


# Improving security surveillance by hidden cameras

Hadia Tazeem<sup>1</sup> · Muhammad Shahid Farid<sup>1,2</sup>  ·  
Arif Mahmood<sup>3</sup>

Received: 23 March 2015 / Revised: 23 November 2015 / Accepted: 8 January 2016 /  
Published online: 20 January 2016  
© Springer Science+Business Media New York 2016

**Abstract** Increasing solicitudes about security demand better, robust and effective solutions. Security cameras are playing a vital role in this regard and the surveillance technology is improving rapidly. However, these cameras are usually installed at obvious and visible locations which are often exploitable by the criminals either by hiding themselves from the camera, choosing an alternative path or deceiving the camera. This situation can be overcome to a large extent if the cameras are installed at hidden places looking through narrow regions, e.g. camera fixed inside the building and looking through the window curtain slits. However, this solution poses new challenges in terms of capturing the video through slits and accumulating the information to a meaningful view. In this paper we propose an effective and robust solution to this problem that automatically extracts the slit regions and merges them over a large number of frames to construct a panoramic view. Moreover, such a security surveillance system will be subjected to the sudden illumination variations. We effectively handle such variations by incorporating robustness in the proposed framework. A large number of experiments are performed on various indoor and outdoor real video sequences. The results demonstrate the effectiveness of the proposed framework.

---

✉ Muhammad Shahid Farid  
farid@di.unito.it  
Hadia Tazeem  
hadia.tazeem@pucit.edu.pk  
Arif Mahmood  
arif.mahmood@uwa.edu.au

<sup>1</sup> Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan

<sup>2</sup> Dipartimento di Informatica, Università degli Studi di Torino, Corso Svizzera 185,  
10149 Torino, Italy

<sup>3</sup> School of Computer Science and Software Engineering, The University of Western Australia,  
35 Stirling Highway, Crawley, Australia

Experiments are also performed to objectively assess the perceptual quality of the resulting panoramic images. Our results are even better than the existing commercial software.

**Keywords** Security surveillance · Mosaicing · Scene stitching · Global flow · Illumination compensation

## 1 Introduction

Demand of better security and surveillance systems has undergone a fast growth in the recent years. It is due to deteriorating political and economic stability in many parts of the world. The number of cameras used for surveillance has been increasing with time. Surveillance cameras were initially used to keep a check on robbers and shoplifters in large stores and trade centers where it was not possible to keep a vigil on every one by a single person. Surveillance by camera has gained gigantic popularity and now it has been used almost everywhere from schools, colleges, hospitals, shopping malls to public places. Recording in these cameras is also used to identify the thieves and robbers in homes, banks or at other victim places. Placement of security camera is a crucial step in implementation of video surveillance technology. Installing a camera on an inappropriate location may create a flaw in the security system and a gap in the field of view and thus it may expose a facility to outside threats and decreases the effectiveness of the security system.

Usually, the security cameras fail to capture the identity of an offender because the presence of cameras is often at obvious location and a suspect may easily take counter measures. Visibility and detection of the security cameras may be exploited by the suspects to avoid being captured by the camera by moving through uncovered corridors or alternatively covering them with big hats, black glasses, facial masks or sheets to avoid their identity being captured. The best place for security cameras is the one that is not visible or accessible. If cameras are fixed at hidden or secret locations, full field of view (FOV) may not be available and in extreme cases, only a part of the FOV will be captured. The small visible portion of the scene in a frame may not yield sufficient useful information about the scene and the objects. However, if the visible scene is mosaiced across a number of frames, sufficient scene information may be obtained which can be used for better decision making. This paper proposes a solution to the problem where a camera is capturing the scene through a narrow region (referred to as *slit* in the rest of the text) with limited field of view (FOV) which significantly limits the available information. An object passing through the FOV may not be completely visible in a single frame. However, by combining information from a group of successive frames one may obtain the complete view of the object. Stitching two or more partially overlapped images into a single big picture is called *image mosaicing* in which the correspondence between the images is estimated and used to merge them into a single image.

Image mosaicing is indeed a complex task that involves not only an automatic correspondence estimation of the input images but a seamless integration is also required. Seamless integration permits the edge consistency across the mosaic that is the edges of the objects in the mosaic must be connected for plausible results. Mosaicing from narrow slits is a challenging task in many respects compared to simple image mosaicing as described below:

- *Limited Field of View (FOV)*: In the proposed scenario, a camera is installed at a hidden place and a narrow region is available of the target view. In this case, a large part of the

captured images is redundant and useless. The slit region must be extracted from the images and integrated so that the complete view of an object can be obtained.

- *Illumination Variation*: Since we deal with the videos, the scene illumination may vary over time resulting in images with different illumination. Integrating such images would result in low quality, imperceptible picture. A viable solution would deem to filter the illumination variation in the video before merge. Illumination filtration will also increase the image registration performance by enabling robust feature matching.
- *Image Registration*: Due to limited FOV image registration and alignment becomes a difficult task. Most of the current mosaicing algorithms require significant overlapped region for correct image registration and alignment. Therefore these algorithms cannot be directly applied for the case of limited FOV. We propose an image integration method to handle limited FOV for the purpose of image mosaicing.

In proposed framework first the slit region is extracted from each frame in the input video sequence to remove the useless redundant frame regions. In the second step, illumination filtration is performed on the slit to filter out any sudden variations in the illumination. In the third step, the geometrical relationship between the slits is estimated and iteratively refined. In the fourth step, the slits are stitched together by using the geometric transformation recovered in the third step. The main contributions of this paper are as follows:

- Proposal of a novel security surveillance system where camera is placed at a non-obvious hidden location with a limited field of view;
- A framework is proposed to extract important content in the form of a pano-ramic view from this limited context video;
- It is noted that due to limited FOV the impact of illumination variation is noticeable in the mosaiced image. To this end, a frequency based illumination filtering is performed before image alignment;
- Global affine warp model is used for image alignment. The model parameters are estimated through an iterative multi-scale algorithm;
- The proposed framework is tested on real sequences captured with limited FOV as well as on synthetic data to analyze the performance both visually and numerically. The results are also compared with the well-known commercial mosaicing tools.

The rest of the paper is organized as follows. In Section 2 a review of the related literature is presented. Section 3 gives an overview of the proposed security system. The slit extraction and illumination filtration is described in Section 4, Slit registration and integration are described in Sections 5 and 6 respectively. The experimental evaluation is presented in Section 7 and the conclusions are drawn in Section 8.

## 2 Related work

In this section we briefly review the state-of-the-art literature related to security surveillance and image mosaicing. Video camera based security surveillance systems are being used everywhere: in law enforcement agencies, at homes, shopping malls, markets, airports, etc. Last decade has witnessed a tremendous growth in security surveillance systems [29]; high quality surveillance devices have been introduced and robust computer vision algorithms realized them into efficient security systems.

The existing security surveillance systems use a single or multiple video cameras that exploit the full field of view to monitor the target area. The videos displayed on monitors

are being observed by a human to detect any unusual, suspicious event. However, automatic surveillance is paramount and has attracted a significant research in the recent years. The surveillance system proposed in [38] extracts motion trajectories from video and analyze them to detect the events. It exploits support vector machines (SVM) and propose kernel boundary alignment (KBA) algorithm for effective leaning and recognition. Video based surveillance system in [22] is proposed for monitoring the movements of aircraft, ships and other commercial vehicles on the ground or in port. In addition to video, the audio is also used to provide a robust surveillance. A wireless based surveillance system for homes was proposed in [13]. It detects the suspicious event and informs the monitors through text and multimedia messages. An automated video surveillance system *Knight* is proposed in [30]. It detects and classifies the objects moving in the field of view without human intervention and reports the important activities in the form of key frames.

Image mosaicing is a challenging problem in respect of seamless integration of the images. A minor unavoidable gray level difference in frames boundaries gains considerable visibility in panoramic mosaic. Furthermore, when objects are moving the mosaicing results in a blurry or ghosted panoramic picture. During the stitching process a composition of the overlapped regions is computed in order to create smooth transition between spatially consecutive regions. Ghosting and exposure are two common artifacts in dynamic mosaics. Uyttendaele et al. [35] proposed a weighted vertex cover algorithm to remove the blur effect visible due to moving objects. The block based adjustment algorithm adjusts the scenes where a single change in exposure would result in an under or overexposed image. To avoid ghosting in the integrated view, the pixel values should be selected from only one of the contributing images. The exposure difference occurs due to change in exposure setting of automatic camera while capturing the sequence of images for mosaic construction. To compensate the exposure effect, a full radiometric camera calibration is performed. Suen et al. [33] proposed an optimization method based on image derivatives to limit the ghosting effect by achieving better image alignment. To minimize the ghosting effect Combined SIFT [19] and Dynamic Programming (CSDP) technique is proposed in [42]. It uses SIFT based matching and dynamic programming with edge-enhanced weighting intensity difference operator to limit the ghosting effect the in the mosaic image. The recent ghost detection and removal methods are reviewed in [32].

Global warping methods e.g. similarity, affine and projective are used to align the images [6, 11, 15, 44] to create a panorama picture. However, these warpings may not produce accurate alignment resulting in shape and area distortions in the stitched image [17]. To address this problem a number of local warping methods have been proposed. A smoothly varying affine (SVA) model [17] computes local affine warps for overlapping regions to improve the alignment accuracy. Moving Direct Linear Transformation (Moving DLT) under projective warp is proposed in [41] to minimize the local inconsistencies in the overlapped regions under the global transform. Projective and similarity warps are combined [4] that uses global projective warping for improved warping accuracy and exploits similarity warp to preserve the local objects' shapes. A robotic setpoint control technique for limited FOV of microscope is proposed in [16] for optical manipulation with unknown trapping stiffness. The research in [14] applies video surveillance for user behavior recognition in workflows. Cristofaro et al. [5] presented a study of minimum time trajectories for a differential drive robot with a fixed and limited field-of-view. Risk assessment is an important activity in video surveillance systems which helps to detect the threats and vulnerabilities in the system and to devise the mitigation strategies. A risk assessment model based on fuzzy cognitive maps is proposed in [34]. Affine warp parameters are recursively computed and

augmented with Kalman filter in [39] to improve the alignment accuracy for video mosaicing. A computationally efficient graph based image mosaicing approach is proposed in [7]. Some recent research in video surveillance with field-of-view constraints can be found in [34].

A variational method for image blending is proposed in [36]. It improves the quality of the blended image by computing a weighting mask for each of the multiple input images. A similar technique using image gradients is proposed in [24]. A nonlinear weighting fusion of input images is also proposed in [43] for seamless integration. Nonuniform exposures may result in visible seams in the mosaic. Litvinov et al. [18] exploited the camera non-uniformity and radiometric response to overcome this problem. A Fourier transform based image blending is proposed in [9] where different level frequencies of the input images are mixed together to obtain seamless integration. Seamless stitching in aerial images [20] is obtained by decomposing the partially overlapped images into high and low frequency images and different mosaicing schemes are used for each frequency component. For low frequencies simple weighted blending is used; for high frequencies an improved seam searching strategy is devised for seamless blending. A good review of state of the art image blending techniques can be found in [27].

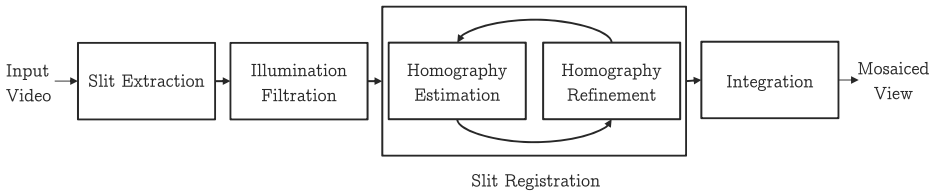
For better stitching Fu and Wang [10] proposed to use Harris operator [12] to find the corner features in the input images which are used for matching with RANSAC method to estimate the transformation matrix for registration to compute the panorama image. SIFT and RANSAC algorithms are also utilized in [31] for seamless stitching to create a panoramic image. The Harris operator with region based matching is proposed in [40] for efficient mosaicing. In particular, they showed that regions matching results in better estimates of rotation, scaling and translation between the two successive images.

### 3 The proposed security surveillance system

We propose a novel security surveillance system that allows the installation of video cameras at non-obvious, hidden locations so that they cannot be tricked by the suspects. Moreover, the camera is capturing the target with a limited field of view. Due to limited FOV the target may not be completely visible in a single video frame; the proposed system exploits the video stream in temporal direction and leverages on image mosaicing to collect the information in a single picture.

The proposed system works in four steps starting from the slit region extraction from the input video. In the second step, the slits are filtered to remove illumination variations in the video which is imperative for good quality of integrated image and also improves the registration step. In the third step, the global geometrical relationship between the input slits is estimated. The estimated relationship is refined through a multi-resolution iterative procedure. The final step is to warp the input slits to a selected reference view and merge them into a single larger image. These steps are described in detail in the following sections. Figure 1 shows the block diagram of the proposed algorithm.

**Camera setup** We assume a single camera setup - the camera is fixed and capturing the scene through a narrow view. The orientation of the view (visible region) can be configured in multiple ways e.g. horizontal or vertical rectangular slit, circular or square hole, etc. This orientation depends on the position of the camera with respect to the target region. Here we assume that camera is installed at a hidden location parallel to the target with a vertical



**Fig. 1** Block diagram of proposed framework

rectangular exposure. Figure 2 shows two sample frames from test video sequences S1 and S2 (see Table 1 for details) captured in this setup through a narrow region exposed between the barriers.

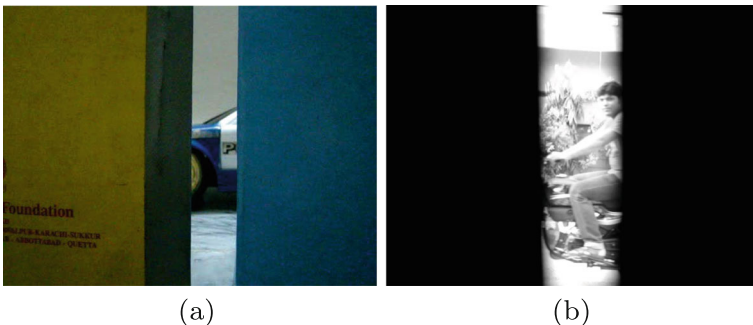
### 4 Slit extraction and illumination filtration

Slit extraction refers to the process of segmenting the region of interest (ROI) from the video frames. As described previously, slit is a narrow target region in the image between the obstacles or barriers. This can be achieved by marking the four corner coordinates of the ROI manually at the system setup time and then use them for slit extraction in the video stream. The vertical lines between the left-top, left-bottom and right-top, right-bottom points are plotted by joining the four input points and used to define the ROI in the image. The method is simple and require minimum user interaction, just once at the system setup time.

The next step after slit extraction is to eliminate the illumination variations in the slits of the video. Illumination filtering is performed through homomorphic filtering in frequency domain on logarithm transformation of the slit. We assume Lambertian object surfaces where pixel intensity in an image is the product of source illumination and surface reflectance [25]. Let  $I^t$  be image at time  $t$  of size  $m \times n$ , the intensity of pixel  $(x, y)$  is:

$$I^t(x, y) = I_i^t(x, y) \cdot I_r^t(x, y) \tag{1}$$

where  $I_i^t(x, y)$  is the illumination component and  $I_r^t(x, y)$  is the reflectance component of the surface. To eliminate the illumination variation the illumination component  $I_i^t(x, y)$  must be filtered out. The low frequencies of the image mainly contribute to the illumination whereas the mid to high frequencies form the reflectance component [1]. Logarithm



**Fig. 2** Sample frames from test video sequence S1 (a), and S2 (b)

**Table 1** Dataset details

Seq. No.	No. Frames	Slit Size	Video Type
S1	140	30 × 480	Outdoor
S2	238	63 × 480	Indoor
S3	98	91 × 480	Outdoor
S4	48	31 × 480	Outdoor
S5	52	74 × 480	Outdoor
S6	35	31 × 480	Outdoor
S7	64	66 × 480	Outdoor
S8	99	74 × 480	Outdoor

transform has been used in contrast enhancement [2]. To separate the reflectance and illumination the image is transformed in logarithm domain and is thresholded with a predefined value. From (1):

$$\log(I^t(x, y)) = \log(I_i^t(x, y)) + \log(I_r^t(x, y)) \tag{2}$$

To separate the two components we transform the resultant image  $\log(I^t(x, y))$  in frequency domain and apply the appropriate filters (low-pass and high-pass) to get the desired frequencies. We use Fourier transform to represent the image in frequency domain:

$$\tilde{I}^t = \mathcal{F}[\log(I^t(x, y))] \tag{3}$$

where  $\mathcal{F}(\cdot)$  is Fourier operator.

To extract the high ( or low) frequencies from the transformed image various filters e.g. Ideal, Butterworth, Gaussian are available. We choose Gaussian filter as it does not exhibit the ringing effect [8]. To remove the low frequencies we apply Gaussian high-pass filter  $H_{hp}$ :

$$\begin{aligned} H_{hp}(u, v) &= 1 - H_{lp}(u, v) \\ &= 1 - \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \end{aligned} \tag{4}$$

where  $1 \leq u \leq m, 1 \leq v \leq n$  and  $\sigma$  is the standard deviation of the Gaussian distribution. The illumination and reflectance components are computed as:

$$I_i^t = \exp(\mathcal{F}^{-1}(H_{lp} * \tilde{I}^t)) \tag{5}$$

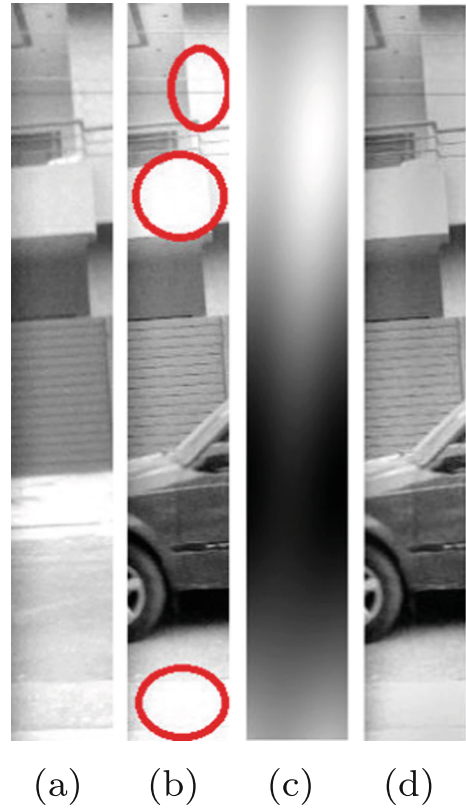
$$I_r^t = \exp(\mathcal{F}^{-1}(H_{hp} * \tilde{I}^t)) \tag{6}$$

where  $\mathcal{F}^{-1}(\cdot)$  represents the inverse Fourier. Finally, the obtained reflectance component  $I_r^t$  represents the illumination compensated image. Figure 3 shows two slits from test sequence S1 with significant illumination difference. Figure 3c shows the illumination component obtained by applying the low-pass filter  $H_{lp}$  with  $\sigma = 0.1$  to transformed image. Figure 3d shows the reflectance component of the image obtained using the corresponding high-pass filter  $H_{hp}$  (4).

### 5 Slit registration

After slit extraction and normalization the next step is to estimate the geometrical relationship between the slits with respect to a reference slit. With our assumption that camera is

**Fig. 3** **a** 1st slit of sequence S1, **b** 48th slit of sequence S1. The *red circles* show the regions with significant illumination variation. **c** Illumination component obtained by after applying proposed filtration, **d** The illumination filtered image (reflectance component)



fixed the relationship between the slits must be a homography (affinity transformation). The homography between two slits is computed using a multi-resolution iterative method. To compute the affine warp between two adjacent slits we use Bergen et.al. [3] hierarchical model. Let  $I(x, y, t)$  be the pixel intensity of the image (in this section, terms image and slit will be used interchangeably) at time  $t$  which has moved by displacement  $[u \ v]^T = [\Delta x \ \Delta y]^T$  at time  $t + \Delta t$ . The relationship can be written as:

$$\begin{aligned} u &= a_1x + a_2y + b_1 \\ v &= a_3x + a_4y + b_2 \end{aligned} \tag{7}$$

or

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{8}$$

⇒

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix} \begin{bmatrix} a'_1 \\ a_2 \\ b_1 \\ a_3 \\ a_4 \\ b_2 \end{bmatrix}$$

⇒

$$\mathbf{u} = \mathbf{Xa} \tag{9}$$



After illumination filtration (Section 4), we assume that the brightness constancy holds, that is:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{10}$$

Assuming that the pixel intensity remains constant and motion is very small, the first order Taylor expansion of (10) gives:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\delta I}{\delta x} \Delta x + \frac{\delta I}{\delta y} \Delta y + \frac{\delta I}{\delta t} \Delta t + H.O.T \tag{11}$$

Ignoring the higher-order terms (H.O.T), from (11) we obtain:

$$\frac{\delta I}{\delta x} \Delta x + \frac{\delta I}{\delta y} \Delta y + \frac{\delta I}{\delta t} \Delta t = 0 \tag{12}$$

or

$$\frac{\delta I}{\delta x} \frac{\Delta x}{\Delta t} + \frac{\delta I}{\delta y} \frac{\Delta y}{\Delta t} + \frac{\delta I}{\delta t} \frac{\Delta t}{\Delta t} = 0 \tag{13}$$

where  $\frac{\Delta x}{\Delta t} = u$  and  $\frac{\Delta y}{\Delta t} = v$  are the  $x$  and  $y$  velocity components.  $\frac{\delta I}{\delta x}$ ,  $\frac{\delta I}{\delta y}$  and  $\frac{\delta I}{\delta t}$  are the  $x$ ,  $y$  and  $t$  derivatives of image intensity  $I(x, y, t)$  respectively. Equation (13) can be written as:

$$f_x u + f_y v + f_t = 0 \tag{14}$$

which can be represented as:

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} [u \ v] = -f_t \tag{15}$$

Let,

$$f_X = \begin{bmatrix} f_x \\ f_y \end{bmatrix} \text{ and } \mathbf{u} = [u \ v]$$

The error in estimation of  $\mathbf{u}$  is calculated as:

$$E(u) = \sum_{\forall \mathbf{x} \in f(x,y)} \left( f_t + f_X^\top \mathbf{u} \right)^2 \tag{16}$$

where  $f(x, y)$  is a small patch around each pixel in which  $\mathbf{u} = [u \ v]$  remains constant. By replacing  $\mathbf{u}$  with  $\mathbf{Xa}$  in (16) from (9), the error term is:

$$E(\mathbf{a}) = \sum_{\forall \mathbf{x} \in f(x,y)} \left( f_t + f_X^\top \mathbf{Xa} \right)^2$$

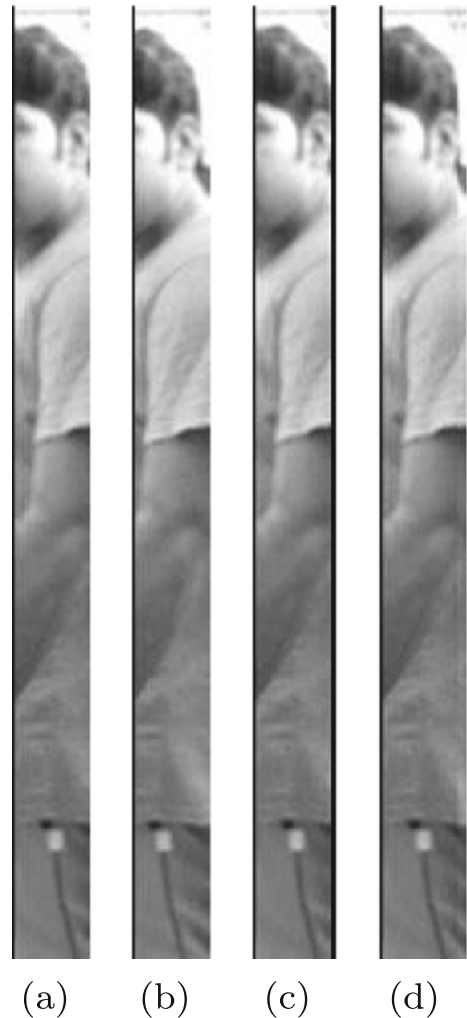
The estimation process is iteratively repeated and the incremental parameters  $\delta a$  are computed as:

$$E(\delta a) = \sum_{\forall \mathbf{x} \in f(x,y)} \left( f_t + f_X^\top \mathbf{X} \delta a \right)^2$$

Ideally, we want error to be zero, that is:

$$\begin{aligned} \sum_{\forall \mathbf{x} \in f(x,y)} \left( f_t + f_X^\top \mathbf{X} \delta a \right)^2 &= 0 \\ \implies \sum f_t + \sum f_X^\top \mathbf{X} \delta a &= 0 \\ \implies \sum f_X^\top \mathbf{X} \delta a &= - \sum f_t \end{aligned}$$

**Fig. 4** **a**  $I^{t-1}$  ( $31 \times 480$ ), **b**  $I^t$  ( $31 \times 480$ ), **c** Warped  $\tilde{I}^t$ , **d** Integrated image ( $35 \times 480$ )



For pseudo inverse, multiplying with  $X^T f_X$  on both sides:

$$[\sum X^T (f_X)(f_X)^T X] \delta a = - \sum X^T f_X f_t \quad (17)$$

which is of the form:

$$Aa = B \quad (18)$$

To obtain the best estimates the estimation process is implemented in multi-scale fashion using image pyramids. The process is iteratively repeated starting from the top level down-sampled image to refine the estimated affine parameters. Figure 4 shows an example where the homography between two slits  $I^{t-1}$  and  $I^t$  is computed and used to warp  $I^{t-1}$  to  $I^t$  and integrate the resultant images. Figure 4d shows the resultant stitched image.



**Fig. 5** Sequence: S1, From *left to right* Few intermediate mosaics, the last image shows the final mosaic

### 6 Mosaic composition

The slit integration process starts from the second slit which is expected to contain novel information than the preceding slit. Extending the same notation used in the previous section, the homography between the image  $I^{t-1}$  and  $I^t$  is computed and the image  $I^{t-1}$  is warped to  $I^t$ . The warped image  $\bar{I}^t$  may contain some holes (empty pixels) compared to  $I^t$  and the region corresponding to these holes in image  $I^t$  is the novel region appeared in the image and is denoted by  $\Omega^t$ . The mosaic  $M$  at time  $t$  is obtained by concatenating the mosaic at time  $t - 1$  with  $\Omega^t$ . This process is described by the following equations:

$$M^t = \begin{cases} I^0 & \text{if } t = 0 \\ [M^{t-1} | \Omega^t] & \text{if } t \geq 1 \end{cases}$$

where  $M^{t-1}$  and  $M^t$  are mosaics at time  $t - 1$  and  $t$  respectively. Mosaic is initialized with first image  $I^0$  at time 0 in the sequence. The black region in Fig. 4c represents the holes and the corresponding region in the image  $I^t$  is  $\Omega^t$  which is integrated with the Fig. 4a to obtain the updated mosaic  $M^t$  shown in Fig. 4d.

An important question about the mosaic composition is to decide the condition of termination of integration process. The termination criterion indeed can be based upon many attributes e.g. mosaic size, amount of significant novel information in the current image with respect to previous image. The simplest approach is to define the size of the mosaic and when the size of the integrated image reaches that threshold a new mosaic is started. An alternative strategy could use a scene change detection algorithm [23, 28] to find the amount of change in two consecutive images and in case the change is significant, more than a pre-defined value, a new mosaic may be started. This decision has large impact on the size of the mosaiced video; too frequent terminations will increase the video size.



**Fig. 6** Sequence: S2, From *left to right* Every 25th image and the last image. The last image shows the final mosaic



**Fig. 7** Sequence: S4, *Left to right* Every 5th image from the sequence and the last image resultant mosaic

## 7 Experiments and results

The proposed algorithm is tested for accuracy and scalability on 13 video sequences including eight real world videos and five synthetic sequences. The results are compared with the well known Microsoft Image Composite Editor [21]. To the best of our knowledge no video dataset matching the proposed framework requirements is publicly available; to this purpose we created a dataset comprising of 8 video sequences. The dataset contains both indoor sequences captured under constant illumination and outdoor videos shot with varying illumination. Slit size is also varied to assess the accuracy of the proposed algorithm at different resolution images. It is observed from experiments that increasing the video resolution improves the matching between the consecutive slits that results in better quality mosaic; however, it may also increase the computational time. Table 1 describes the details of the each test video sequence.

In all experiments the illumination filtering is performed through Gaussian low-pass filter with  $\sigma = 0.1$ . The number of levels  $\kappa$  and the number of iterations at each level  $\tau$



**Fig. 8** Sequence: S7, *Left to right* Every 5th image from the sequence and the last image resultant mosaic



**Fig. 9** Sequence: S8, *Left to right* Every 5th image from the sequence and the last image resultant mosaic

performed in affine warp estimation may vary depending on the image resolution and the required amount of alignment accuracy. The value of  $\tau$  varies from 1 to 5 whereas  $\kappa$  is set to 3 in all experiments. Figure 5 shows the first experiment in which the hidden camera has captured a car passing through the target field of view. Camera is looking through the narrow slit of only 30 pixels wide. The multi-scale iterative registration is performed at 3 levels of pyramid with 5 iterations at each level. The size of the resultant mosaic image is  $622 \times 480$  pixels constructed from the sequence of 140 slits. Figures 6, 7, 8 and 9 show the results of four more experiments. Each figure shows few slits from the sequence and the final mosaiced image. Figure 10 shows the final mosaics obtained in other 3 experiments. These experiments show that the visual quality of the mosaics constructed through proposed method is fairly good.

The proposed algorithm is implemented in Matlab. For each experiment the execution time is also computed which includes the alignment time, mosaic construction and file I/O time. The execution time of slit segmentation is not included as it is performed just once at camera setup time. The reported time is based on Intel(R) core-i3 CPU 2.53 GHz, on 64-bit operating system with 4GB RAM. Table 2 presents the details of the experiments.



**Fig. 10** *Left to right* Final mosaic obtained in sequence S3, S5 and S6

**Table 2** Experiment Details.  $\kappa$  is the number of levels,  $\tau$  is the number of iterations at each level. Time represents the execution time per frame

Seq#	$\kappa$	$\tau$	Mosaic size	Time(sec)
S1	3	5	622 × 480	0.30
S2	3	1	774 × 480	0.19
S3	3	3	382 × 480	0.32
S4	3	1	219 × 480	0.11
S5	3	2	269 × 480	0.26
S6	3	1	167 × 480	0.11
S7	3	5	318 × 480	0.30
S8	3	1	477 × 480	0.24

## 7.1 Objective quality assessment and comparison

To quantitatively evaluate the quality of the integrated image we performed 5 experiments on synthetic dataset. Each image is divided into a number of slits with different amount of overlapped region. Figure 11 shows such an example where an image is divided into overlapped slits, which are then stitched together to reconstruct the image. The reconstructed image is compared with the original image considered as ground truth to estimate the quality of the reconstruction. Since, all the slits belong to the same image illumination is constant across them therefore illumination filtering is not applied. Moreover, we compare the results with well known mosaicing tool ‘Microsoft Image Composite Editor’ (ICE Version 1.4.4.0) [21] to assess the accuracy and visual quality of the obtained mosaic. We also tried to use Autostitch<sup>1</sup> and Panorama Maker<sup>2</sup> tools for comparison however they failed to construct panorama from such small size slits. In fact most of the algorithms for image panorama construction require significantly larger sized images to be stitched. For very small sized slits as in our case, the performance of these algorithms deteriorate.

To objectively assess the quality of the mosaic it is compared with the original image using *Peak Signal to Noise Ratio* (PSNR) metric and well-known visual quality metric ‘*Structural SIMilarity*’ index (SSIM) [37]. PSNR estimates the statistical difference between the two images whereas SSIM is widely used for perceptual image quality assessment. SSIM is considered to produce better estimates due to its high correlation with human visual system (HVS). Therefore, we use both metrics in objective quality assessment of the proposed algorithm. The PSNR between panorama image  $I'$  and the corresponding original image  $I$  is computed as follows:

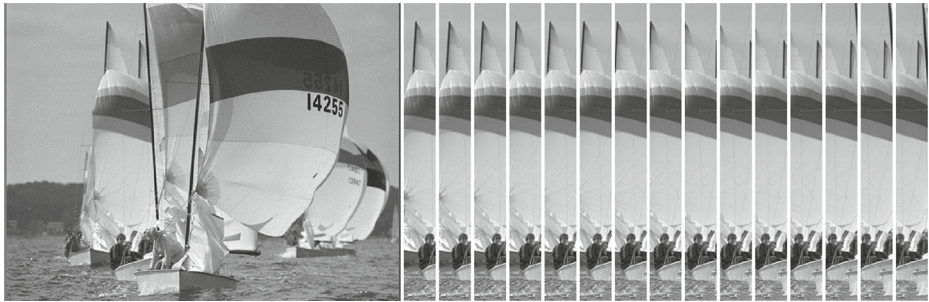
$$PSNR = 10 \log_{10} \left( \frac{255}{\sqrt{MSE}} \right) \quad (19)$$

where MSE is ‘mean square error’ computed as:

$$MSE(I', I) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [I'(i, j) - I(i, j)]^2$$

<sup>1</sup><http://www.cs.bath.ac.uk/brown/autostitch/autostitch.html>

<sup>2</sup><http://www.arcsoft.com/panorama-maker/>



**Fig. 11** A test image (*the leftmost*) with few slits obtained by divided into it into overlapped images

where  $M \times N$  is the size of the images. The SSIM index is computed as:

$$MSSIM(I', I) = \frac{1}{W} \sum_{i=1}^W SSIM(x_i, y_i) \tag{20}$$

where  $x_i$  and  $y_i$  are local windows of images  $I'$  and  $I$ .  $W$  is the number of total local windows of the image. The similarity index between the local windows pair  $x_i$  and  $y_i$  is computed as:

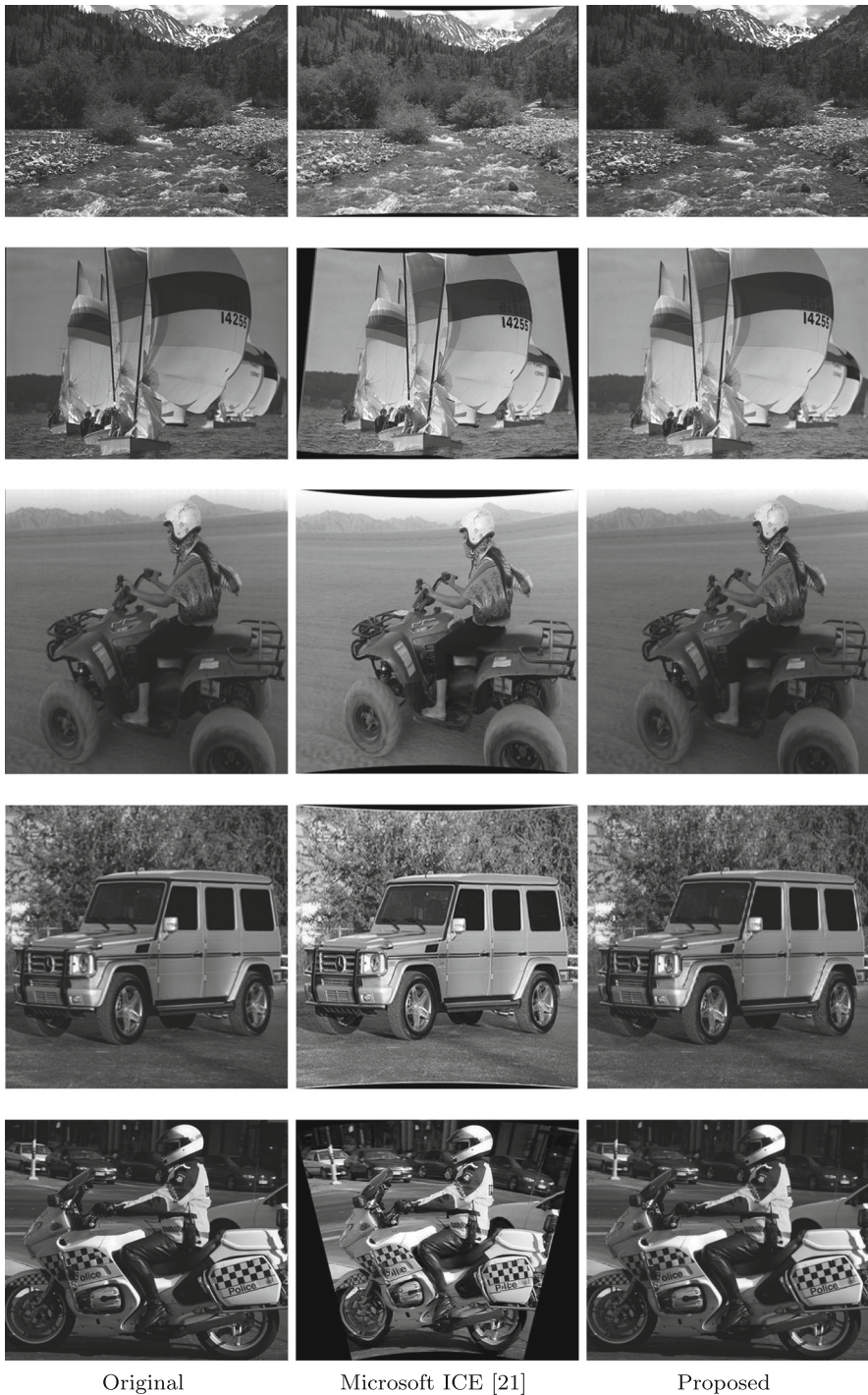
$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{21}$$

Here,  $\mu_x, \mu_y$  are the means,  $\sigma_x^2, \sigma_y^2$  are the variances of the windows  $x$  and  $y$  respectively.  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ .

In each experiment the mosaics are computed using the proposed technique and Microsoft ICE and compared with the corresponding original image to compute PSNR and SSIM scores. The obtained quality scores are listed in Table 3. In each experiment the proposed technique achieved better score in both quality metrics. In test sequence T3 and T4, the difference in PSNR is between 2 to 4 dB however, this difference is more significant in T1, T2 and T5. Similar results can be noted in SSIM values. This poor performance of ICE in these three experiments is due to the shape and area distortions. This can be visually observed in the respective mosaics shown in Fig. 12. From this we note that the mosaicing techniques that use feature based matching like SIFT and RANSAC may not produce plausible panoramas in the scenario at hand; due to very small size slits they get few matching

**Table 3** Objective Evaluation and Comparison with ICE. NS is total number of slits, S-Size is slit size, M-Size the final mosaic size. In all experiments  $\kappa = 1$  and  $\tau = 2$  is used

Seq.	NS	S-Size	M-Size	Proposed		ICE [21]	
				PSNR	SSIM	PSNR	SSIM
T1	237	40×384	512×384	21.93	0.8761	14.22	0.2828
T2	236	41×384	512×384	28.54	0.9728	12.81	0.3822
T3	143	21×450	447×450	19.71	0.5446	15.22	0.5221
T4	105	21×330	327×330	17.14	0.3712	14.65	0.3129
T5	208	19×545	640×545	37.07	0.9926	10.00	0.1183



Original

Microsoft ICE [21]

Proposed

**Fig. 12** Visual quality comparison. In each row original image (*left*), Microsoft Image Composite Editor result (*middle*) and Proposed technique result (*right*)



points which turns in poor alignment. The proposed technique that relies on flow estimation to compute the warp parameters performs better alignment resulting in reasonably good quality mosaics. Figure 12 shows the results of proposed technique and ICE on the synthetic dataset. The first two images in this dataset are taken from TID2008 database[26], while others are publicly available on internet. Both the objective and visual evaluation show the ability of the proposed framework to produce superior quality mosaics.

## 8 Conclusion

In this paper a novel security framework is proposed that leverages on mosaicing in the context of hidden security and surveillance video cameras. The proposed framework permits the construction of a larger view from small images referred as ‘slits’ captured by a hidden camera with a limited field of view to meet the increasing demands of effective security surveillance systems. The slits are extracted from the video and are registered through a multi-scale iterative algorithm. The registered slits from a number of frames are integrated to obtain a panoramic view. Moreover, to limit the effect of instant illumination variations homomorphic filtering in frequency domain is applied to slits prior alignment. The illumination compensation on one hand improved the registration accuracy and on the other hand it helped to produce a perceptually better quality panorama picture. Large number of experiments were performed to measure the subjective as well as objective quality of the integrated images. The results show the effectiveness of the proposed solution. In future, we plan to extend the proposed approach for more challenging scenarios, e.g., surveillance with moving cameras.

## References

1. Aach T, Dumbgen L, Mester R, Toth D (2001) Bayesian illumination-invariant motion detection. In: Proceedings of the international conference image process. (ICIP), vol 3, pp 640–643
2. Agaian S, Silver B, Panetta K (2007) Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE Trans Image Process* 16(3):741–758
3. Bergen J, Anandan P, Hanna K, Hingorani R (1992) Hierarchical model-based motion estimation. In: Proceedings 2nd European conference on computer vision, ECCV '92. Springer, London, pp 237–252
4. Chang CH, Sato Y, Chuang YY (2014) Shape-preserving half-projective warps for image stitching. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 3254–3261
5. Cristofaro A, Salaris P, Pallottino L, Giannoni F, Bicchi A (2014) On time-optimal trajectories for differential drive vehicles with field-of-view constraints. In: IEEE 53rd annual conference on decision and control (CDC), pp 2191–2197
6. Dornaika F, Chakik F (2012) Efficient object detection and tracking in video sequences. *J Opt Soc Am A* 29(6):928–935
7. Elibol A, Gracias N, Garcia R, Kim J (2014) Graph theory approach for match reduction in image mosaicing. *J Opt Soc Am A* 31(4):773–782
8. Farid M, Mahmood A (2012) Image morphing in frequency domain. *J Math Imaging Vis* 42(1):50–63
9. Farid M, Mahmood A (2014) An image composition algorithm for handling global visual effects. *Multimed Tools Appl* 71(3):1699–1716
10. Fu Z, Wang L (2014) Optimized design of automatic image mosaic. *Multimed Tools Appl* 72(1):503–514
11. Guizar-Sicairos M, Thurman ST, Fienup JR (2008) Efficient subpixel image registration algorithms. *Opt Lett* 33(2):156–158
12. Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings Alvey vision conference, pp 23.1–23.6

13. Hou J, Wu C, Yuan Z, Tan J, Wang Q, Zhou Y (2008) Research of intelligent home security surveillance system based on zigbee. In: International symposium on intelligent information technology application workshops, pp 554–557
14. Kosmopoulos DI, Doulamis ND, Voulodimos AS (2012) Bayesian filter based behavior recognition in workflows allowing for user feedback. *Comput Vis Image Underst* 116(3):422–434. Special issue on Semantic Understanding of Human Behaviors in Image Sequences
15. Li B, Wang W, Ye H (2013) Multi-sensor image registration based on algebraic projective invariants. *Opt Express* 21(8):9824–9838
16. Li X, Cheah CC (2015) Robotic cell manipulation using optical tweezers with unknown trapping stiffness and limited fov. *IEEE/ASME Trans Mechatronics* 20(4):1624–1632
17. Lin WY, et al. (2011) Smoothly varying affine stitching. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 345–352
18. Litvinov A, Schechner YY (2005) Radiometric framework for image mosaicking. *J Opt Soc Am A* 22(5):839–848
19. Lowe D (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
20. Lu T, Li S, Fu W (2014) Fusion based seamless mosaic for remote sensing images. *Sensing and Imaging* 15(1):101
21. Microsoft Image Composite Editor. <http://research.microsoft.com/ivm/ice/>. Version 1.4.4.0
22. Monroe D (2003) Ground based security surveillance system for aircraft and other commercial vehicles. US Patent 6,545,601
23. Mubasher M, Farid M, Khaliq A, Yousaf M (2012) A parallel algorithm for change detection. In: Proceedings 15th international multitopic conference (INMIC), pp 201–208
24. Ng M, Wang W (2013) A variational approach for image stitching ii: using image gradients. *SIAM J Imaging Sci* 6(3):1345–1366
25. Phong B (1975) Illumination for computer generated pictures. *Commun ACM* 18(6):311–317
26. Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Carli M, Battisti F (2009) Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Adv Mod Radioelectron* 10(4):30–45
27. Prados R, Garcia R, Neumann L (2014) State of the art in image blending techniques. In: Image blending techniques and their application in underwater mosaicing. Springer, pp 35–60
28. Radke R, Andra S, Al-Kofahi O, Roysam B (2005) Image change detection algorithms: a systematic survey. *IEEE Trans Image Process* 14(3):294–307
29. Raty T (2010) Survey on contemporary remote surveillance systems for public safety. *IEEE Trans Syst, Man, Cybern C* 40(5):493–515
30. Shah M, Javed O, Shafique K (2007) Automated visual surveillance in realistic scenarios. *IEEE Multimed* 14(1):30–39
31. Silva F, Hiraga A, Artero A, Paiva M (2014) StitchingPHm-A new algorithm for panoramic images. *Pattern Recog Image Anal* 24(1):41–56
32. Srikantha A, Sidib D (2012) Ghost detection and removal for high dynamic range images: recent advances. *Signal Process Image Communc* 27(6):650–662
33. Suen ST, Lam EY, Wong KK (2007) Photographic stitching with optimized object and color matching based on image derivatives. *Opt Express* 15(12):7689–7696
34. Szwed P, Skrzynski P, Chmiel W (2014) Risk assessment for a video surveillance system based on fuzzy cognitive maps. *Multimed Tools Appl*:1–24
35. Uyttendaele M, Eden A, Skeliski R (2001) Eliminating ghosting and exposure artifacts in image mosaics. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), vol 2, pp II–509
36. Wang W, Ng M (2012) A variational method for multiple-image blending. *IEEE Trans Image Process* 21(4):1809–1822
37. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
38. Wu G, Wu Y, Jiao L, Wang YF, Chang EY (2003) Multi-camera spatio-temporal fusion and biased sequence-data learning for security surveillance. In: Proceedings of the ACM international conference on multimedia, pp 528–538
39. Xu Z (2013) Consistent image alignment for video mosaicing. *Signal Image Video Process* 7(1):129–135
40. Zagrouba E, Barhoumi W, Amri S (2009) An efficient image-mosaicing method based on multifeature matching. *Mach Vis Appl* 20(3):139–162
41. Zaragoza J, Chin TJ, Brown M, Suter D (2013) As-projective-as-possible image stitching with moving dlt. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 2339–2346

42. Zeng L, Zhang S, Zhang J, Zhang Y (2014) Dynamic image mosaic via sift and dynamic programming. *Mach Vis Appl* 25(5):1271–1282
43. Zeng L, Zhang W, Zhang S, Wang D (2014) Video image mosaic implement based on planar-mirror-based catadioptric system. *Signal Image Video Process* 8(6):1007–1014
44. Zitov B, Flusser J (2003) Image registration methods: a survey. *Image Vis Comput* 21(11):977–1000



**Hadia Tazeem** received her B.S. degree in Mathematics and Computer Science, M.Sc. and M.Phil. degrees in Computer Science from the University of the Punjab, Lahore, Pakistan in 2007, 2009 and 2013 respectively. She is working as Assistant Professor in School of Systems and Technology at University of Management and Technology and aiming for a Ph.D. in Computer Vision. Her research interests include computer vision and image processing, in particular video surveillance and image and video mosaicing.



**Muhammad Shahid Farid** received the B.S., M.Sc., and M.Phil. (Hons.) degrees in computer science from the University of the Punjab, Lahore, Pakistan, in 2004, 2006, and 2009, respectively, and the Ph.D. in computer science from University of Torino, Turin, Italy, in 2015. He is currently a Lecturer with the Punjab University College of Information Technology, Lahore, Pakistan. His research interests broadly include image segmentation, image morphing, and 3DTV, in particular, 3D video representation and coding, virtual view synthesis, and quality assessment of 3D videos.



**Arif Mahmood** received his Masters and the Ph.D. degrees in Computer Science from Lahore University of Management Sciences in 2003 and 2011 respectively. During MS he obtained Gold Medal for the best academic performance. Currently he is Postdoc researcher in the Department of Computer Science and Engineering, Qatar University, Doha. His current research direction is counting and tracking individuals in extremely dense crowds. Before this he worked as Research Assistant Professor with the School of Mathematics and Statistics, the University of the Western Australia. He worked on characterizing structure of complex networks using Machine Learning techniques. Before that he was Research Assistant Professor with the School of Computer Science and Software Engineering, UWA and worked on hyperspectral object recognition and action recognition using depth images. His major research interests are in Computer Vision and Pattern Recognition. More specifically he has performed research in data clustering, classification, action and object recognition using image sets.