PhD Forum: Correlation Coefficient based Template Matching for Indoor People Tracking

Junzhi Guan*, Peter Van Hese*, Jorge Oswaldo Niño-Castañeda*, Sebastian Gruenwedel*,
Dirk Van Haerenborgh†, Dimitri Van Cauwelaert*†, Peter Veelaert*† and Wilfried Philips*
*Image Processing and Interpretation, TELIN, Ghent University/IBBT, Belgium
†Vision Systems, University College Ghent, Belgium
Email: junzhi.guan@telin.ugent.be

***Abstract*—One of the most popular methods to extract information from an image sequence is template matching. The principle of template matching is tracking a certain feature or target over time based on the comparison of the content of each frame with a simple template. In this article, we propose an correlation coefficient based template matching which is invariant to linear intensity distortions to do correction or verification of our existing indoor people tracking system [1].**

I. Introduction

Template matching appears in a number of applications in the field of signal processing, image processing, and computer vision [2]. The basic principle is that the position of a given pattern is determined by a pixel-wise comparison of the image with a given template that contains the desired pattern [3]. For this, the template is shifted u discrete steps in the x direction and v steps in the y direction of the image, and then the comparison is caculated over the template area for each position (u, v). To caculate this comparison, template matching normally uses simple similarity measures such as the sum of absolutely differences (SAD) or the sum of squared differences (SSD). However, these measures are not invariant to brightness and contrast variations, which often occur in indoor people tracking. A more robust similarity measure is a correlation coefficient, which is invariant to linear intensity distortions [2]. In this paper, we use correlation-coefficient-based template matching(CCBTM) to correct or verify our existing indoor people tracking system.

In section II, the system overview is given. The application of CCBTM is described in section III. Results and discussion are given in section IV. Finally, conclusion is given in section V.

II. System Overview

The existing system is constructed from multiple calibrated smart cameras and one server which estimates ground plane positions of people [1]. Each camera performs the task of tracking people in the image plane based on the foreground mask and feedback from the server. Only the bounding boxes (position of people in image plane) are sent to the server. The server makes use of this information to estimate the position of people in real-world coordinates using a Bayesian filter approach. Ground plane positions together with predefined person's width and height are used to construct a person's cuboid, which is back-projected to each camera view as image bounding boxes. Here we use CCBTM to correct or verify the 2-D tracking hypotheses in each camera.

III. Method

For previous frame, the tracking system gives the information about the bounding box as follows:

$$bbox_{n-1} = [x_{n-1}, y_{n-1}, w_{n-1}, h_{n-1}] \tag{1}$$

Then the template A is extracted from the previous frame based on the $bbox_{n-1}$ using the following equation:

$$A = I_{n-1}(x_{n-1} : x_{n-1} + w_{n-1}, y_{n-1} : y_{n-1} + h_{n-1}) \tag{2}$$

In order to search the person in the current frame n, we define:

$$x_s = x_{n-1} \pm d_x, y_s = y_{n-1} \pm d_y, \tag{3}$$

in which, d_x, d_y represent the search ranges in the x, y direction within image coordinate. Then a set of sub image can be extracted from the current frame using the following equation:

$$B_{x_s,y_s} = I_n(x_s : x_s + w_{n-1}, y_s : y_s + h_{n-1}) \tag{4}$$

Then the correlation coefficient between the template A and sub image B_{x_s,y_s} is computed based on the following equation:

$$\rho_{x_s,y_s} = \frac{\sum\sum AB_{x_s,y_s}}{\sqrt{(\sum\sum A^2)(\sum\sum B^2_{x_s,y_s})}} \tag{5}$$

Then find the maximum value of ρ_{x_s,y_s}, and the corresponding $\hat{x_s}, \hat{y_s}$ using the following equation:

$$(\hat{x_s}, \hat{y_s}) = \arg\max_{(x_s,y_s)} \rho_{x_s,y_s} \tag{6}$$

In order to evaluate the acuracy of the current tracking system, we compare the result with the ground truth. For frame n, the ground truth gives the information about the bounding box as follows:

$$bbox^g_n = [x^g_n, y^g_n, w^g_n, h^g_n] \tag{7}$$

the tracking system gives the information about the bounding box as follows:

$$bbox_n = [x_n, y_n, w_n, h_n] \tag{8}$$

The error between ground truth and the tracking system is computed as follows:

$$e_x^T = x_n^g - x_n \tag{9}$$

$$e_y^T = y_n^g - y_n \tag{10}$$

$$e^T = \sqrt{(e_x^T)^2) + (e_y^T)^2} \tag{11}$$

and the error between ground truth and CCBTM is computed as follows:

$$e_x^{CCBTM} = x_n^g - \hat{x_s} \tag{12}$$

$$e_y^{CCBTM} = y_n^g - \hat{y_s} \tag{13}$$

$$e^{CCBTM} = \sqrt{(e_x^{CCBTM})^2) + (e_y^{CCBTM})^2} \tag{14}$$

IV. Results and Discussion

In order to evalute the performance of the tracking system and CCBTM, the video sequence which we conduct research on has illumination changes during the tracking. A typical frame of our system shown in Fig. 1. The red bounding box is the result of the tracking system, the dark blue one is the ground truth and the light blue one is the result of CCBTM.

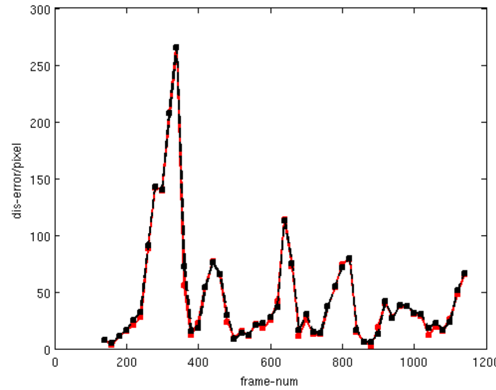
Fig. 1. Typical frame

The error is depicted in Fig. 2. It can be seen from Fig. 2(a) that the accuracy of the CCBTM and the tracking system is almost the same. It is clear that the error of CCBTM is smaller than the tracking system in frame 740 where illumination changes as shown in 2(b) which means that CCBTM is invariant to illumination changes, although the performance of both methods is not that good. The main reason is that the template for CCBTM is updated every frame based on the result of the current tracking system, so the performance of the CCBTM depends highly on the tracking system.

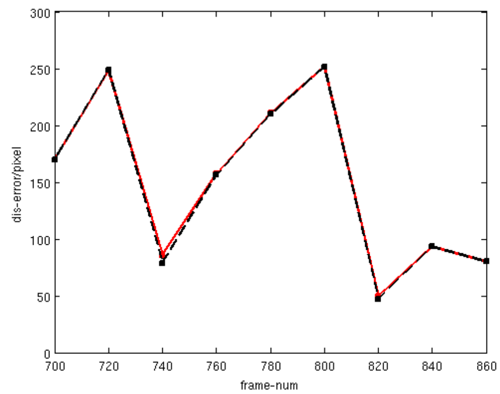
V. Conclusion

The performance of the tracking system and CCBTM is evaluated in our current indoor people tracking system. Results show that the CCBTM is more invariant to illumination changes and has the potential to correct or verify the 2-D tracking hypotheses in each camera.

Research [3], [4] illustrated that the template matching method can also be used to handle occlusion in indoors people tracking system. Future research will be conducted on implementing the CCBTM independently with the purpose of of improving the 2-D tracking hypotheses in each camera and handing occlusion based on CCBTM.



(a)



(b)

Fig. 2. Error of tracking system and CCBTM: (a) error in euclidean distance through the whole tracking period. (b) zoom in error in euclidean distance during a period which has illumination changes. Red curve - error between ground truth and tracking system e^T, black curve - error between ground truth and CCBTM e^{CCBTM}.

Acknowledgment

This work has been supported by the iCocoon project of the Flemish Interdisciplinary Institute for Broadband Technology (IBBT).

References

[1] S. Grünwedel, V. Jelaa, J. Nio-Castaeda, P. Van Hese, D. Van Cauwelaert, P. Veelaert, and W. Philips, "Decentralized tracking of humans using a camera network," in *Proceedings of SPIE, Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques*, vol. 8301. SPIE, 2012.

[2] A. Mahmood and S. Khan, "Correlation-coefficient-based fast template matching through partial elimination," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2099 –2108, april 2012.

[3] F. Jurie and M. Dhome, "Real time robust template matching," in *in British Machine Vision Conference 2002*, 2002, pp. 123–131.

[4] H. Nguyen, M. Worring, and R. van den Boomgaard, "Occlusion robust adaptive template tracking," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, 2001, pp. 678 –683 vol.1.