Image Registration for Gait Analysis of Farm Animals

Jaemin Son[1,a)] Yasushi Makihara[1,b)] Yasushi Yagi[1,c)] Morito Shiohara[2,d)]

Abstract: Development of gait recognition technology has opened new realm of application, gait diagnosis, which aims to diagnose disease related with gait. This challenge, if successfully applied to farm animals, will drastically reduce the cost for managing farm animals. As the first step to analyzing gait, silhouettes of subjects are essential from which most of gait features are extracted. Importantly, a sequence of images from a static perspective is desirable for background subtraction-based extraction. However, a static camera is unavailable in the farms due to probable infection of diseases from animals and only hand-held cameras are being allowed. In this study, we propose a method that registers images from a hand-held camera in order to overcome such constraints in the farms. For this purpose, we exploited fences to estimate the position of the camera and aligned frames to based on the initial frame. An experiment showed that registration worked well to some extent in terms of adjacent frames, although accumulation errors still remained.

Keywords: Vanishing point, Image registration, Gait recognition, Farm animals, Hand-held camera

1. Introduction

The outstanding development of Information Technology (IT) has directly advanced quality of daily lives furnishing with various convenient applications geared with complex functions and providing accommodative environment connected to network anytime anywhere. The IT also improved the efficiency of work in the primary industry by replacing human labors with intelligent machinery on mundane or dangerous work. One of the unconquered realm that human labors still have dominantly played a major role but its amount of work to be done outweighs limited resource of labors may include diagnosis of farm animals. Literally, flocks of farm animals cannot be fully under the care of veterinarians. Diagnosing disease for the farm animals with a matured IT technology, if possible, would bring about a new evolution in the agricultural industry where accurate, thorough and inexpensive diagnosis for pervasive entities comes into reality.

One of the farm animals that need thorough and constant care is cattle. It is said that about a third of cattle that are affected by hoof disease would die earlier than healthy cattle when not treated appropriately. Considering that cattle are expensive and invaluable resource in farming, such a loss should be avoided.

There is previous research that attempts automatic diagnosis of hoof diseases for cattle with the IT. Okada et al. [13] proposed an easier detection method for hoof diseases where they examined the effectiveness of tri-axial accelerometer and gait score in different kinds of hoof diseases. It was possible to cluster cattle with hoof diseases from healthy cattle based on vector sum of acceleration and locomotion score that is marked by veterinarians. However, attaching acceleration sensors to all entities of cattle takes enormous amount of labors and costs. In this situation, currently explored field, gait diagnosis based on video can provide solutions that can reduce costs, if successfully applied.

Such video-based gait analysis has been studied in the biometrics field by many researchers around the globe [8], [12], [22]. As the recognition performance improves, gait recognition techniques become to be applied to diagnosing disease. Gait recognition techniques that were used in individual identification now play an role in disease identification. Changing the subject for identification, we hope to see whether it is possible to apply the same techniques to animals, especially cattle.

The gait recognition techniques divides two folds, model-based and appearance-based approaches. The model-based approaches [1], [5], [7], [21] try to fit the gait information into a model. Therefore, accurate models that can extract prominent features of gait is essential in this approach. The appearance-based approaches [3], [9], [17], [20], [22] basically assume that gait features that are useful for recognition might exist many parts of body including legs, arms, back, neck and combination of two or more.

Most of the appearance-based approaches rely on silhouettes extracted by background subtraction. In order to extract silhouette of moving objects from background subtraction, static background should be presumed. Thus, the silhouette-based approach requires static perspective for images in usual cases. However, places such as barns and cowshed where static camera cannot be fetched due to the probable infection of disease from farm animals, hand-held cameras are only the alternative and perspective of the camera changing throughout the filming is inevitable. Since the camera keeps moving while filming, image registration process, which align every frame into uniform perspective, is nec-

[1] Osaka University
[2] FUJITSU LABORATORIES LTD
[a)] son@am.sanken.osaka-u.ac.jp
[b)] makihara@am.sanken.osaka-u.ac.jp
[c)] yagi@am.sanken.osaka-u.ac.jp
[d)] shio@jp.fujitsu.com

essary.

As the preprocessing step, this study focuses on the approach to aligning frames that are excerpted from a sequence of video. Section 2 introduces general methods of image registration and Section 3 briefly explains constraints of the proposed method. In section 4, the proposed method would be explained. Experimental results are shown in Section 5. In section 6, we will discuss its limitation and future work.

2. Related Work

Image registration refers to the process of aligning two images captured in different camera poses and positions. Reference image acts as a standard to which other images are aligned. Largely, there are two ways for image registration, area-based method and feature-based method [23]. Area-based method focuses on finding a transformation model that the best expresses correspondence between two images. A conventional criteria for area-based method is normalized Cross Correlation, which measures correspondence between reference image and input image.

An advantage of area-based approach is its accuracy. Since it compares all the pixels to find the optimal correspondence, the search process is thorough. As disadvantages, calculation time elongates due to thorough search and the method might fail when illumination changes drastically.

On the other hands, feature-based approach makes the best of characteristic structures. Features can be considered as a compact descriptor for spatial information. Further, it would be more robust against changes in illumination and computationally efficient than area-based approach. Feature-based approach consists of four steps. First feature points are detected. This step poses much importance as overall performance largely depends on robustness and accuracy of feature descriptor. Then, feature points on the input and reference images are matched. Next, transformation model is estimated based on matching result. Finally, interpolation techniques compensate for the input image that may have no information corresponding to reference image after transformation. As a common drawback for feature-based approach is that the reciprocal feature points are sometimes hard to retrieve when structural distortions exist in images and shows inferior approach results to area-based approach.

Handy access to OpenCV function of invariant feature point detection that are robust to changes in rotation and scale [4] as well as function of feature points approach [11] enables quick implementation of the method. However, the sequence of images may entail motion blurs and needs to be de-interlaced, by which corresponding feature points would be hardly extracted between frames. This limitation hampers to apply feature-based approach in frame alignment for a video sequence.

3. Problem Setting

The proposed approach assumes that all frames in a image sequence contain horizontal and vertical structures and other structures are negligible compared to the orthogonal structures. Scenes of cowshed, places where fence is constructed for protection purpose, may satisfy the requirements mentioned above. Using horizontal poles and vertical poles of fence, one can estimate the camera position including translation and rotation with respect to the world coordinate system. An example of a cowshed is presented in Fig. 1. Note that a sequence of images captured by a hand-held camera includes motion blur thus, shows lower sharpness in resolution than a sequence of images that captured by a static camera.

Fig. 1 Example of cowshed where vertical and horizontal poles of fence serve as orthogonal structures in real world. These poles are to be exploited in estimating vanishing point.

4. Proposed Method

The process of image registration can be divided into two parts. As the first part, we should know the position of the camera with respect to the scene. When the prerequisite described in the previous section is satisfied, vanishing points estimated from the vertical and horizontal structures would provide cue for estimating the position of the camera. The estimated position of the camera helps re-project the image onto another plane which are equal to the image captured from a different perspective. In our approach, we chose a view point of the camera so as to be orthogonal to a plane spanned by the fence poles, since cattle walk along the fence and the most characteristic parts of the gait, legs for the farm animals, could be observable.

As the second part, other frames except the initial frame is aligned to the previous frame rather than converting their perspective with vanishing points since vanishing point detection fails to show satisfiable performance for frames that include motion blur. As image sequences may contain motion blur, accurate estimation of vanishing points becomes unavailable due to lack of sharpness which leads to failure of edge detection. This shortcoming becomes more untenable when the orthogonal structures from which straight segments need to be extracted have similar color with surroundings as such in our cases.

4.1 Vanishing Point Detection

In the settings that meet the requirements addressed in Section 3, vanishing points can be detected where straight lines converge. Straight lines need to be as credible as it can inform us which edges are parts of horizontal and vertical structures in real world. For this reason, we chose to not rely on standard hough transformation (SHT) which run fast with simple algorithm but the results might entail irrelevant lines as it weighs only the number of pixel on a line without taking connectivity of edges into

account. To resolve this problem, we extract straight line segments from the scene and estimate the vanishing points based on the level of consensus that these segments generate. Many work has been done on extracting straight line segments from images [6], [19]. In order to estimate vanishing points, we voted each segments on Gaussian sphere proportional to its length on the assumption that the intersection of longer lines represents more likelihood to be vanishing point than intersection of shorter lines. Various work has been done surrounding vanishing point detection [2], [10], [15], [16].

4.2 Transformation of viewpoint

Upon vanishing points being known, the perspective of the image could be arbitrarily converted. Projection matrix, $P \in \mathbb{R}^{3\times4}$, maps a point in world coordinate system $\tilde{\boldsymbol{M}} = [X, Y, Z, 1]^T$ to a point on the image plane $\tilde{\boldsymbol{m}} = [u, v, 1]^T$ as follows :

$$s\tilde{\boldsymbol{m}} = P\tilde{\boldsymbol{M}} \tag{1}$$

Moreover, the projection matrix P can be decomposed as follows:

$$P = K[R|t] \tag{2}$$

where $K \in \mathbb{R}^{3\times3}$ is intrinsic camera matrix, $R \in \mathbb{R}^{3\times3}$ is rotation matrix which becomes known from vanishing points and $t \in \mathbb{R}^{3\times1}$ is translation vector which is set to zero for the initial frame in our methods. Intrinsic camera matrix K can be described when skew is zero and focal length is invariable for x, y direction as follows :

$$K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where u_0, v_0 are x,y coordinate of the image center. Rotation matrix R can be formalized as follows :

$$R = \begin{bmatrix} v_x^h & v_x^v & v_x^d \\ v_y^h & v_y^v & v_y^d \\ v_z^h & v_z^v & v_z^d \end{bmatrix} \tag{4}$$

where v^h, v^v, v^d denote horizontal, vertical, depth vanishing points. Rotation matrix R satisfies following equation :

$$RR^T = R^TR = I \tag{5}$$

where I is identity matrix. Therefore, in order to estimate the position of the camera, more than two exact vanishing points should be detected so that all elements can be retrieved [14].

We can define world coordinate system to have projection vector perpendicular to the plane that vertical and horizontal poles of fence span as shown in Fig. 2. It is known that homography, $\boldsymbol{H} \in \mathbb{R}^{3\times3}$ can be derived as :

$$H = K(I + \frac{tn^T}{f})R^{-1}K{-}1 \tag{6}$$

by which,

$$\tilde{\boldsymbol{m}'} = H\tilde{\boldsymbol{m}} \tag{7}$$

where $\tilde{\boldsymbol{m}'}$ and $\tilde{\boldsymbol{m}}$ denote image coordinates for world coordinate system and the current camera coordinate system respectively.

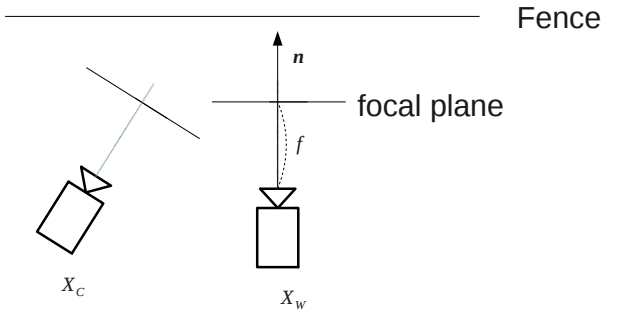


Fig. 2 Bird's eye view for filming situtation. $\boldsymbol{X_W}$ denotes world coordinate system and $\boldsymbol{X_C}$ denotes camera coordinate system for current frame. World coordinate system is defined based on the initial frame. Homography directly relates between the image coordinates for two cameras.

Problem of aligning frames then can be reduced to finding optimal variables for homography matrix. The variables to be modulated include rotation matrix and translation vector. Rotation vector r can replace rotation matrix R with Rodrigues conversion rule as follows :

$$R = I + \frac{\sin\|\mathbf{r}\|}{\|\mathbf{r}\|}[\mathbf{r}]_\times + \frac{(1-\cos\|\mathbf{r}\|)}{\|\mathbf{r}\|^2}[\mathbf{r}]_\times^2 \tag{8}$$

where $[\mathbf{r}]_x$ is the cross-product matrix for $\mathbf{r}$ which can be described as:

$$[\mathbf{r}]_\times = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} \tag{9}$$

4.3 Frame Alignment

Instead of feature-based registration method that would extract characteristic points in images and match them to get transformation function, we explored to apply area-based registration method in this study because feature point cannot be always successfully detected due to motion blur which remains edges and corners unclear.

Choice of cost function holds importance as it decides how stably final solution can be found. First of all, comparison of pixel values can have various forms. In our approach, the square sum of intensity for RGB channels for each pixel is adopted to make use of difference in color.

For convenience, define vector $\boldsymbol{d}(i, j)$ with respect to the point (i, j) as :

$$\boldsymbol{d}(i, j) = \boldsymbol{I}_{prev}(i, j) - \boldsymbol{I}_{curr}(i, j) \tag{10}$$

where $\boldsymbol{I}_{prev}(i, j) \in \mathbb{R}^3$ and $\boldsymbol{I}_{curr}(i, j) \in \mathbb{R}^3$ denote a vector of intensity values for RBG channels at position (i, j) in previously aligned frame and current frame to be aligned.

Certainly, only pixels inside the region where two image regions intersect should be compared. Comparing pixel values is nonsense when either of which has no information at all. We also want intersection region to be larger to prevent comparison of small identical region.

Additionally, we should only count for difference between pixels of stable background excluding pixels of moving objects and unstable pixels of background if necessary. On the assumption that the noise pattern follows additive white Gaussian noise (AWGN), pixels whose all RGB channels follow Gaussian dis-

tribution with the zero mean are counted as static background pixels. Let BG denote a set of pixels in static background. Then, a point on the image $(i, j) \in BG$, if all elements of $d(i, j)$ follow Gaussian distribution $N(0, \sigma^2)$ which can be formulated as :

$$BG = \{(i, j) | \forall k,\ d_k(i, j) \sim N(0, \sigma^2)\} \tag{11}$$

where $d_k(i, j)$ means the kth element of vector $\boldsymbol{d}(i, j)$.

Finally, we define a set of stable background region by setting the threshold as 3σ.

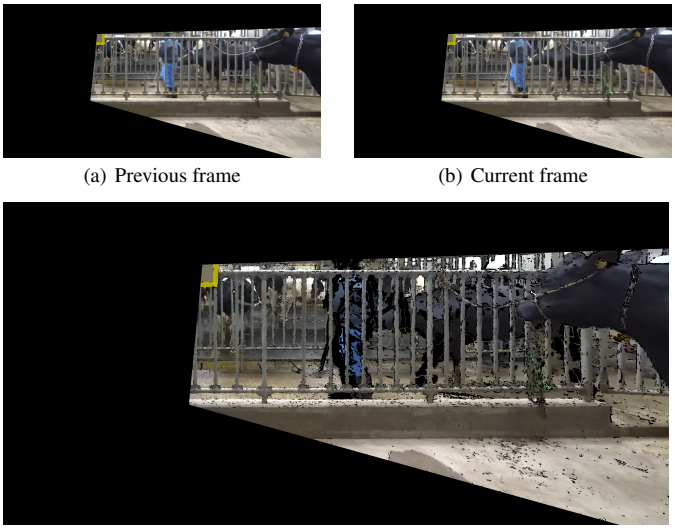
$$BG = \{(i, j) | \forall k,\ |d_k(i, j)| < 3\sigma\} \tag{12}$$

Considering that the ratio of static background would not drastically change frame to frame when temporal interval between frames is millisecond scale and assuming that previous frame is aligned favorably, the ratio of static background in previous alignment can be used as the ratio of static background in current alignment. Our experiments confirmed that this ratio showed no severe fluctuation.

Summing up all requirements above, we designed the cost function that can be thought of in the simplest form :

$$cost = \sum_{(i,j) \in BG} \frac{\|\boldsymbol{d}(i, j)\|_2^2}{|BG|} \tag{13}$$

where $|BG|$ means the area of the set BG. An example of the pixels that act as static background was depicted in Fig. 3.



(a) Previous frame (b) Current frame

(c) Static background

Fig. 3 (b) Pixels in current frame without moving objects are compared between previous frame and current frame. (c) Pixels that are considered to be static background are color-coded. Moving parts, such as legs of the man and forefoot of the cow, were excluded from comparison.

In optimizing the cost, brute force search would take enormous amount of calculation cost. In order to efficiently find optimal equilibrium, we addressed gradient descent method. Among methods of gradient descent, conjugate gradient [8], [12] was adopted since it generally shows faster convergence than steepest gradient descent that search the direction that would lessen the cost. It is important to note that cost function is non-linear and not continuous. Non-linear property of the cost function infers numerous local minimums that potentially attract the search to a halt and discontinuity implies sensitive dependency on step size. We addressed pyramid images in matching to prevent the optimization from ending up with one of local minimums. Six variables to adjust are three elements of rotation vector and three elements of translation vector.

For accurate calculation of gradient and step size, it is desirable to unify the scale between rotation and translation. For this reason, we addressed the amount of movement for optical center. For instance, when focal length is f and one element of rotation vector changes as much as Δr, the optical center moves as much as $f\Delta r$ which is equivalent to movement of an element in translation vector as much as $f\Delta r$. Even though rotational effect is greater than translational effect for rotation moves points proportional to the distance from the optical center, this difference becomes negligible when Δx is remained small. Then unified scale can be introduced as follows :

$$\mathbf{x} = (fr_1, fr_2, fr_3, t_1, t_2, t_3)^T \tag{14}$$

where f denotes focal length. Increment for the kth element can be also introduced :

$$\Delta\mathbf{x}_k = \lambda(\delta_{1k}, \delta_{2k}, \delta_{3k}, \delta_{4k}, \delta_{5k}, \delta_{6k})^T \tag{15}$$

where δ_{ij} denotes Kronecker delta and λ in the Eq. (15) regulates the scale of increment. Partial derivative with respect to the kth element of $\mathbf{x}$ can be described as followings :

$$\frac{\partial s(\mathbf{x})}{\partial x_k} = \frac{s(\mathbf{x} + \Delta\mathbf{x}_k) - s(\mathbf{x} - \Delta\mathbf{x}_k)}{\|2\Delta\mathbf{x}_k\|_2} \tag{16}$$

where function $s(\mathbf{x})$ denotes cost function previously introduced. Then gradient of cost function can be described as :

$$\nabla s = (\frac{\partial s(\mathbf{x})}{\partial x_1}, \frac{\partial s(\mathbf{x})}{\partial x_2}, \frac{\partial s(\mathbf{x})}{\partial x_3}, \frac{\partial s(\mathbf{x})}{\partial x_4}, \frac{\partial s(\mathbf{x})}{\partial x_5}, \frac{\partial s(\mathbf{x})}{\partial x_6})^T \tag{17}$$

Along the direction of gradient, line search method seeks the minimum point based on brent method in our method. Experimentally, λ more than 1 failed to search correct gradient direction which results in poor alignment while subpixel λ under 0.5 showed stable alignment result.

5. Experimental Result

5.1 Set Up

To see overall performance for our method, we experimentally registered one sequence of video that captured cattle followed by trainers. Frame-rate for the video sequence is 29 fps and original image resolution is 1920×1080 pixels. Examples of input images for our method are shown in Fig. 4.

In frame alignment, we employed 84 images in which cattle pass through the aisle under the control of the trainer. We set a value for step size λ to 0.125.

5.2 Result

Results of frame alignment are provided in Fig. 5. Even though the alignment results showed satisfiable correspondence between neighborhood frames, we could find overall background

Fig. 4 An example of a sequence of input images. A cow walk through the aisle under the guidance of a trainer from right side of the camera to left side.

are shifted throughout the sequence. This phenomenon is may also be caused by mapping non-planar 3D world to 2D plane where the camera movement entails translation. Possible solution to completely solve this problem shall be to construct 3D structure of the scene. However, since constructing 3D structure takes redundant information that would be needed for gait recognition, we aimed to align each frame maintaining position of front fence. In order to see the effect of inconsistent background over time, we constructed a panoramic background image after all frames registered as shown in Fig. 6. The panoramic background image is constructed with the median value for each pixel. To note that the initial frame starts from the right side of the image, blurred poles of fence in left side would indicate the effect of sliding background.

The degree of shift by the accumulation errors for the poles of fence can be witnessed in $x-t$ cross section for the sequence as shown in Fig. 7. We could confirm that about 20 pixels are shifted for vertical poles of fence which should nearly zero ideally.

6. Discussion

Current method needs improvement in terms of both performance and complexity. An easy way to achieve both aspects is to first align each frame to previous frame based on feature points and then use of the rotation and translation of homography matrix as a initial value for pixel-based alignment. This change would guarantee improved alignment result than only with feature point based method and drastically reduce the calculation time as good initial point has been available.

Another issue that could be improved may include selecting standard objects to align and aligning to these structures. For instance, poles of vertical fence failed to show uniform straightness in $x-t$ cross section in the proposed method as shown in Fig. 7. We expect these poles could be artificially aligned to stay static in the cross section image. The first step to achieve that goal, we designed dynamic programming that would find the trajectory of same pixel value that minimizes the cost within a window whose size is heuristically set. The concept is drawn in Fig. 8.

The cost function should penalize the difference of intensity between points as well as the difference of intensity between initial point and each point and constrain movements. The cost function is as follows :

$$\begin{aligned} cost(x_1, y_1) = \sum_{t=t_0}^{T} & (I(x_{t+1}, y_{t+1}, t+1) - I(x_t, y_t, t))^2 \\ & (I(x_{t+1}, y_{t+1}, t+1) - I(x_1, y_1, t_0))^2 \\ & + \lambda((x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2) \end{aligned} \tag{18}$$

One Example of retrieved trajectory in a vertical pole of fence are shown in Fig. 9.

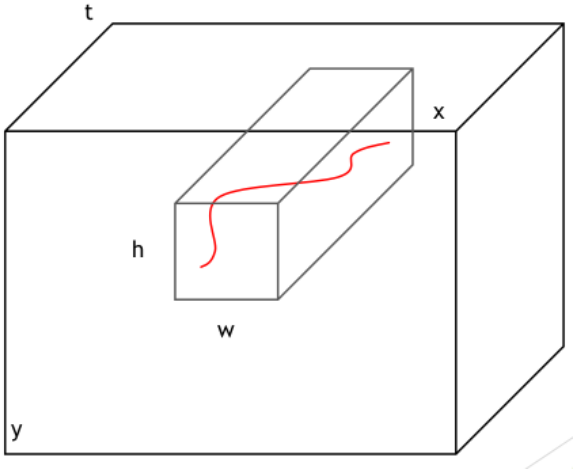


Fig. 8 A concept diagram for finding the trajectory within temporal-spatial volume.Proper window size of h and w is heuristically set and trajectory that minimizes cost inside the window would be sought with dynamic programming.

Fig. 9 Trajectory of a point in a vertical pole of fence on $x-t$ cross section. Vertical axis denotes x axis and horizontal axis denotes temporal axis. The trajectory starts with the center on the left side and ends at right side. It is confirmed that positional information of slightly shifted fence throughout the sequence is captured in the trajectory.

We look forward to realigning frames with information retrieved from trajectory on various positions.

7. Conclusion

We have discussed a method that registers images taken from a hand-held camera under special situation where orthogonal structures in real world, such as vertical and horizontal poles of fence in barns and cowshed, are dominant features in images. Appearance-based gait recognition techniques, especially for those that need silhouette for objects of interest by background subtraction, can be applied with the help of sound result of image registration. The result has shown further work to be done as poles of fence slightly move throughout the sequence due to accumulation errors. We expect our future work that realign based on trajectory of poles of fence would resolve this problem.

Acknowledgments The cattle in the video sequences used in this paper were owned by National Agriculture and Food Research Organization, Hokkaido Agricultural Research Center, Dairy Production Research Division.

References

[1] Ariyanto, G. and Nixon, M.: Marionette mass-spring model for 3D gait biometrics, *Proc. of the 5th IAPR International Conference on Biometrics*, pp. 354–359 (2012).

(a) Initial frame (b) Result for the initial frame

(c) 30th frame (d) Result for 30th frame

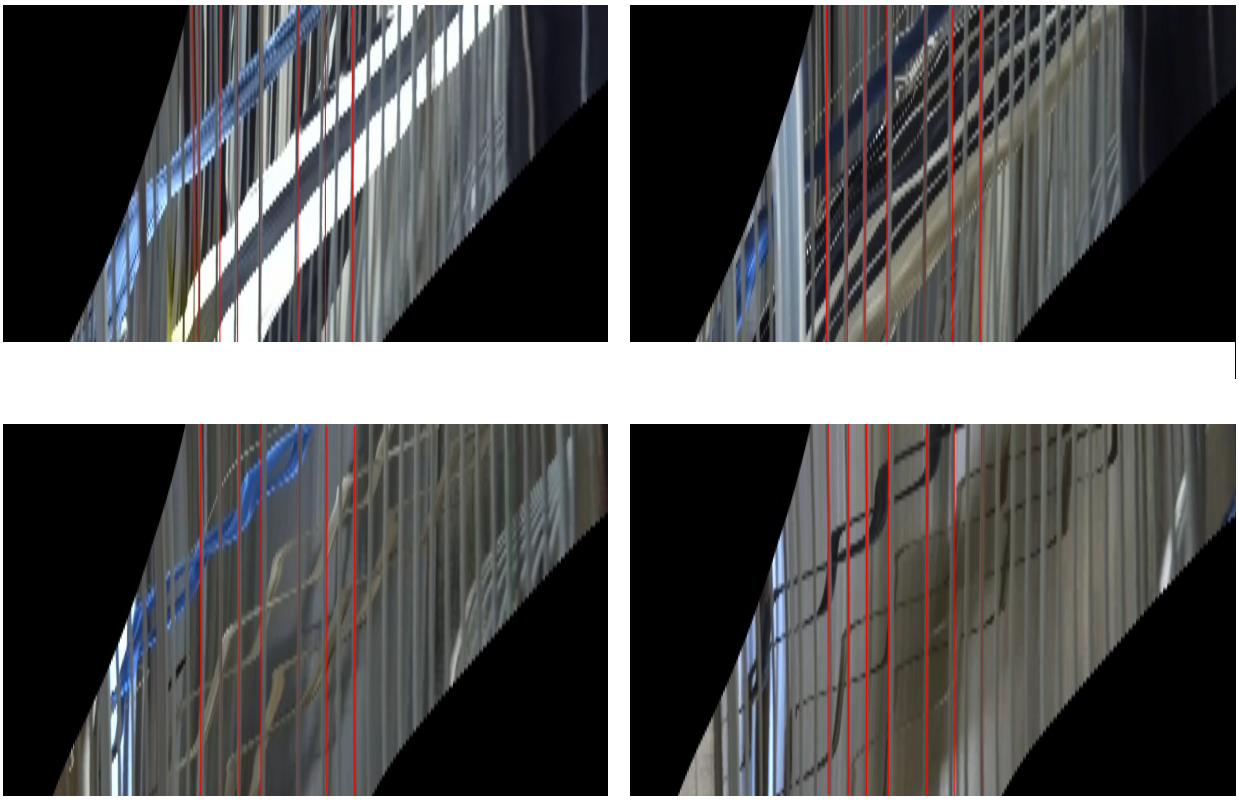
(e) 60th frame (f) Result for 60th frame

(g) Final frame (h) Result for final frame

Fig. 5 (a) Horizontal and vertical poles of fence are dominant structures in the scene that shape orthogonal structure in real world from which vanishing point is estimated. Little motion blur is desirable, as accurate detection for straight line segment largely relies on sharpness of images. (b) An image whose perspective is converted so as to rectify orthogonal structures in real world. Ideally, vertical poles should become vertically upright and horizontal poles should horizontally flat after transformation. (c),(d),(e),(f),(g),(h) indicate that results for converting perspective entail error as poles of fence fail to demonstrate appropriate rectification. The misalignment exacerbates as the frame index increases which suggests accumulative error in alignment.

[2] Barnard, S. T.: Interpreting perspective images, *Artificial intelligence*, Vol. 21, No. 4, pp. 435–462 (1983).
[3] Bashir, K., Xiang, T. and Gong, S.: Gait recognition without subject cooperation, *Pattern Recognition Letters*, Vol. 31, No. 13, pp. 2052–2060 (2010).
[4] Bay, H., Tuytelaars, T. and Van Gool, L.: Surf: Speeded up robust features, *Computer Vision–ECCV 2006*, Springer, pp. 404–417 (2006).
[5] Bobick, A. and Johnson, A.: Gait Recognition using Static Activity-specific Parameters, *Proc. of the 14th IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 423–430 (2001).
[6] Burns, J. B., Hanson, A. R. and Riseman, E. M.: Extracting straight lines, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, No. 4, pp. 425–455 (1986).
[7] Cunado, D., Nixon, M. and Carter, J.: Automatic Extraction and Description of Human Gait Models for Recognition Purposes, *Computer Vision and Image Understanding*, Vol. 90, No. 1, pp. 1–41 (2003).
[8] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316– 322 (2006).
[9] Lam, T. H. W., Cheung, K. H. and Liu, J. N. K.: Gait flow image: A silhouette-based gait representation for human identification, *Pattern Recognition*, Vol. 44, pp. 973–987 (online), DOI: http://dx.doi.org/10.1016/j.patcog.2010.10.011 (2011).
[10] Magee, M. J. and Aggarwal, J. K.: Determining vanishing points from

Fig. 6 Panoramic background image. After aligning all frames, value for each pixel is set to median pixel bje
val



(c) (d)

Fig. 7 (a),(b),(c),(d) $x-t$ cross secition cut by red lines from top to bottom in Fig. 6. Vertical axis represents temporal axis and horizontal axis represents spatial axis in x direction. Red lines in (a),(b),(c),(d) are vertically straight. If vertical poles of fence are static throughout the time, gray pixel of poles would draw straight lines parallel to the red lines. It is noticeable that vertical poles of fence are not completely static and subtly being shifted throughout the sequence. This disposition can be confirmed with fence moving in the video sequence of transformed images. Observably, the tendency of shift is not uniform for every pole. For instance, (a) shows more shift than others even though pixels reside in same pole. This result suggests that simple translational revision would not resolve the 'moving fence' problem.

perspective images, *Computer Vision, Graphics, and Image Processing*, Vol. 26, No. 2, pp. 256–267 (1984).

[11] Muja, M. and Lowe, D. G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration., *VISAPP (1)*, pp. 331–340 (2009).

[12] Nixon, M. S., Tan, T. N. and Chellappa, R.: *Human Identification Based on Gait*, Int. Series on Biometrics, Springer-Verlag (2005).

[13] Okada, K., Kobayashi, H., Hanada, N., Hiranuma, H., Hayashi, N., Arashi, Y., Chida, H., Deguchi, Y. and Sato, S.: Detection of hoof diseases in cattle using a triaxial accelerometer, *3 åŁå*, Vol. 2, No. 4, pp. 183–188 (2011).

[14] Orghidan, R., Salvi, J., Gordan, M. and Orza, B.: Camera calibration using two or three vanishing points, *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, IEEE, pp. 123–130 (2012).

[15] Quan, L. and Mohr, R.: Determining perspective structures using hierarchical Hough transform, *Pattern Recognition Letters*, Vol. 9, No. 4, pp. 279–286 (1989).

[16] Rother, C.: A new approach to vanishing point detection in architectural environments, *Image and Vision Computing*, Vol. 20, No. 9, pp. 647–655 (2002).

[17] Sarkar, S., Phillips, J., Liu, Z., Vega, I., ther, P. G. and Bowyer, K.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 27, No. 2, pp. 162–177 (2005).

[18] Sugaya, Y. and Kanatani, K.: Extracting moving objects from a moving camera video sequence, *Proceedings of the 10th Symposium on Sensing via Imaging Information*, pp. 279–284 (2004).

[19] Von Gioi, R. G., Jakubowicz, J., Morel, J.-M. and Randall, G.: On straight line segment detection, *Journal of Mathematical Imaging and Vision*, Vol. 32, No. 3, pp. 313–347 (2008).

[20] Wang, C., Zhang, J., Wang, L., Pu, J. and Yuan, X.: Human Identification Using Temporal Information Preserving Gait Template, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 11, pp. 2164 –2176 (online), DOI: 10.1109/TPAMI.2011.260 (2012).

[21] Wang, L., Ning, H., Tan, T. and Hu, W.: Fusion of Static and Dynamic Body Biometrics for Gait Recognition, *Proc. of the 9th International Conference on Computer Vision*, Vol. 2, pp. 1449–1454 (2003).

[22] Wang, L., Tan, T., Ning, H. and Hu, W.: Silhouette analysis-based gait recognition for human identification, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 25, No. 12, pp. 1505–1518 (online), DOI: 10.1109/TPAMI.2003.1251144 (2003).

[23] Zitova, B. and Flusser, J.: Image registration methods: a survey, *Image and vision computing*, Vol. 21, No. 11, pp. 977–1000 (2003).