

Pan Correction through Overlap Estimation in a Multi-Camera Environment

S. Mohamed Mansoor Roomi, V. Manojkumar and M. Senthilarasi
Department of Electronics and Communication,
Thiagarajar College of Engineering, Madurai, India

Abstract: Problem statement: Multiple cameras are employed for surveillance of larger environment. In such a case there is a need to maintain overlap in the adjacent cameras for correct object registration. Overlap may get disturbed by natural or manual factors. This study proposed an automatic camera pan correction by determining the area of overlap from multi-view images. **Approach:** A closed loop system which used feature extraction using SIFT, feature matching using descriptor ratio method and Mean Absolute Error (MAE) over Gaussian scale space, followed by overlap estimation is implemented for restoring the camera position. **Results:** The proposed method was experimented with the datasets acquired in the environment where surveillance involves two cameras. The matched points of the two images were used to calculate the overlap percentage. The overlap percentage estimated by the surveillance server was communicated to the pan controller to re-orient the camera to its original position. **Conclusion:** The proposed algorithm identified the robust and distinctive features that are invariant to translation, rotation and scaling. These features help in the accurate estimation of overlap percentage, which is further used to automatically correct the pan of the camera.

Key words: Overlapping region, pan, SIFT, MAE, Gaussian scale space

INTRODUCTION

In multiple images the location of common objects must be correctly registered, as this is the first step to many computer vision systems. Overlap in the Field Of View (FOV) of multiple cameras is needed for correct registering and labeling of an object in an environment monitored by a system with multiple cameras. Change in overlap in FOV due to camera pan will affect the correct working of such systems. A method automatically performing pan correction still remains one of the field's biggest problems. A primary use of information available in initial overlap is the basis for estimating camera pan and its correction. A common approach to the problem is to first detect a number of distinguishable features independently in each image and then to determine which features originate from the same region in the scene. Shape dependent features by edge detection (Mikolajczyk *et al.*, 2003) can be used, but discarding the poor edges is not easier. The image can be segmented into the regions and segmented regions can be used as features for matching (Mittal and Davis, 2001). Invariant moments and point based approaches are also used. But the earlier works show that the points extracted by SIFT algorithm gives better features (Mikolajczyk and Schmid, 2003). Scale

Invariant Feature Transform (SIFT) detects features over varied scales of an image (Lowe, 1999; 2004). RANSAC can be used to reduce the iterative processing to find points (Chen *et al.*, 1998). An algorithm to estimate Pan, Tilt and Zoom (PTZ) parameters of a PTZ camera are obtained from Frame-2-Frame (F2F) correspondences at different sampling rates for a real-time video surveillance and automatic object tracking system. PTZ estimation accuracy is maintained as long as its error is small (Wu *et al.*, 2006). A wide area surveillance system that detects tracks and classifies moving objects across multiple cameras has been presented (Javed *et al.*, 2003).

Image variance density is used to optimally estimate camera pan and tilt values by incrementally refining image registration using overlapping images from prior frames. Performance of this algorithm is reduced with increase in noise and fast moving objects (Song *et al.*, 2006). A method of automatically performing the registration of two range images that have significant overlap is discussed (Roth, 1999). In the study (Trajkovic, 2002), the camera position and orientation is determined by pointing the camera at several points in the area which is at the height of the camera and by applying the linear algorithm.

Corresponding Author: S. Mohamed Mansoor Roomi, Department of Electronics and Communication,
Thiagarajar College of Engineering, Madurai, India

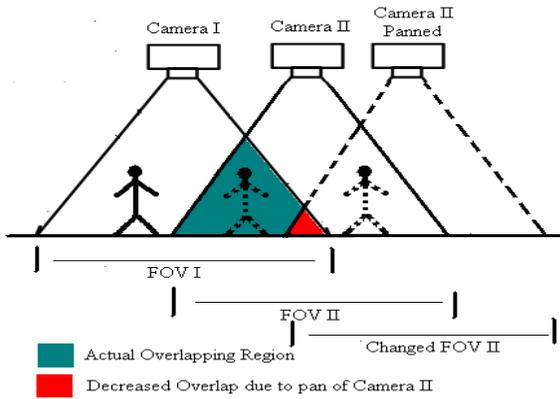


Fig. 1: Change in overlap FOV of adjacent cameras

The estimation of camera position and orientation is done only if the height of the camera and the initial points assumed are known. In the proposed study, feature matching is done through two steps: First being matching the SIFT descriptors by using descriptor ratio method followed by matching using correlation in scale space. In this study, attention is paid to the correction of camera movement in the situation when one of the cameras is moved due to an unknown cause. This makes the change in FOV of a camera which is depicted in the Fig. 1.

MATERIALS AND METHODS

This study proposes a method to automatically detect the camera pan due to external disturbances and restores the camera to its original position. The proposed method has four stages viz. Feature Extraction, Feature Matching, Overlap Estimation followed by final stage of camera pan correction. The overall flow of the proposed method is depicted in Fig. 2.

Feature extraction: Feature points are extracted from an image using SIFT algorithm proposed by Lowe (2004; 1999). The SIFT consists of four steps namely Detection of scale space extrema, Local extrema detection, Orientation assignment and Keypoint Descriptor formation which is explained below.

Building Gaussian scale space: Gaussian Scale space $L(x, y, \sigma)$ is constructed by convolving an initial image $I(x, y)$ with Gaussians of different variance $G(x, y, \sigma)$ which is expressed in Eq. 1. This contributes an octave. Next octave of Gaussian scale space is built from the initial image by down sampling it by a factor of 2 and this process is repeated. Adjacent Gaussian images are

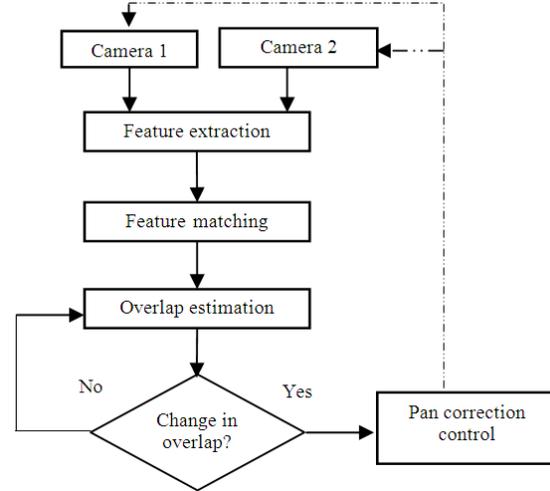


Fig. 2: Methodology

subtracted to produce the Difference of Gaussian $D(x, y, \sigma)$ as in Eq. 3. This approach of constructing Gaussian scale space makes the algorithm to be invariant of scaling. Interest points for SIFT features correspond to local extrema of Difference-of-Gaussian (DoG) images at different scales:

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (1)$$

where, \otimes is the convolution operation in x and y and:

$$G(x, y, \sigma) = \frac{1}{(2\pi\sigma^2)} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

Local extrema detection: Keypoints are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 26 neighbors in 3×3 regions at the current and adjacent scales. The keypoints with low contrast points and in edge responses are excluded.

Orientation assignment: For each image sample $L(x, y)$, the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ is computed using the pixel differences as shown below in Eq. 4 and 5. The key point orientation is determined from a gradient orientation histogram. This histogram is computed from the contribution of each neighboring pixel which is weighted by the gradient magnitude and a Gaussian window with a ' σ ' that is 1.5 time the scale of the key point. The peaks in the histogram which is greater than 80% of the maximum peak value are considered for feature extraction:

$$m(x, y) = \sqrt{\left((L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2 \right)} \quad (4)$$

$$\theta(x, y) = \tan^{-1} \left\{ \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right\} \quad (5)$$

Key point descriptor: A keypoint descriptor is formed using the gradient magnitude and orientation at each image sample point in a region around the keypoint location. The 128 element feature vector of a keypoint is obtained by 4×4 location descriptor with 8 orientation bins each. These descriptors are normalized to make it illumination invariant as normalization reduces the brightness and contrast changes. The feature points that are extracted through SIFT are robust and invariant to translation, rotation, scale and affine transforms.

Feature matching: Feature matching phase comprises of Descriptor Ratio method using the SIFT feature descriptors followed by Correlation based matching in the varied scales of image as shown in Fig. 3.

Descriptor ratio method: The descriptors obtained from SIFT are used for feature matching by descriptor ratio method. Dot product of descriptors of each keypoint in Camera 1 (Left) image with descriptors of all keypoints in Camera 2 (Right) image is computed and then inverse cosine is taken:

$$F = \cos^{-1} \left(D_i * \sum_{j=1}^n D_j \right) \quad (6)$$

Where:

- * = The dot operator
- D_i and D_j = Keypoint descriptors in camera 1 and 2 images each of size 1×128
- 'n' = The number of keypoints in camera 2 image

This approach is computationally cheaper and effective. The resulting angle values are sorted into a vector. The ratio of least value of angle vector 'c' and its successor value 'd' are compared with the threshold 'Th' (ranging from 0.1-0.9) for feature matching which is expressed below. This procedure is repeated as many times as the number of keypoints in camera 1 image:

$$\frac{c}{d} \leq Th \quad (7)$$

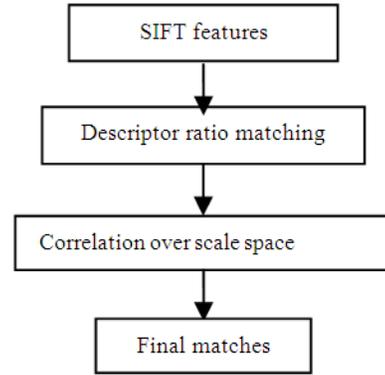


Fig. 3: Feature matching

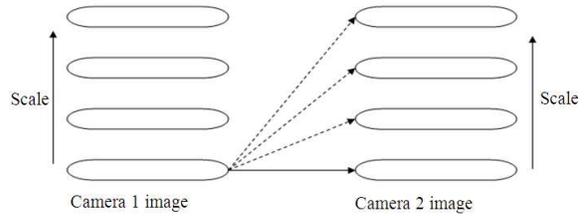


Fig. 4: Correlation in scale space

Correlation over scale space: Correlation over scale space is used for feature matching of keypoints which are obtained as matched features from the previous step. This matching is done over the different scale space of the Camera 1 and 2 image as shown in Fig. 4.

Let (x_i, y_i) be the coordinates of matched features in camera 1 image and (x'_i, y'_i) be the coordinates of matched feature points in camera 2 image, 'R'. Consider a neighborhood of size $(2r+1)$ for each feature points. 'N' is the number of scales in scale space of reference image and sensed image. The Mean Absolute Error (MAE), 'e' using the neighborhood of each point is given by Eq. 8:

$$e(i(k,l)) = \frac{\sum_{u=-r}^r \sum_{v=-r}^r \left(L_{(k,l)}(x_i + u, y_i + v) - R_{(k,l)}(x'_i + u, y'_i + v) \right)}{(2r+1)^2} \quad (8)$$

Where:

- $i = 1, 2, 3, \dots, M$
- $k = 1, 2, 3, \dots, N$
- $l = 1, 2, 3, \dots, N$

Error 'e' is calculated for all the points in all the scale pairs of 'k' and 'l'. For each matched point the scale pair ('k' and 'l') producing the minimum error is detected.

For ‘M’ feature points, there will be ‘M’ pair of scales that produces the minimum error. Now the pair of scales that produces the most number of minimum errors among the ‘M’ pairs is detected and is considered to be the closest scale between the camera 1 and 2 image. Let ‘CsL’ and ‘CsR’ be the closest scales of reference image and sensed image. After finding the closest pair of scales, the mean of errors M_e for all the feature point pair in the closest scale pair is computed as given in Eq. 9:

$$M_e = \left\{ \frac{\sum_{i=1}^M e(i(k = CsL, l = CsR))}{i} \right\} \quad (9)$$

The pair of M points in closest scale pair, having their error less than the mean error is considered to be set of final matched points (m):

$$m = \{M : M \in e(i(k = CsL, l = CsR))\} \quad (10)$$

Only the true feature points that are located in common FOV of both cameras gets matched. Those points that are located outside the overlapping regions are classified as mismatches and are rejected from further processing.

Overlap estimation: The points that are deduced as matched feature points are used to determine the percentage of overlap. The region where the feature points get matched in both Camera 1 (Left-L) and 2 (Right-R) images is the overlap region ($L \cap R$) as shown in Fig. 5. The overlap percentage in both Camera views is the percentage composition of ($L \cap R$) in both the images considered together:

$$\text{Overlap}(\%) = \frac{L \cap R}{(L \cup R) - (L \cap R)} \times 100 \quad (11)$$

Pan correction control: The camera pan (unwanted movement due to external forces) is corrected by the change in the percentage of overlap as shown in Fig. 6. An electrical signal is generated based on the change in overlap percentage and is sent to the external hardware through the parallel port. External hardware senses the electrical signal and runs a motor to correct the position of camera.

Change in overlap: The initial overlap between the Camera 1 and 2 images is IOP. If the camera undergoes

a pan then the two cameras have a new overlap percentage of NOP. The deviation of the overlap percentage is said to be the error ξ in overlap:

$$\xi = \text{IOP} - \text{NOP} \quad (12)$$

Let $R_{\text{mat}}(X)$ and $P_{\text{mat}}(X)$ be the x-coordinates of any matched feature point in reference image and panned image. The direction of pan of camera can be deduced by calculating the difference between the coordinate values:

$$\text{dir} = R_{\text{mat}}(X) - P_{\text{mat}}(X) \quad (13)$$

An LUT (Look up Table 1) has been formulated for this purpose. The look up Table 1 matches the possible error overlaps with the corresponding time period needed to regain the camera back to its original position in such a way that the error gets nullified. Once the error gets nullified, it is said that the two cameras cover a particular region with the predefined overlap.

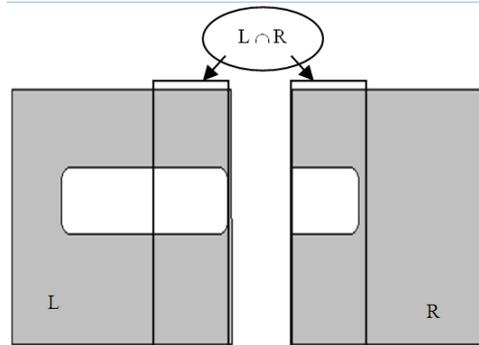


Fig. 5: Overlap percentage calculation

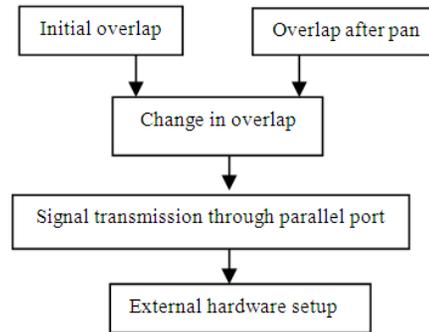


Fig. 6: Framework of pan correction stage

Table 1: Percentage of overlap in camera 1 and 2

IOP in camera 1	IOP in camera 2	NOP in camera 1 after pan	NOP in camera 2 after pan
19.514	30.081	9.777	8.6587



Fig.7: Hardware setup for stepper motor control for pan correction

Signal transmission through parallel port: The parallel port cable DB25 is connected to the parallel port of the controller. The Control signal to drive the external hardware circuitry is sent through the parallel port of computer.

External hardware setup: External hardware setup consists of a parallel port data cable which connects the parallel port of the computer system to the hardware setup and the Stepper Motor. The stepper motor attaches the camera for controlling the panning mechanism. The error signal is sent through the parallel port to the hardware setup. The setup buffers and amplifies the signal to drive the stepper motor to correct the camera pan. The signal is sent through the 2nd, 3rd, 4th and 5th pin of the parallel port of the computer. The hardware setup consists of a buffer 74HC244 which receives the data from the computer at 2nd, 4th, 6th and 8th pins. The buffered data are sent through 13th, 15th, 17th and 19th pins. The buffered data is then fed as input to the SLA4061. The SLA4061 amplifies the signal and then the amplified signal is used to control the stepper motor. The hardware setup is shown in the Fig. 7.

RESULTS

The proposed method is experimented in the environment where two cameras (Camera 1 and 2) are employed to acquire a scene. The images obtained from these cameras have a constant overlap percentage in their FOV. The scene captured by the camera 1 and 2 is shown in Fig. 8a and b. The overlap percentage between these two images is estimated as per the proposed algorithm. For experimentation purpose, the camera 2 is subjected to have a change in its pan.



Fig. 8a: FOV of camera 1 (Left)



Fig. 8b: FOV of camera 2 (right)



Fig. 8c: FOV of panned camera 2

This disturbance in pan makes the FOV of the camera 2 to change. Hence this change in FOV of the camera results in the deviation of its overlap percentage from the initial stage. The FOV of the panned camera 2 is shown in Fig. 8c.

DISCUSSION

Feature points are extracted separately in the camera 1, 2 and panned camera 2 images using the SIFT algorithm. The extracted points from these images are matched with one another using descriptor ratio method and correlation ratio method. The matching is done between the image pairs of reference Camera 1 and 2 image (pair1), Camera 1 and panned Camera 2 image (pair2) and reference Camera 2 image and panned Camera 2 images (pair3). The final matched points are shown in Fig. 9a-f as white colored '*' marks along with their corresponding red colored labels.



Fig. 9a: Final matched points in camera 1 image computed from image pair1



Fig. 9b: Final matched points in camera 2 image computed from image pair1



Fig. 9c: Final matched points in camera 1 image computed from image pair2



Fig. 9d: Final Matched points in panned camera 2 image computed from image pair2



Fig. 9e: Final matched points in reference camera 2 image computed from image pair3



Fig. 9f: Final matched points in panned camera 2 image computed from image pair3



Fig. 10a: Overlapping region in camera 1 image computed using image pair1



Fig. 10e: Overlapping region in camera 2 image computed using image pair3



Fig. 10b: Overlapping region in camera 2 image computed using image pair1



Fig. 10f: Overlapping region in panned camera 2 image computed using image pair3



Fig. 10c: Overlapping region in camera 1 image computed using image pair2



Fig. 11: Change in overlap estimation



Fig. 10d: Overlapping region in panned camera 2 image computed using image pair2

The coordinate values of the matched feature between the reference camera 2 image and panned camera 2 image are used to detect the direction of pan. The reduction in percentage of overlap is calculated and is used to generate the electrical signal to control the stepper motor.

In the Fig. 11, the region of overlap that got reduced is indicated by red region. The blue region indicates the new overlap region. The common area of overlap in reference image of camera 2 and panned image is indicated by white colored vertical lines. The direction of pan correction is estimated from the overlap percentage. And finally the electrical signal is transmitted to the external hardware setup to correct the camera pan.



Fig. 12a: Camera 1 (left)



Fig. 12e: Overlapping region in camera 1



Fig. 12b: Camera 2 (right)



Fig. 12f: Overlapping region in camera 2



Fig. 12c: Matched feature points in camera



Fig. 12d: Matched feature points in camera 2

The proposed algorithm is tested for the video data taken using two web cameras simultaneously in Digital Signal Processing Lab in Electronics and Communication Department. Acquired video is converted to AVI format for simulation purpose. The resolution of the video is 320×240 and the frame rate is $15 \text{ frames sec}^{-1}$. The Fig. 12a and b show the first frame captured using two cameras namely Camera 1 and 2. The initial overlapping region in the camera views is determined by the proposed algorithm as shown in Fig. 12e and f. This overlapping region of initial frame of a video can be used for correct labeling of an object in view of multiple cameras for video tracking.

CONCLUSION

The proposed work aims at automatically correcting the camera pan by estimating the percentage of overlap between adjacent camera images. The algorithm implemented in this study has the advantage of identifying the robust and distinctive features that are invariant to translation, rotation and scaling. Descriptor ratio method used for feature matching reduces the computation complexity. The issue due to different zooms of camera is resolved by using Correlation over scale space. The overlapping error is used to correct the

panned camera orientation. Extensive simulations were carried out on a variety of images and the proposed algorithm was found to perform with accuracy.

REFERENCES

- Chen, C.S., Y.P. Hung and J.B. Cheng, 1998. A fast automatic method for registration of partially-overlapping range images. Proceedings of the 6th International Conference on Computer Vision, Jan. 4-7, IEEE Xplore Press, Bombay, pp: 242-248. DOI: 10.1109/ICCV.1998.710725
- Javed, O. and Z. Rasheed, O. Alatas and M. Shah, 2003. KNIGHTTM: A real time surveillance system for multiple overlapping and non-overlapping cameras. Proceedings of the International Conference on Multimedia and Expo, Special Session on Multi-Camera Surveillance Systems, July 6-9, IEEE Computer Society, Baltimore, pp: 1-4. DOI: 10.1.1.3.3529
- Lowe, D.G., 1999. Object recognition from local scale-invariant features. Proceeding of the International Conference on Computer Vision, Sept. 20-25, IEEE Computer Society, Washington DC., USA., pp: 1150-1157.
<http://portal.acm.org/citation.cfm?id=850924.851523>
- Lowe, D.G., 2004. Distinctive image features from scale-invariant key points. *Int. J. Comput. Vis.*, 60: 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94
- Mikolajczyk, K. and C. Schmid, 2003. A performance evaluation of local descriptors. *IEEE Trans. Patt. Anal. Mach. Intell.*, 27: 1615-1630. DOI: 10.1109/TPAMI.2005.188
- Mikolajczyk, K., A. Zisserman and C. Schmid, 2003. Shape recognition with edge-based feature. Proceedings of the British Machine Vision Conference, Sept. 9-11, Springer, Norwich, UK., pp: 779-788.
<http://www.robots.ox.ac.uk/~vgg/publications/papers/mikolajczyk03a.pdf>
- Mittal, A. and L. Davis, 2001. Unified multi-camera detection and tracking using region-matching. Proceedings of the IEEE Workshop on Multi-Object Tracking, (MOT'01), IEEE Xplore Press, Vancouver, BC, pp: 3-10. DOI: 10.1109/MOT.2001.937975
- Roth, G., 1999. Registering two overlapping range images. Proceedings of the 2nd International Conference on 3-D Digital Imaging and Modeling, (DIM'99), IEEE Xplore Press, Ottawa, Canada, pp: 191-200. DOI: 10.1109/IM.1999.805349
- Song, D., N. Qin, K. Goldberg, 2006. A minimum variance calibration algorithm for pan-tilt robotic cameras in natural environments, robotics and automation. Proceedings of the IEEE International Conference on Robotics and Automation, May 15-19, IEEE Computer Society, Florida, USA, pp: 3449-3456. DOI: 10.1109/ROBOT.2006.1642229
- Trajkovic, M., 2002. Interactive calibration of a PTZ camera for surveillance applications. Proceedings of the 5th Asian Conference on Computer Vision, Jan. 23-25, NAKIDO, Melbourne, pp: 507-514.
- Wu, S., T. Zhao, C. Broaddus, C. Yang and M. Aggarwal, 2006. Robust pan, tilt and zoom estimation for PTZ camera by using meta data and/or frame-to-frame correspondences. Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision, Dec. 5-8, IEEE Xplore Press, Singapore, pp: 1-7. DOI: 10.1109/ICARCV.2006.345423