, and can operate either online or offline based on different applications [6]. DIS can also produce more precise stabilization results and operate either individually or be combined with EIS or OIS to improve the stabilization performance of EIS or OIS [7].

Generally a DIS system consists of two principal units: motion estimation (ME) and motion correction (MC), as shown in Fig. 1. The ME unit estimates the global motion parameters (GMP) between every two consecutive frames of the input image sequence. With these GMP, the MC unit then generates the correcting motion parameters (CMP) needed to compensate for the jitter of a frame and warps the frame to create a more visual stable image sequence. If no intended camera motion exists, CMP is equal to the accumulated motion parameters (AMP) of GMP. Otherwise, CMP is produced by passing AMP through a low-pass filter so as to remove high frequency jitters and retain smooth intended camera motions.

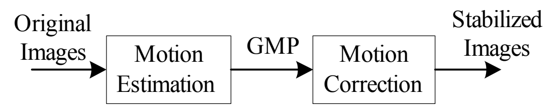


Fig. 1. Block diagram of DIS system.

ME unit plays the most important role in DIS system and its estimation precision is a decisive factor for the overall stabilization performance of the system. The challenges of ME we are facing may include: 1) the complex motions in the real environment, not only translation, but also rotation and scaling, 2) large rotational motions, 3) the influence of the foreground, and 4) the smooth background to result in false motion vectors. This paper aims at presenting a novel circular block matching based ME algorithm which can accurately estimate the global translational, rotational and scaling motions even when the motions are very large. In this algorithm, the contributions can be summarized as: 1) rotation invariant features based scheme, 2) circular block matching, and 3) gradient descent based GMP refinement.

For estimating the GMP of current frame, feature-based block matching is first performed for considered circular blocks in the frame to generate the local motion vectors (MVs) of the block centers. With these local MVs and their corresponding block positions in current and reference frames, a linear system is then constructed and solved by repeated least squares (RLS) to generate the coarse GMP which is finally refined by gradient descent based iteration process. We conducted both simulations and real experiments to verify the performance of our DIS system. The simulation results show that the estimation errors of scaling, rotational and translational motion parameters are less than 0.005, 0.08 degree and 0.05 pixels respectively

The remaining part of this paper is organized as follows. Section II gives a brief review of the previously proposed ME techniques for DIS. The details of the proposed ME algorithm is described in section III. Section IV presents some experimental results and Section V concludes the paper.

II. Review of the ME Techniques for DIS

According to the motion models being considered, the already proposed global ME techniques for DIS can be roughly divided into two categories: 1) two-dimensional stabilization (2D) techniques which deal with translational

[1] This research is jointly supported by National Natural Science Foundation of China under Grant No. 60472028 and Doctoral Program Research Foundation of Ministry of Education of China under Grant No. 20040003015

Contributed Paper
Manuscript received January 15, 2006

jitter only and 2) multi-dimensional (MD) stabilization techniques which aim at stabilizing more complicated fluctuations in addition to translation. Most of the existing algorithms [1], [3]-[6], [8]-[22] fall into the first category because translation is the most commonly encountered motion and the complexity of estimating translation parameters is relatively low for real-time stabilization. In the second category, the majority of algorithms [2], [23]-[26], [28], [29] considered similarity motion model, while a few algorithms considered affine motion model [7], [30] or perspective motion model [27], [31].

A. 2D Stabilization Techniques

For 2D stabilization, block matching with mean absolute difference (MAD) or mean square error (MSE) criterion is the traditional way to detect the global motion vector between two successive frames. In this scheme, each image is divided into squares or rectangles of a certain dimension, and motion search is then performed to find the MV for each block in current frame. With these local MVs, clustering methods [1], [4], [9], [11] are used to find the dominate MV which is regarded as the global MV.

Due to the heavy computational cost of block matching with full pixel information, several schemes have been proposed to speed up the process with only partial pixel information, e.g. bit-plane matching [12] and gray-coded bit-plane matching [16] schemes, representative point matching (RPM) schemes [3], [6], [13]-[15], [20] and selected areas matching (SAM) schemes [3], [6], [9], [12], [16]. In the SAM scheme, the selected areas were generally located on the borders and/or corners of the current frame, based on the assumption that in most cases, the foreground objects are located on the center of the image and the surroundings of the image are the best candidates for background motion detection [6]. Therefore, SAM technique can not only reduce the computation complexity, but lower the impact of foreground object movement to the estimation accuracy of background motion.

Besides block matching techniques with full or partial pixel information, some feature matching based techniques, e.g. characteristic curves matching [10], edge pattern matching [21], [22] and phase correlation [8], have also been proposed for 2D image stabilization. Like other 2D stabilization techniques, they are lack of adaptation to high dimensional motions.

B. MD Stabilization Techniques

Theoretically, pixel based block matching schemes cannot be used for MD stabilization because the pixels in the block have different translational offsets. But when the motions are all very small except translation, the block matching results can be approximately regarded as the MVs of the block centers [2].

The more effective methods for MD stabilization are feature based ones. Several optical flow based techniques have been proposed for image stabilization [7], [24], [26], [30]. The optical flow constraints adopted in these algorithms are the first-order approximation of Taylor series, which means that motions can not be very large. Otherwise, the ignored higher-order terms of Taylor series can lead to significant estimation errors [32]. To reduce the estimation error, a multi-resolution, iterative process algorithm was proposed in [7]. Large translational motion in highest resolution images are scaled down to small motions in the lowest resolution images. Therefore estimating from the lowest resolution images to the highest resolution images in a coarse-to-fine fashion can deal with large translational motion. However this algorithm can not manage large rotational motion because the rotational motions in the highest and lowest resolution images are the same.

Feature points tracking schemes can deal with complicated motions. In [25], some feature points are extracted in the first frame and then tracked in every subsequent frame of the sequence, which implies that this algorithm does not work when intended motion exist because the feature points will vanish when the intended motions are accumulated. In [28], after feature points extraction, feature tracking is performed by traditional block matching scheme. Therefore, it can not deal with large rotational motion. Some stabilization techniques using multiple visual cues, e.g. lanes and vanishing points, have been presented in [27] and [31]. But these visual cues based algorithms are only suitable for special applications where the considered cues appear in each frame.

Some DIS algorithms utilize the information in transformed domain. An algorithm based on stationary wavelet transform was proposed in [23]. It first estimated the translational motion by projecting the vertical and horizontal details of stationary wavelet decomposition in level 2. Using the translation parameters as the initial value (the initial rotation angle and scaling factor are set to 0 and 1 respectively), an iterative scheme was proposed to estimate similarity motion parameters. The problem is that the iteration may fall into local optimal solution if the true rotation angle or scaling factor is very large. In [29], FFT phase correlation and log-polar magnitude spectra representation are utilized to obtain translation, rotation and scale parameters. The two disadvantages of this algorithm are: 1) the resolution of the estimated rotation angle is 1.4 degree which is not very precise, and 2) the rotational jitter can not be removed entirely if the rotational motion does not always have its origin at the center of the image frame.

In this paper, we proposed a novel MD image stabilization algorithm based on circular block matching scheme. This algorithm can efficiently deal with rotational and scaling motions in addition to translational motion, even when the sequence has both large rotational motion and intended panning motion, where the ever proposed MD image stabilization algorithms can not work effectively.

III. CIRCULAR BLOCK MATCHING BASED ME

Generally, in a captured video sequence, the rotational and translational motions between two successive frames are small.

But when intended rotation and translation exist, the global motions between two successive frames may be very large depending on the camera's moving speed. Rectangular block motion search is an effective matching scheme to deal with large translational motions but it does not work when rotation motion exceeds a certain small range, which will be shown in section IV. Considering that circle is the only shape which can contain the same contents in current frame and reference frame at any rotation angle, we adopt circular block for block matching to guarantee to produce true local MV for the pixel at block center no matter how large the rotation angle is. On the other hand, pixel matching scheme is not suitable for circular block matching because of the rotation motion. Therefore feature matching scheme is proposed in our DIS system and the selected features should be rotation invariant.

The detailed block diagram of the proposed DIS system is shown in Fig. 2. The ME unit in this system consists of three modules: feature image generation, circular block matching and GMP generation. After feature image generation, rotation-invariant-features based circular block motion search is performed between two consecutive incoming frames to produce the local MVs for the centers of considered circular blocks in current frame. With these local MVs and their corresponding block positions in current frame and reference frame, a linear system is constructed. The GMP is generated by solving the linear system with repeated least-squares and then refined by a gradient descent based iteration scheme. The MC unit consists of CMP generation module and image warping module. In addition, an image buffer is utilized in this system to store and provide image data for the processing of other modules. The techniques of MC unit can be referred to the previous paper [33]. The remaining part of this section describes the detailed techniques of the proposed ME algorithm.

A. Feature Images Generation and Feature Extraction

Many features can be utilized for circular block matching so long as they are rotation invariant. Generally, the more the features are used, the more precise the matching results are, but the higher the computational complexity is. To guarantee sufficient matching precision and relatively low matching complexity, four features, which may not be optimum but are demonstrated to be effective, are experimentally selected in the proposed DIS system. These four features are extracted from two feature images of the corresponding frame (frame t): intensity image I_t and gradient image G_t. To ensure the rotation invariant property of extracted features, the gradient operator should also be rotation invariant. Therefore, Laplacian of Gaussian (LoG) operator is adopted to generate the gradient image because Laplacian operator is rotation invariant [28] and Gaussian filter is circularly symmetric [34]. Practically, only strong edges are trusty for block matching, so a step function is used to remove weak edges from G_t:

$$G_t(x,y)=\begin{cases} G_t(x,y) & if\ \ G_t(x,y)\geq T \\ 0 & if\ \ G_t(x,y)<T. \end{cases} \tag{1}$$

Where, T is a threshold which is empirically set to 15 in the proposed DIS system.

The four adopted features, f_1, f_2, f_3 and f_4, are extracted by (2) from the block area (BA) which contains N pixels.

$$\begin{cases} f_1=\sum\limits_{(x,y)\in BA}(G_t(x,y)>0)\ ?\ 1\ :\ 0 \\ f_2=\dfrac{1}{N}\sum\limits_{(x,y)\in BA}G_t(x,y) \\ f_3=\dfrac{1}{N}\sum\limits_{(x,y)\in BA}I_t(x,y) \\ f_4=\dfrac{1}{f_1}\sum\limits_{\substack{(x,y)\in BA\& \\ G_t(x,y)>0}}[(x-\bar{x})^2+(y-\bar{y})^2]. \end{cases} \tag{2}$$

Where,

$$\bar{x}=\frac{1}{f_1}\sum_{\substack{(x,y)\in BA\& \\ G_t(x,y)>0}}x,\quad \bar{y}=\frac{1}{f_1}\sum_{\substack{(x,y)\in BA\& \\ G_t(x,y)>0}}y. \tag{3}$$

Here, f_1 means the number of strong edge pixels, f_2 and f_3 stand for the average gradient and average intensity respectively, f_4 represents the coordinate dispersion of the strong edge pixels. It is obvious that these four features, which form a feature vector as shown in (4), are all rotation invariant.

$$\vec{f}=(f_1,\ f_2,\ f_3,\ f_4)^T \tag{4}$$

B. Circular Block Matching

For a circular block in current frame, block matching is performed by two-dimensional motion search to find the best match block in reference frame and obtain the local MV of the block center. Block motion search process is depicted in Fig. 3. Circle A is an appointed circular block in current frame, B' is a circular block in reference frame whose center coordinate is the same as that of A in current frame. The rectangular area bounded with dotted line is the search window, and C' is a

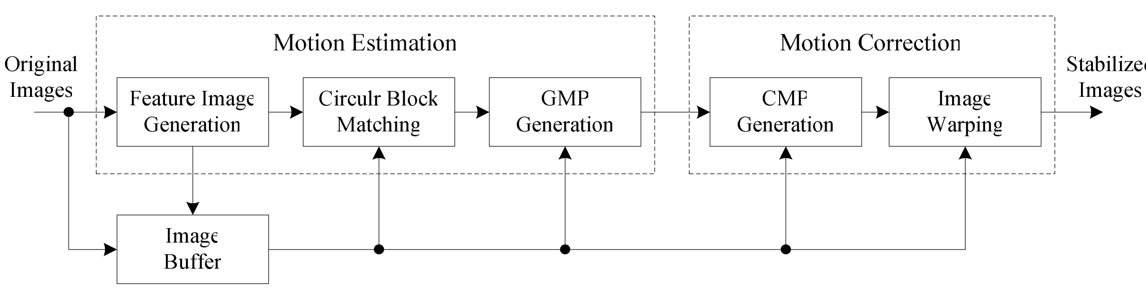


Fig. 2. Block diagram of the proposed DIS system

candidate block in the search window. The 1-norm of the difference vector between the feature vectors of block A and block C' is used as the matching criterion. Assuming that in the candidate block set S, A' is the best matching block to A, then (mv_x, mv_y) shown in Fig. 3 is the local MV of the center of block A, and block A' can be represented by the following optimization issue:

$$A' = \arg\min_{C' \in S} \{ \| \bar{f}_A - \bar{f}_{C'} \|_1 \}. \tag{5}$$

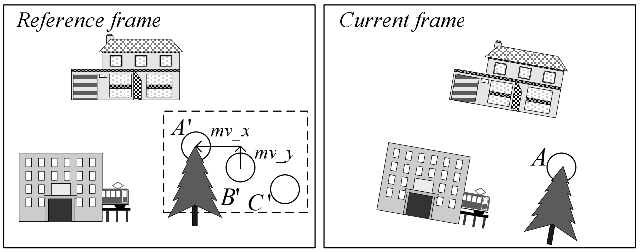


Fig. 3. Circular block motion search process

Multiple circular blocks in current frame are used to perform block matching to obtain local MVs. The general way to select circular blocks is to divide the current frame into n by n squares and use the inscribed circles of every square as the considered circular blocks. For each circular block, full search is the optimal solution to detect the local MV within the search window. However, block motion search is a very time-consuming module which may prevent the proposed DIS system from being utilized in real-time applications. Therefore two methods are schemed to degrade the processing complexity. One is to reduce the number of circular blocks used for matching, and the other is to use fast motion search instead of full search.

For the sake of reducing number of blocks, only part regions in current frame are used for block selection. These regions are located near the borders of the frame, as shown in Fig. 4, based on the assumption that in most cases, the foreground object is located on the center of the frame whereas the area near the frame borders is background [6]. The reason that the selected regions are not on the borders is that the local MVs of border blocks are not quite reliable.

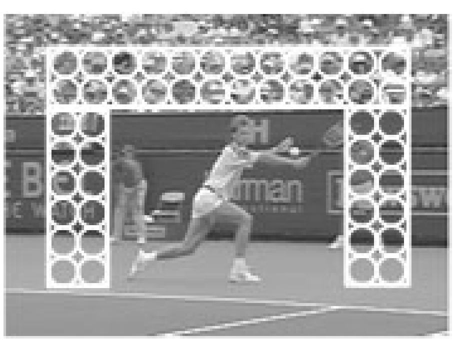


Fig. 4. Regions for circular block selection.

Fast motion search, which is popularly used in video compression, reduces the matching complexity by searching a subset of the searching window. In video compression, motion search aims at minimizing the mean absolute prediction errors rather than finding the true MVs. So a fast motion search algorithm for video coding is regarded as being efficient so long as it can greatly improve the coding speed with only slight compression performance loss, even if it generates wrong MVs. Quite the contrary, in DIS system, motion search aims at finding the true local MVs to ensure the precision of the generated GMP. Therefore, the fast motion search should output the MV the same as or very close to that produced by full search. A fast motion search scheme is proposed in our DIS system by utilizing the extracted rotation invariant features f_1 of circular blocks.

The main idea of the proposed fast motion search is that in searching process, if the difference between the feature f_1 of the block in current frame and that of a candidate block in search window is bigger than a threshold, then the neighboring candidate blocks of this candidate block can be removed from the search window, because they are unlikely to be the best matching block. How to set the threshold, as well as what kind of neighboring candidate blocks should be removed, is described as follows.

Let f_1^O be the feature of block O in current frame, $f_1^{C_1}$ and $f_1^{C_2}$ be the features of two candidate overlapping blocks C_1 and C_2 in search window, f_1^C be the feature of the best matching block C, r be the radius of circular blocks and $2d$ be the center distance between C_1 and C_2, as shown in Fig. 5.

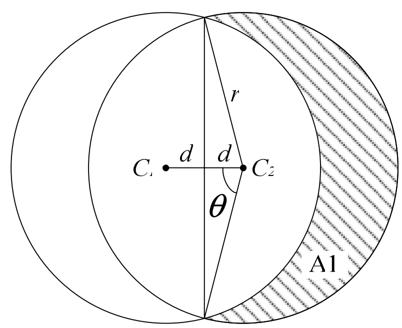


Fig. 5 Relation between two candidate circular blocks

In Fig. 5, the ratio k of the shaded area A1 to the area of the whole circle is defined by:

$$\begin{aligned} k &= \frac{\pi r^2 - 2(\theta r^2 - d\sqrt{r^2 - d^2})}{\pi r^2} \\ &= 1 - \frac{2}{\pi}[\cos^{-1}(\frac{d}{r}) - \frac{d}{r}\sqrt{1 - (\frac{d}{r})^2}]. \end{aligned} \tag{6}$$

The relation between k and d/r is approximately linear when d/r is no more than 0.4, as shown in Fig. 6. So k can be approximately represented by:

$$\begin{cases} k = 5d/4r & \text{if } d \le 0.4r \\ k < 5d/4r & \text{if } d > 0.4r. \end{cases} \tag{7}$$

Then we can get:

$$| f_1^{C_1} - f_1^{C_2} | \le k\pi r^2 \le \frac{5\pi dr}{4}. \tag{8}$$

In addition, let C_1 be the block currently being checked and:

$$\begin{cases} | f_1^O - f_1^{C_1} | = k_1 \pi r^2 \\ | f_1^O - f_1^C | < \Delta \pi r^2. \end{cases} \tag{9}$$

Where, Δ is a pre-defined small factor, then we can get:

$$\begin{aligned} | f_1^{C_1} - f_1^C | &= | (f_1^O - f_1^{C_1}) - (f_1^O - f_1^C) | \\ &\geq | f_1^O - f_1^{C_1} | - | (f_1^O - f_1^C) | > (k_1 - \Delta)\pi r^2. \end{aligned} \tag{10}$$

Comparing (8) and (10), we can draw the conclusion that C_2 can not be the best matching block if:

$$(k_1 - \Delta)\pi r^2 > \frac{5\pi dr}{4}, \tag{11}$$

i.e.

$$d < \frac{4(k_1 - \Delta)r}{5}. \tag{12}$$

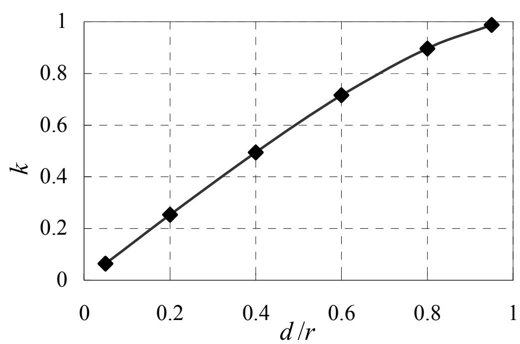


Fig. 6. Relation between k and d/r.

The fast motion search can be achieved by utilizing the results analyzed above. Namely, after checking a candidate block C_1, its neighboring candidate blocks which satisfy (12) can be removed from the search window. Obviously, this fast motion search can generate the local MV almost the same as that generated by full search.

The drawback of block matching is that not all local MVs are reliable. It is likely to produce wrong MV that searching a homogenous block on homogenous area. Therefore, any circular block in current frame whose f_1 is less than a threshold is not used for block matching since it contains too little edge information.

C. GMP Generation

The global translational, rotational and scaling motions are considered in the proposed DIS system. These motions can be modeled by the following similarity motion model:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}. \tag{13}$$

Where, (a, b, c, d) are the global motion parameters, (x, y) the coordinate in the current frame and (x', y') the coordinate in the reference frame.

Assuming that after circular block matching, we have n local MVs (mv_x_i, mv_y_i) together with their corresponding image positions (x_i, y_i) and (x_i', y_i') for i= 1, 2, …., n in two consecutive frames. Combing the similarity motion model with these local MVs and their corresponding image positions, a linear system, as shown in (14), can be constructed. The GMP can be produced by solving the linear system.

$$\begin{pmatrix} x_1 & -y_1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & -y_n & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_n & x_n & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \\ y_1' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} x_1 + mv\_x_1 \\ \vdots \\ x_n + mv\_x_n \\ y_1 + mv\_y_1 \\ \vdots \\ y_n + mv\_y_n \end{pmatrix}. \tag{14}$$

Generally, this linear system has no exact solution because it is an ill-posed problem. But least-squares method can be used to produce the approximation of the GMP by minimizing the sum of squared errors (*SSE*) defined as follows:

$$SSE = \sum_{i=1}^{n} [(ax_i - by_i + c - x_i')^2 + (bx_i + ay_i + d - y_i')^2]. \tag{15}$$

A few odd local MVs may be generated in circular block matching process if some blocks in current frame are selected from the area containing the foreground or consisting of repeated or similar patterns. These odd local MVs are very harmful to the estimation results of GMP if they are used in least-squares method. Therefore, an RLS algorithm, which is named trimmed least squares in [24], is used to reduce the affection of the odd local MVs.

In RLS algorithm, the least-squares is first performed to generate an estimated GMP $(\hat{a}, \hat{b}, \hat{c}, \hat{d})^T$ with all the local MVs and the corresponding image positions. Then the squared estimation error (*SEE*) of each position is defined as:

$$SEE_i = (\hat{a}x_i - \hat{b}y_i + \hat{c} - x_i')^2 + (\hat{b}x_i + \hat{a}y_i + \hat{d} - y_i')^2. \tag{16}$$

A local MV is regarded as being odd if its corresponding *SEE* satisfies:

$$SEE_i > \frac{T^2}{n} \sum_{j=1}^{n} SEE_j, \tag{17}$$

where T is a constant which can be empirically set to 1.5. Then the detected odd MVs are discarded and the remainders are used for next least-squares. The above process is repeated until no odd local MV is detected any more. The output of the last least squares is regarded as the final output of RLS.

Though the output parameters of RLS are very close to the true GMP, some small differences may still exist between them because: 1) the local MVs are integers which may be not very precise, and 2) the local MVs may be slightly affected by scaling motion. To refine the output of RLS, a gradient descent based iteration algorithm is utilized. The iteration is performed to minimize the merit function E defined as:

$$E = \sum_i e_i^2 = \sum_i [I_t(x_i, y_i) - I_{t-1}(x_i', y_i')]^2. \tag{18}$$

Let $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T = (a, b, c, d)^T$, then E is a function of β and can be well approximated by its second order Taylor series [35]:

$$E(\beta) = \eta - \zeta^T \cdot \beta + \frac{1}{2} \beta^T \cdot \mathrm{A} \cdot \beta. \tag{19}$$

Where η is a constant, $\boldsymbol{\zeta}$ is a gradient vector and **A** is a Hessian matrix. The Levenberg-Marquardt algorithm can be applied to solve this non-linear optimization problem and produce the following iteration [36]:

$$\beta_k = \beta_{k-1} + \mathrm{H}^{-1}\chi. \tag{20}$$

Where β_k and β_{k-1} are the motion parameter vectors at iteration k and k-1, and the coefficients of matrix **H** and vector χ are given by:

$$\begin{cases} \mathrm{H} = \frac{1}{2}\mathrm{A} = [h_{sl}] = [\sum_i \frac{\partial e_i}{\partial \beta_s}\frac{\partial e_i}{\partial \beta_l}] \\ \chi = -\frac{1}{2}\zeta = [\chi_s] = [-\sum_i e_i \frac{\partial e_i}{\partial \beta_s}]. \end{cases} \quad 1 \le s, l \le 4 \tag{21}$$

Using the output of RLS as the initial value β_0, the motion parameter vector β is updated iteratively until a convergence is reached. Because the initial value is very close to the global minimum, the accurate convergence can be reached with no more than five iterations generally.

IV. Experimental Results

A. Block size setting

The setting of circular block size can affect the performance of the DIS system. Bigger block introduces higher complexity of block motion search. Smaller block can reduce the complexity of block motion search, but contains less information for block matching which may degrade the matching precision. On the other hand, the block size can also affect the upper and lower bounds of the scaling motion that can be managed by the proposed DIS system, because when scaling motion exists, bigger block size introduces more differences between the content of a block in current frame and that of the corresponding block in reference frame and thus degrades the matching precision.

Fig. 7 shows the upper and lower bounds of scaling motion at different circular block radiuses. When the block radius is set to 8 pixels, the system achieves the best adaptation to scaling motion. Therefore, the circular block radius is set to 8 pixels in the DIS system and at this time, the range of the scaling factor that can be managed by the proposed DIS system is 0.84 to 1.14.

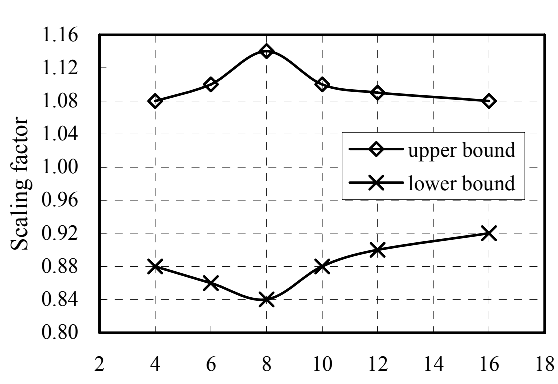


Fig. 7. Relation between block radius and scaling factor limits

B. Performance of the proposed ME technique

The performance of the proposed DIS system is evaluated with two schemes. In the first scheme, we captured a stable video sequence by mounting the camera on a steady tripod and then added some jitter motions to the video sequence with ground truth GMP to generate a fake jitter video. The performance of the DIS technique is evaluated by comparing the estimated GMP with the ground truth GMP. In the second scheme, we captured a jitter video by holding the camera by hand. Then the difference image between each stabilized frame and its previous stabilized frame is produced to demonstrate the stabilization results.

To show the estimation precisions of different kind of motions more intuitionisticly, we rewrite the similarity motion model in (13) as:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \gamma \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}. \tag{22}$$

Where γ is scaling factor, θ is rotation angle, and c and d are horizontal and vertical translation parameters respectively.

Four cases of motion combinations are considered for evaluation: 1) translational motion only, 2) translational and rotational motions, 3) translational and scaling motions, and 4) all three kinds of motions. The evaluation results of these four cases are shown in Table I ~ IV respectively. The motion parameters of RLS outputs are also presented in the tables together with those of final outputs. From the results, we can see that 1) the RLS outputs are very close to the final outputs and 2) the final outputs are more precise than RLS output, and this demonstrates that 1) the circular block matching based GMP generation algorithm works effectively and 2) the gradient descent based iteration scheme is helpful to improve the estimation precision. The estimation errors shown in these tables are defined as the absolute differences between the ground truth and the final estimated parameters. The results show that the estimation errors for scaling, rotational and translational motions are less than 0.005, 0.08 degree and 0.05 pixels respectively, i.e., our proposed DIS system can stabilize images effectively in all four cases.

TABLE I
Estimation Results with Ground Truth (Translation Only)

	γ	θ (degree)	c (pixel)	d (pixel)
Ground truth	1.00	0.00	15.50	6.25
RLS output	1.0005	0.0612	15.4331	6.0323
Final output	1.0005	0.0612	15.4979	6.2416
Estimation error	0.0005	0.0612	0.0021	0.0084

TABLE II
Estimation Results with Ground Truth (Translation + Rotation)

	γ	θ (degree)	c (pixel)	d (pixel)
Ground truth	1.0000	8.50	12.00	-6.50
RLS output	1.0016	8.5792	12.0012	-6.9253
Final output	1.0016	8.5792	11.9973	-6.5148
Estimation error	0.0016	0.0792	0.0027	0.0148

TABLE III
ESTIMATION RESULTS WITH GROUND TRUTH (TRANSLATION + SCALING)

	γ	θ (degree)	c (pixel)	d (pixel)
Ground truth	1.07	0.00	-6.50	7.00
RLS output	1.0728	0.0243	-6.6028	6.7214
Final output	1.0728	0.0243	-6.5207	6.9711
Estimation error	0.0028	0.0243	0.0207	0.0289

TABLE IV
ESTIMATION RESULTS WITH GROUND TRUTH (FULL MOTIONS)

	γ	θ (degree)	c (pixel)	d (pixel)
Ground truth	1.08	-8.00	5.20	4.70
RLS output	1.0821	-8.0548	5.2533	4.2339
Final output	1.0821	-8.0548	5.1610	4.6790
Estimation error	0.0021	0.0548	0.0390	0.0210

Some stabilization results for a real captured jitter video are shown in Fig. 8. The top-left and top-right images in the figure are two successive frames in the original video sequence. We stabilized the right image (current frame) to the left image (reference frame) with our proposed algorithm and the stabilization result is shown in the middle of the figure. The bottom-left image in the figure is the difference image between the two original successive frames and the bottom right image shows the difference between the stabilized frame and the reference frame. In the difference images, only the pixels with absolute difference higher than 10 are displayed [28]. The results show that the differences on the background area are greatly reduced, which verifies the efficiency of the proposed DIS technique.

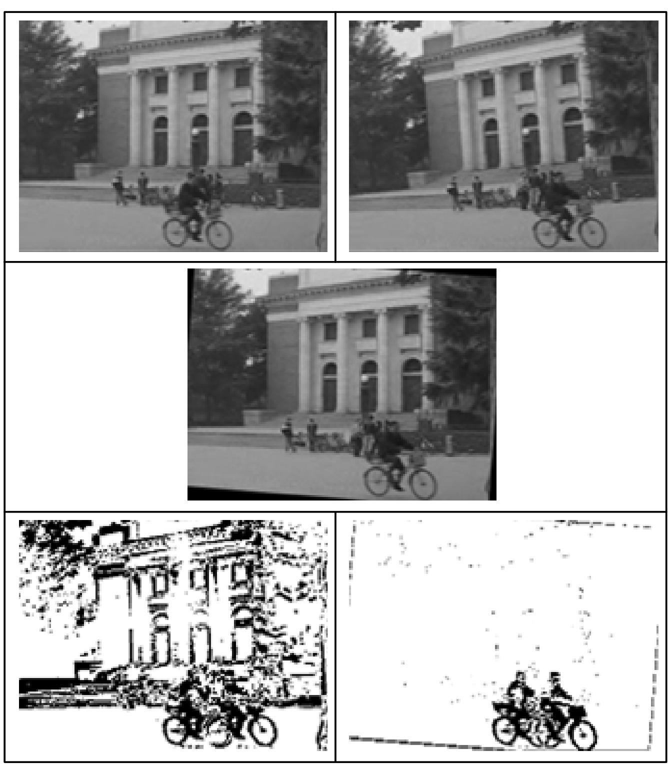


Fig. 8. Stabilization results for a real captured jitter video

C. Performance comparison

In the proposed DIS system, the block motion search based scheme guarantees that the large translational motions can be dealt with, and the rotation invariant features based circular block matching technique ensures that large rotational motions can also be handled. To further demonstrate the performance of the proposed algorithm, we compared our circular block matching (CBM) based ME algorithm with the optical flow (OF) based ME algorithm presented in [24] in two aspects: estimation precision and processing capabilities for translational, rotational and scaling motions. Fig. 9-11, which present the estimation errors of translation parameters, rotation angles and scaling factors respectively, show the performances of these two algorithms. For scale, rotation and translation parameters, the estimation precisions of CBM based algorithm are about 0.005, 0.08 degree and 0.05 pixels respectively and those of OF based algorithm are about 0.01, 0.6 degree and 0.3 pixels respectively. It can be concluded that the proposed CBM based algorithm can generate more precise GMP compared with OF based algorithm.

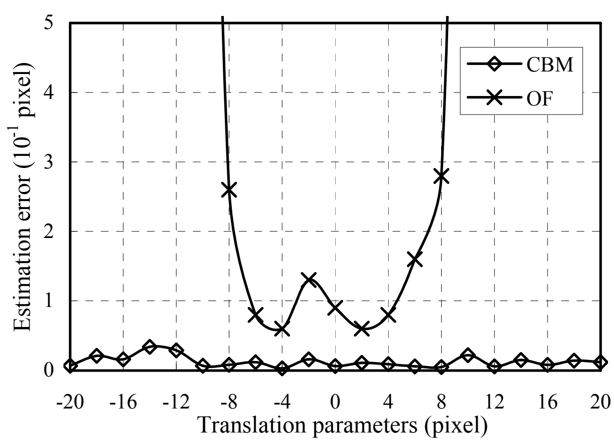


Fig. 9. Estimation errors of translation parameters

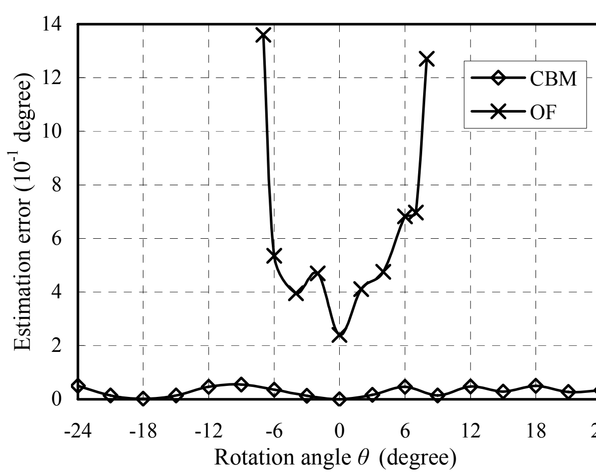


Fig. 10. Estimation errors of rotation angles

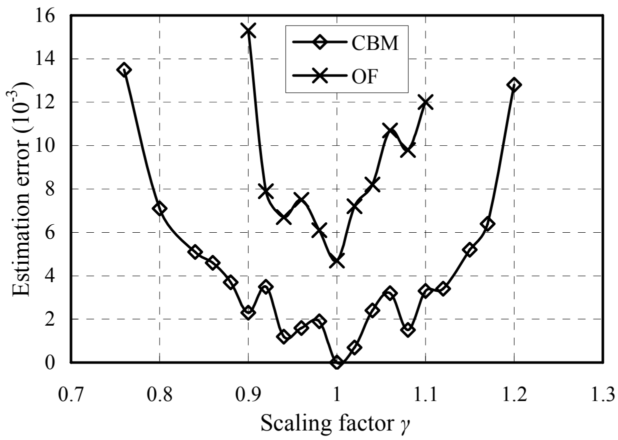


Fig. 11. Estimation errors of scaling factors

The processing capabilities of the CBM based and OF based algorithms are shown in Table V. To explain why the circular block is adopted instead of rectangular block in the proposed system, we replaced the circular block matching module in the proposed DIS system with the traditional rectangular block matching (RBM) scheme. The processing capability of this RBM based algorithm is also presented in the table. The results show that the CBM based algorithm has the highest processing capabilities among the three algorithms. Theoretically, the CBM based algorithm can deal with any translational and rotational motions. But in practice, its processing capabilities to translational and rotational motions are limited by the size of searching window for CBM.

TABLE V
PROCESSING CAPABILITIES TO DIFFERENT MOTIONS

	γ	$\lvert\theta\rvert$ (degree)	$\lvert c\rvert$ *or* $\lvert d\rvert$ (pixel)
CBM	0.84 – 1.14	> 25	> 20
OF	0.92 – 1.08	< 7	< 10
RBM	0.86 – 1.12	< 7	> 20

By utilizing the computational complexity reduction schemes proposed in section III and implementing some modules, e.g. Gaussian-Laplacian image generation, with Intel SIMD instructions, the proposed DIS system can perform real-time processing for QVGA format videos on a PC with Pentium IV 3.0GHz CPU and 1GB RAM.

V. CONCLUSIONS

DIS is a very useful technique for improving the visual quality of the video sequences captured by compact and light weight digital video cameras. In this paper, a brief review of the ever proposed 2D and MD DIS algorithms is first given. Then we proposed a novel circular block matching based DIS algorithm which can deal with scaling, rotational and translational motions. In this algorithm, circular block matching is first performed to generate local MVs of the block centers by utilizing some rotation invariant features. Then with these local MVs and their corresponding image positions, a linear system is constructed and solved by repeated least-squares to generate the GMP which is then refined by a gradient descent based iteration scheme. The experimental results show that the proposed CBM based algorithm can generate more precise motion parameters and has higher processing capabilities for different motions compared with OF based algorithm.

ACKNOWLEDGMENT

We thank Prof. Shang-Hong Lai for verifying the experimental results related to the optical flow based DIS algorithm presented in [24].

REFERENCES

[1] Y.C. Peng, H.A. Chang, H. Chen, and C.J. Kao, "Digital image stabilization and its integration with video encoder," *IEEE Consumer Communications and Networking Conference, 2005,* Jan. 3-6, 2005, pp. 544-549.

[2] Y.T. Tse and R. L. Baker, "Global zoom/pan estimation and compensation for video compression," *IEEE Proc. of International Conference on Acoustics, Speech, and Signal Processing, 1991*, April 14-17, 1991, vol. 4, pp.2725-2728.

[3] A. Engelsberg and G. Schmidt, "A comparative review of digital image stabilising algorithms for mobile video communications," *IEEE Trans. Consumer Electron.,* vol. 45, no. 3, pp. 592-597, Aug., 1999.

[4] C.K. Liang, Y.C. Peng, H.A. Chang, C.C. Su, and H. Chen, "The effect of digital image stabilization on coding performance," *IEEE Proc. of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing,* Oct. 20-22, 2004, pp. 402-405.

[5] L. Marcenaro, G. Vernazza, and C. S. Regazzoni, "Image stabilization algorithms for video-surveillance applications," *IEEE Proc. of International Conference on Image Processing, 2001*, Oct. 7-10, 1991, vol. 1, pp.349-352.

[6] S.C. Hse, S.F. Liang, and C.T. Lin, "A robust digital image stabilization technique based on inverse triangle method and background detection," *IEEE Trans. Consumer Electron.,* vol. 51, no. 2, pp.335-345, May, 1999.

[7] M. Hansen, P. Anandan, K. Dana, G. Wan der Wal, and P. Burt, "Real-time scene stabilization and mosaic construction," *IEEE Workshop on Applications of Computer Vision, 1994,* Dec. 5-7, 1994, pp. 54-62.

[8] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation," *IEEE Trans. Consumer Electron.,* vol. 49, no. 4, pp.1320-1325, Nov., 2003.

[9] F. Vella, A. Castorina, M. Mancuso, and G. Messina, "Digital image stabilization by adaptive block motion vectors filtering," *IEEE Trans. Consumer Electron.,* vol. 48, no. 3, pp.796-800, Aug., 2002.

[10] S. Piva, M. Zara, G. Gera, and C. S. Regazzoni, "Color-based video stabilization for real-time on-board object detection on high-speed trains," *IEEE Proc. of International Conference on Advanced Video and Signal Based Surveillance, 2003*, July 21-22, 2003, pp.299-304.

[11] K.S. Choi, J.S. Lee, J.W. Kim, S.H. Lee, and S.J. Ko, "An efficient digital image stabilizing technique for mobile video communications," *IEEE Proc. of International Conference on Consumer Electronics, 2000*, June 13-15, 2000, pp.299-304.

[12] S.J. Ko, S.H. Lee, and K.H. Lee, "Digital image stabilizing algorithms based on bit-plane matching," *IEEE Trans. Consumer Electron.,* vol. 44, no. 3, pp.796-800, Aug., 1998.

[13] Y. Egusa, H. Akahori, A. Morimura, and N. Wakami, "An electronic video camera image stabilizer operated on fuzzy theory," *IEEE Proc. of International Conference on Fuzzy Systems, 1992*, Mar. 8-12, 1992, pp.851-858.

[14] J.C. Tucher and A. de Sam Lazaro, "Image stabilization for a camera on a moving platform," *IEEE Proc. of Pacific Rim Conference on Communications, Computers and Signal Processing, 1993*, May 19-21, 1993, vol. 2, pp.734-737.

[15] Y. Egusa, H. Akahori, A. Morimura, and N. Wakami, "An application of fuzzy set theory for an electronic video camera image stabilizer," *IEEE Trans. Fuzzy Systems, 1995*, vol. 3, no. 3, pp.351-356, Aug., 1995.

[16] S.J. Ko, S.H. Lee, S.W. Jeon, and E.S. Kang, "Fast digital image stabilizer based on gray-coded bit-plane matching," *IEEE Trans. Consumer Electron.,* vol. 45, no. 3, pp.598-603, Aug., 1999.

[17] S. Erturk, "Image sequence stabilisation: motion vector integration (MVI) versus frame position smoothing (FPS)," *IEEE Proc. of International Conference on Image and Signal Processing and Analysis, 2001,* June 19-21, 2001, pp. 266-271.

[18] K. Ratakonda, "Real-time digital video stabilization for multi-media applications," *IEEE Proc. of International Symposium on Circuits and Systems, 1998,* May 31- June 3, 1998, vol. 4, pp.69-72.

[19] M.K. Gullu and S. Erturk, "Fuzzy image sequence stabilisation," *IEE Electronics Letters,* vol. 39, no. 16, pp.1170-1172, Aug., 2003.

[20] K. Uomori, et al., "Automatic image stabilizing system by full-digital signal processing," *IEEE Trans. Consumer Electron.,* vol. 36, no. 3, pp.510-519, Aug., 1990.

[21] J.K. Paik, Y.C. Park, and D.W. Kim, "An adaptive motion decision system for digital image stabilizer based on edge pattern matching," *IEEE Trans. Consumer Electron.,* vol. 38, no. 3, pp.607-616, Aug., 1992.

[22] J.K. Paik, Y.C. Park, and S.W. Park, "An edge detection approach to digital image stabilization based on tri-state adaptive linear neurons," *IEEE Trans. Consumer Electron.,* vol. 37, no. 3, pp.521-530, Aug., 1991.

[23] H.R. Pourreza, M. Rahmati, and F. Behazin, "An electronic digital image stabilizer based on stationary wavelet transform (SWT)," *IEEE Proc. of International Conference on Image Processing, 2003*, Sept. 14-17, 2003, vol. 2, pp.383-386.

[24] H.C. Chang, S.H. Lai, and K.R. Lu, "A robust and efficient video stabilization algorithm," *IEEE Proc. of International Conference on Multimedia and Expo, 2004*, June 27-30, 2004, vol. 1, pp.29-32.

[25] A. Censi, A. Fusiello, and V. Roberto, "Image stabilization by features tracking," *IEEE Proc. of International Conference on Image Analysis and Processing, 1999*, Sept. 27-29, 1999, pp.665-667.

[26] J.Y. Chang, W.F. Hu, M.H. Cheng, and B.S. Shang, "Digital image translation and rotation motion stabilization using optical flow technique," *IEEE Trans. Consumer Electron.,* vol. 48, no. 1, pp.108-115, Feb., 2002.

[27] Y.S. Yao, P. Burlina, R. Chellappa, and T.H. Wu, "Electronic image stabilization using multiple visual cues," *IEEE Proc. of International Conference on Image Processing, 1995*, Oct. 23-26, 1995, vol. 1, pp.191-194.

[28] C. Morimoto and R. Chellappa, "Fast electronic digital image stabilization," *IEEE Proc. of International Conference on Pattern Recognition, 1996*, Aug. 25-29, 1996, vol. 3, pp.284-288.

[29] S. Erturk, "Translation, rotation and scale stabilization of image sequences," *IEE Electronics Letters,* vol. 39, no. 17, pp.1245-12462, Aug., 2003.

[30] S. Srinivasan and R. Chellappa, "Image stabilization and mosaicking using the overlapped basis optical flow field," *IEEE Proc. of International Conference on Image Processing, 1997,* Oct. 26-29, 1997, vol. 3, pp. 356-359.

[31] Y.M. Liang, H.R. Tyan, H.Y. Mark Liao, and S.W. Chen, "Stabilizing image sequences taken by the camcorder mounted on a moving vehicle," *IEEE Proc. of Intelligent Transportation Systems, 2003,* vol. 1, pp.90-95.

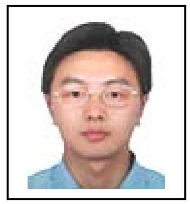
[32] Milan Sonka, Vaclav Hlavac, and Roger Boyle, *Image Processing, Analysis, and Machine Vision,* 2nd ed., Plenum Press: Brooks/Cole, 1993, pp. 685-689.

[33] L. Meng, X.G. Lin, L.D. Xu, and F.W. Fu, "Video stabilization for digital camera," *Proc. of 7th International Conference on Signal Processing*, *2004*, Aug. 31-Sep. 4, 2004, pp. 1119- 1122.

[34] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision,* 2nd ed., Plenum Press: Brooks/Cole, 1993, pp. 81-88.

[35] E.M.L. Beale, *Introduction to optimization*, Plenum Press: John Wiley & Sons, 1988, pp.3-7.

[36] H.K. Cheung and W.C. Siu, "Fast global motion estimation for sprite generation," *IEEE Proc. of International Symposium on Circuits and Systems, 2002,* May 26-29, 2002, vol. 3, pp. III-1 - III-4.



Lidong Xu received the B.S. degree in Electronic Engineering from Tsinghua University in China in 2001. He is currently a PH.D. candidate in Electronic Engineering from Tsinghua University in China. In 2004, he worked as an intern in Intel China Research Center for more than 3 months. He was a session chair of the fifth IEEE International Symposium on Communications and Information Technologies (2005). His research interests are in the areas of digital video processing and video compression.

Xinggang Lin (SM'94) received the Ph.D. degree and M.S. degree both in information science from Kyoto University in Japan in 1986 and 1982, respectively, and the B.S. degree in Electronic Engineering from Tsinghua University in China in 1970. In 1970 to 1978, he was an engineer at Hunjiang (Baishan) Electronic Instrument Corporation in Northeast China in charge of R&D of test apparatus of semiconductor devices. He joined the Department of Electronic Engineering at Tsinghua University in Beijing, China in 1986 where he has been a full professor since 1990. In 1997 he visited the City University of Hong Kong as a senior research fellow for half a year. For his achievements in National High-tech R&D Program (863 Program), he received "Great Contribution Award" from Ministry of Science and Technology of China in 1992, and "Promotion Awards of Science and Technology" from Beijing Municipality for his excellent researches in image processing and analysis. He was a General Co-chair of the second IEEE Pacific-Rim Conference on Multimedia (2001), and a technical/organizing committee member of many international conferences such as VCIP, PCS, etc. He is now the associate editor of IEEE Transactions on Circuits and Systems for Video Technology, a senior member of IEEE and the Chinese Institute of Electronics, a fellow of China Institute of Communications, and has published over 140 referred conference and journal papers in diversified research fields including digital image/video processing and analysis, pattern recognition, video coding and communications, multimedia man-machine interface, hardware implementation of signal processing systems, etc.