Automatic alignment of infrared video frames for equipment leak detection

Liming Zhou*, Yousheng Zeng

Providence Engineering and Environmental Group LLC, Baton Rouge, LA, USA

Abstract

Petroleum refining and petrochemical industries start using infrared (IR) cameras to detect volatile organic compounds (VOC) leaking out of process equipment. However, further quantitative processing of these video image data or automatic recognition of VOC plumes are hindered by unaligned video frames owing to the slight vibrations of the camera. An automatic method is developed to align the IR video frames as a preprocessing procedure for other possible video processing methods. The alignment method is based on a two-dimensional spatial Fourier transform. The accuracy can reach fractional pixels in estimation of translational shift and 1–2° for rotational shift. Temporal Fourier transform of actual industrial tests of IR videos is performed with both unaligned and aligned video frames. The results indicate that only after the alignment of the video frames, the camera motion interferences on VOC plume identification can be eliminated or minimized, and the VOC plume can be identified through investigating the characteristic flickering frequency power in the temporal Fourier transform. This alignment method provides a useful tool for IR or other optical video image data preprocessing purposes.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Video frame alignment; Volatile organic compounds leak detection; Image registration; Leak detection and repair

1. Introduction

Since the 1980s, the U.S. industries including petroleum refineries and organic chemical manufacturers are required to periodically detect and repair leaks of volatile organic compounds (VOC) from equipment components such as valves, connectors, pump seals, etc. This requirement is referred to as leak detection and repair (LDAR) program. The current work practice entails using a portable flame ionization detector (FID) monitor to check ("sniff") the seal around the components for possible leaks [1]. Because of a large number of components (hundreds of thousands) at a single facility, such a work program results in very high labor cost even if it is only performed quarterly. The work is tedious. Possible operator errors also reduce the reliability of the results.

Recently, infrared (IR) video cameras have been investigated for VOC leak detection as a cost-effective alternative [2,3]. The approach of using IR cameras for leak detection is often referred to as "Smart LDAR". These cameras are operated at a predetermined wavelength band with strong VOC absorptions. When VOC is emitted from a leaking component, its temperature drops during the initial expansion, causing a temperature difference between the VOC plume and the surrounding air. The temperature difference generates a net radiation difference by the gas absorption and an image can be formed on the detector array of the camera. The intensities of the image pixels depend on the gas concentration and camera sensitivity. This kind of cameras can be operated manually by operators to scan all possible leaking components. Usually, several seconds of video length are enough for an operator to identify the leaking gas plume within an area covered by the viewfinder of the camera. The camera can also be mounted at a fixed location to continuously monitor a large operational area within a refinery or a chemical plant. The fugitive VOC emissions due to equipment leak can be reduced if the leakers can be found and repaired promptly since the duration of leaking is minimized.

To further reduce labor cost, it is desirable to process the IR video automatically. The IR video data processing includes automatic identification of VOC plume in a non-attendant manner, quantification of the relationship between the image and

---

* Corresponding author. Tel.: +1 2257667400; fax: +1 2257667440.
 E-mail address: limingzhou@providenceeng.com (L. Zhou).

the actual VOC leaking rate, and compression of the video images to save data storage space. In order to accomplish some of these automated tasks, frames in the IR video need to be closely aligned to the same frame of reference. A video footage consists of many frames. These frames in a raw video footage are usually not in the same reference spatially as the camera shift positions during recording. When the camera experience moving, shaking, or vibration, a frame may capture a slightly different scene compared to its immediate preceding frame or the captured scene may be rotated. The unstable camera may be caused by the operator, process equipment, or strong wind if it is mounted on a tall structure. For practical industrial applications, this vibration becomes inevitable and the unaligned frames will make those image processing algorithms fail. Therefore, the video frames have to be aligned before a quantitative processing can be performed.

To transform an image to match another image is called image registration. The alignment of the video frames includes a series of image registration process. At present, all image processing software packages use manually selected control points for image registration and only provide local registration. Reddy and Chatterji [4] summarized different automatic registration methods and improved the efficiency of the frequency domain (FFT based) registration method. This method is a global registration and is not biased by moving foreground objects in the video images [5]. Xie et al. [6] used this method for registration of satellite images. In this study, an automatic registration method based on frequency domain is used for video frame alignment purposes.

In this work, the Reddy algorithm will be implemented in MATLAB and tested for both simulated data and IR video data from real industrial applications. A temporal frequency based smoke (or plume) detection method is used to evaluate the effect of the video frame alignment.

2. Method

The spatial fast Fourier transform (FFT) based image registration algorithm relies on the Fourier shift theorem [4]. For two images I_1 and I_2 with only difference by shift dx in the horizontal direction and dy in the vertical direction, as shown in Eq. (1), their Fourier transforms are related by Eq. (2):

$$I_2(x, y) = I_1(x - dx, y - dy) \qquad (1)$$

$$F_2(\xi, \eta) = e^{-j2\pi(\xi\ dx + \eta\ dy)} F_1(\xi, \eta) \qquad (2)$$

where ξ and η is the vertical and horizontal frequency, respectively, and j is the imaginary number symbol.

The phase shift R of the two images I_1 and I_2 is defined as

$$R = e^{-j2\pi(\xi\ dx + \eta\ dy)} = \frac{F_1(\xi, \eta)F_2^*(\xi, \eta)}{|F_1(\xi, \eta)||F_2(\xi, \eta)|} \qquad (3)$$

where * represents the complex conjugate and $||$ represents absolute value. For certain frequencies, the absolute value may be close to zero and then the R value needs to be directly set to zero. However, this situation is rare in image processing. Then

an inverse Fourier transform to the phase shift results in a Dirac delta function with the offset same to the translational motion:

$$\delta(x - dx, y - dy) = F^{-1}(R) = F^{-1}(e^{-j2\pi(\xi\ dx + \eta\ dy)}) \qquad (4)$$

By finding the location of the maximum value, the translational amount can be determined. To register scaled and rotated images, the $|F(\xi, \eta)|$ needs to be converted from Cartesian rectangular coordinates into log–polar coordinates $F\ lp(\log\ \rho, \theta)$ as indicated in Eqs. (5) and (6). The center of the new image will be the low frequency components of $|F(\xi, \eta)|$. Then, Eqs. (3) and (4) can be used again to compute the scaling and rotations:

$$x = e^{\log(\rho)}\ \cos(\theta) \qquad (5)$$

$$y = e^{\log(\rho)}\ \sin(\theta) \qquad (6)$$

A bilinear interpolation is used to find the value on the log–polar grids from the original rectangular grids, and the values outside of the original grids are set to zero [6]. To find the new maximum value $M(x,y)$, corresponding to an $F\ lp(\log\ \rho, \theta)$ value on a grid point, the four adjacent intensities $M_{j,k}$, $M_{j+1,k}$, $M_{j,k+1}$, and $M_{j+1,k+1}$ on original grid points $(j,k), (j+1,k) (j,k+1)$, and $(j+1, k+1)$ are used:

$$M(x, y) = M_{j,k}(1 - t)(1 - u) + M_{j+1,k}t(1 - u)$$
$$+ M_{j,k+1}(1 - t)u + M_{j+1,k+1}tu \qquad (7)$$

where t and u is the fractional parts of x and y, respectively.

The above method for estimation of translations gives an accuracy of one pixel. Gibson et al. [7] developed a method to improve the accuracy to fractional pixels. His basic idea is to find the two grid points (x_1, y_1) and (x_2, y_2) between which the true translation amount is located. The true translation between the two points is estimated by the following equations:

$$x = \frac{w_{x1}x_1 + w_{x2}x_2}{w_{x1} + w_{x2}} \qquad (8)$$

$$y = \frac{w_{y1}y_1 + w_{y2}y_2}{w_y + w_{y2}} \qquad (9)$$

where

$$w_{xi} = f(|F(x_i, y_1)|) + f(|F(x_i, y_2)|) \qquad (10)$$

$$w_{yi} = f(|F(x_1, y_i)|) + f(|F(x_2, y_i)|) \qquad (11)$$

In Eqs. (10) and (11), f is an empirical function. The best f function is described in Gibson et al. [7]. For general image registration purposes, the fractional pixel accuracy is good enough. However, the IR video clip processing may still be influenced by the fractional pixel error and a special filter is used in this study to eliminate this influence.

The algorithm for image registration used in this work is summarized into the following steps:

1. Choose a reference frame I_1 for the video clip to be aligned. The frame I_2 is to be aligned to I_1.
2. Apply FFT to images I_1 and I_2 to obtain the Fourier transforms F_1 and F_2.

3. Transform the absolute values of F_1 and F_2 from Cartesian coordinates into log–polar coordinates to obtain $F\ lp_1$ and $F\ lp_2$.
4. Apply FFT to $F\ lp_1$ and $F\ lp_2$ and calculate the phase transfer R by Eq. (3) and then calculate P, the inverse Fourier transfer of R by Eq. (4).
   4.1 Find the maximum $|P|$ value at (x_1, y_1).
   4.2 Among the four points $(x_1 \pm 1, y_1 \pm 1)$, select point (x_2, y_2) with the largest $|P|$ value.
   4.3 Use Eqs. (8)–(11) with $f(z) = z^\alpha$ and $\alpha = 1.55$ to find the scaling and rotation factors. Rescale and rotate the original images to update I_2.
5. Apply FFT for I_1 and I_2 and calculate the phase transfer ratio of these two Fourier transforms.
6. Calculate the inverse Fourier transform obtained in Step 5. Use similar procedures in Steps 4.1–4.3 with $\alpha = 0.65$ to calculate the shift. Then do reverse shift to image I_2.
7. An erosion-dilation filter is used for the difference image of I_1 and I_2. The erosion filter is a process using the minimum value of all eight neighboring pixels and the current pixel to replace the current pixel value. The dilation filter is a process using the maximum value of all eight neighboring pixels and the current pixel to replace the current pixel value. The filtered difference image is then added back to I_1 to generate the finalized I_2. This step is optional and different filter may be used for different processing purposes.
8. Replace I_2 with the next frame to be aligned, go to 2. If no more frames to align, end.

When doing image registration, one image needs to be set as a target and the other will be transformed to match the target. The transformed image will produce missing values and in this work, these missing values are filled with the target values at the same positions. For the video frames, translations and rotations are likely to happen while scaling is unlikely since the focal length will not be influenced by external disturbances.

Therefore, although scaling factors are calculated, the scaling effects will not be considered in both simulations and actual IR video alignment tests so as to avoid any possible unnecessary errors in estimating scaling factors.

3. Results and discussion

3.1. Simulated data

In Fig. 1a and b, are two images need to be aligned in this simulation. Image (a) is the target and (b) will be transformed to match a. Both Images (a) and (b) have the width of 123 pixels and height of 96 pixels. Image c is a direct overlay of a and b. Image b is created by translating Image a to the right by 9 pixels, translating to the bottom by 5 pixels and rotating anticlockwise by 17°.

Fig. 2 shows the Dirac delta function for the rotation and the shift. The maximum value in the upper figure is at (15, 1), indicating the rotation is 15° and no scaling. The maximum value in the lower figure is at (92, 115). Since the two values are exceeding their respective half width and half height
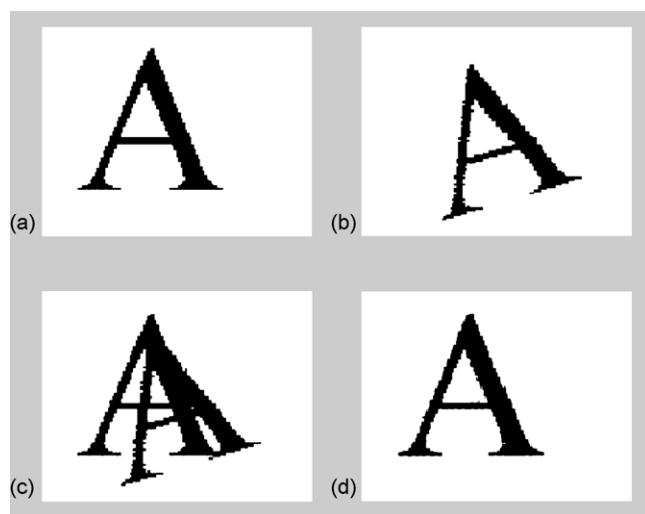

Fig. 1. Simulated images and alignment: (a) target image, (b) image to be aligned, (c) overlay of target image and unaligned image, and (d) overlay of target image and aligned image.
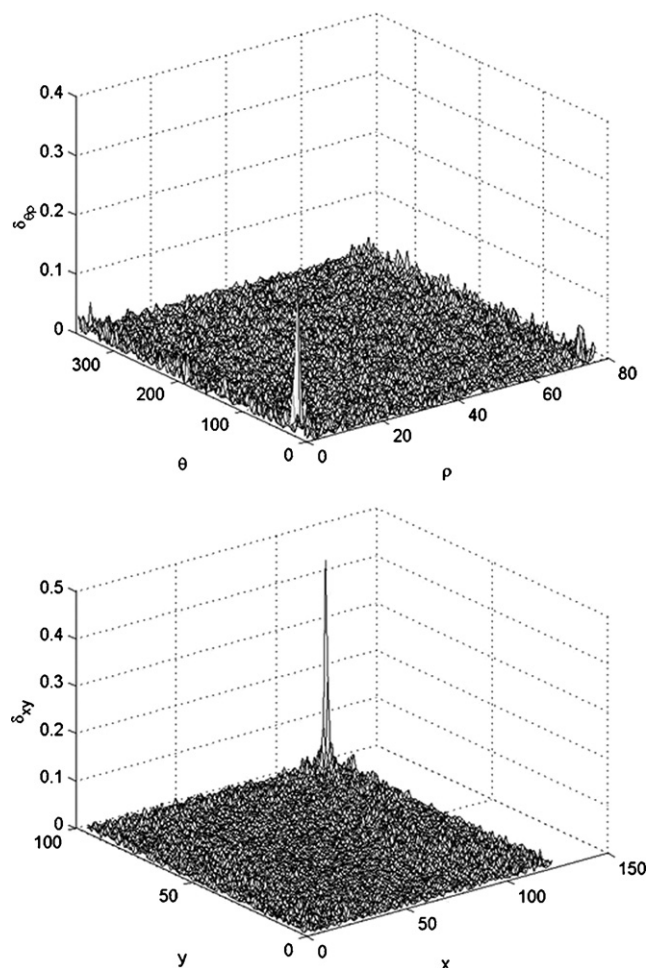

Fig. 2. Delta functions of aligning simulated images: (up) delta functions to determine rotation and scaling and (down) delta functions to determine translation.

Fig. 3. Alignment of actual IR videos—left column: "tank" video; right column: "pipe" video, (top) selected frame images from the videos, (middle) "Flickering" images without alignment, and (bottom) "Flickering" images with alignment.

of the image, Image b in Fig. 1 should be shifted toward the origin and the amount of shift is the width and height minus the maximum value location, *i.e.* horizontally 9 pixels and vertically 5 pixels (where the height and width is added by one pixel since a location exactly at the height or width indicates one pixel shift).

After the interpolation, the calculated translation is 9.28 and 5.37 pixels and the calculated rotation is 16.2°. After applying these transforms to Image b, the transformed image is overlaid on Image a and Image d is obtained (without using the filter in Step 7). From Image d we can see that Image b is well aligned to Image a. For the simulated images, the interpolation by Gibson [7] gives a little worse estimation for the translations but better estimation for the rotation. Since the shift in real video frames probably contains fractional pixels other than exact pixels, Gibson's method will give a better estimation most of the time.
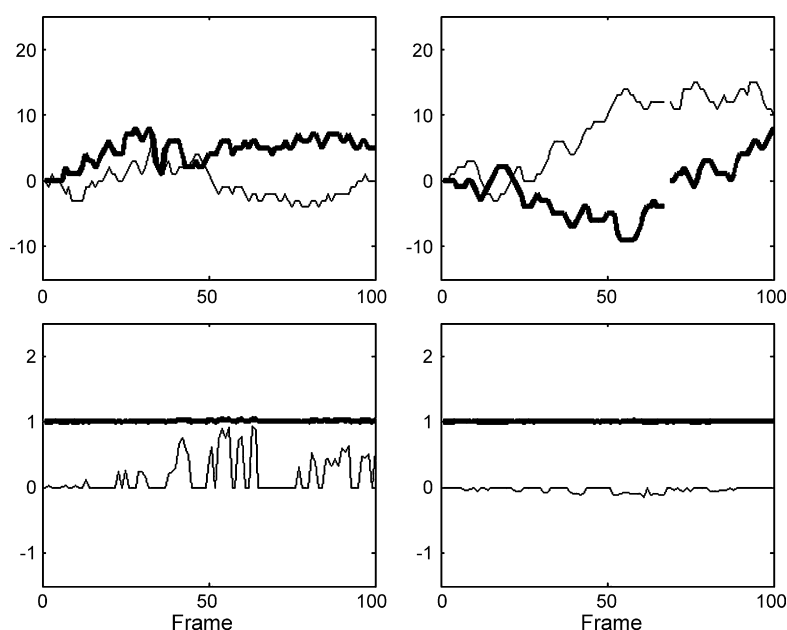

Fig. 4. Translational shift and rotational shift of actual IR videos—left column: "tank" video; right column: "pipe" video, (top) translational shift (thick line-vertical, thin line-horizontal), and (bottom) rotational and scaling shift (thick line-scaling, thin line-rotation).

3.2. Actual video data

Two IR video clips taken from a chemical plant are analyzed. The IR video camera is manufactured by FLIR Systems. Each of the video contains 100 frames at 23 frames s^{-1}. The top two images in Fig. 3 are snapshots from each individual video and the snapshot images are presented to provide a general idea of the VOC leaking situations. The left contains one leaker at the top of the tank (the video is named "tank" hereafter) and the right has two leaking valves (the video is named "pipe"). The "pipe" video has a blank frame and that frame is replaced by its previous frame. Each video was taken with the IR camera pointed to the leaking equipment. However, the frames in each video were constantly shifting due to the movement of the camera operator. The VOC plumes in the videos are easily recognizable by human eyes. The challenge is to recognize the VOC plume using some image processing systems without human intervention. This task is even more challenging when the video frames are constantly shifting and rotating due to an unsteady camera.

When the gas is released into the air, the concentration fluctuates at certain frequencies caused by atmospheric turbulence, similar to fire and smoke motions in the air. This characteristic flickering frequency is at 1–5 Hz. The pixel intensity at a certain location of all frames forms a time series. Frequency based method, such as wavelet or Fourier transform, can be used to identify smoke in videos through processing the pixel intensity time series [8]. Without any frame alignment, Fourier transforms are performed on the frames of original video clips "tank" and "pipe" respectively. The 1 Hz Fourier power forms a new "flickering" image. As shown in the two middle images of Fig. 3, without alignment, the Fourier transform shows no obvious sign of leaking gas but high power of tank edges and pipes caused by camera motions.

By using the algorithm described in Section 2 (with Step 7), the frames of the two videos are aligned to their respective first frames. Although one of the other frames may be the optimal target with the minimum shift to all frames, the shifts are random and the effect of choosing the optimal target frame is not significant. During the alignment process, the amount of translational and rotational shift is recorded and presented in Fig. 4. The horizontal and vertical shift may be up to approximately 15 pixels. The scaling factors are always 1, suggesting the scaling effect can be neglected. There is a little rotation up to 1° in the "tank" video.

After the alignment, Fourier transform is performed to form the "flickering" image. In the bottom row images of Fig. 3, the "flickering" images show hot spots for the leaking gas. The noisy intensities of equipment edges caused by the camera motions are minimized and the effect of Fourier transform is greatly improved by aligning the video frames.

Since this method will be used for real time video processing, the computation time is important. The MATLAB program used in this study usually took several seconds to complete the alignment. The current temporal Fourier transformation is performed for all pixels. If some pixels can be excluded by preprocessing methods, the overall computation speed may be improved.

4. Conclusion

Spatial Fourier transform method is used to align IR video frames for VOC leak detection. After alignment, temporal frequency analysis can identify the "flickering" frequency characteristics with the interferences by the camera motion eliminated or minimized. Further efforts to find better VOC plume identification method will be made to improve the ability of plume identification and to handle moving objects captured in the video.

The video frame alignment method introduced in this work is fast and robust. As a preprocessing method, it will also be useful for a wide range of other video data processing purposes, including, but not limited to, hyperspectral video images, VOC emission rate quantification based on IR camera videos, and other video processing applications involving plume-like targets.

References

[1] U.S. EPA, "Protocol for Equipment Leak Emission Estimates", EPA-453/R-95-017, November 1995.
[2] ENVIRON, 2004. "Development of Emissions Factors and/or Correlation Equations for Gas Leak Detection, and the Development of an EPA Protocol for the Use of a Gas-imaging Device as an Alternative or Supplement to Current Leak Detection and Evaluation Methods", Final Report to Texas Council on Environmental Technology and the Texas Commission on Environmental Quality, October, 2004.
[3] M. Lev-On, H. Taback, D. Epperson, J. Siegell, L. Gilmer, K. Ritterf, Methods for quantification of mass emissions from leaking process equipment when using optical imaging for leak detection, Chem. Eng. Prog. V25 (2006) 49–55.
[4] B.S. Reddy, B.N. Chatterji, A FFT-based technique for translation, rotation, and scale-invariant image registration, IEEE Trans. Image Process. 5 (8) (1996).
[5] Davis, J. Mosaics of Scenes with Moving Objects, IEEE Proceedings of the CVPR'98, June 1998 Santa Barbara, CA. Also available: http://graphics.stanford.edu/~jedavis/panorama/.
[6] H. Xie, N. Hicks, G.R. Keller, H. Huang, V. Kreinovich, Comput. Geosci. 29 (2003) 1045–1055.
[7] S. Gibson, V. Kreinovich, L. Longpre, B. Penn, S.A. Starks, Intelligent mining in image databases, with applications to satellite imaging and to web search, in: A. Kandel, H. Bunke, M. Last (Eds.), Data Mining and Computational Intelligence, Springer, Berlin, 2001, pp. 309–336.
[8] B.U. Töreyin, Y. Dedeoğlu, A.E. Çetin, Wavelet based real-time smoke detection in video, in: Proceedings of the 13th European Signal Processing Conference EUSIPCO 2005, Antalya, 4–8 Eylül 2005.