Object Reacquisition Using Invariant Appearance Model

Jinman Kang, Isaac Cohen and Gérard Medioni
IRIS, Computer Vision Group, University of Southern California
Los Angeles, CA 90089-0273
`{jinmanka|icohen|medioni}@iris.usc.edu`

Abstract

We present an approach for reacquisition of detected moving objects. We address the tracking problem by modeling the appearance of the moving region using stochastic models. The appearance of the object is described by multiple models representing spatial distributions of objects' colors and edges. This representation is invariant to 2D rigid and scale transformation. It provides a good description of the object being tracked, and produces an efficient blob similarity measure for tracking. Three different similarity measures are proposed, and compared to show the performance of each model. The proposed appearance model allows to track a large number of moving people with partial and total occlusions and permits to reacquire objects that have been previously tracked. We demonstrate the performance of the system on several real video surveillance sequences.

1. Introduction

Video surveillance is a popular application of video processing. Detection and tracking of moving objects and constitute the main problems and a large number of solutions have been proposed. However, partial or total occlusions of moving objects, as well as crowded environments remain still hard to process automatically. A partial solution consists of developing an object reacquisition method for persistently track moving objects after erroneous detections or occlusions.

A large number of papers have been published on video tracking. In [9], authors propose to use disparity and color information for individual person segmentation and tracking. In [5] and [6], segmented body parts or silhouettes are used for tracking multiple people. In [2], a multi-class statistical model of a person and of the background is used for tracking and gesture recognition. The limitations of these approaches are the lack of adaptability to temporary occlusions and invariance of the object description when the scene contains large number of moving objects.

To address this problem, each object has to be represented by a persistent object appearance model. An object appearance model is represented by a set of distinctive features such as color, shape, or texture. A typical shape-based appearance model for tracking relies on active con-

tour [10]. This method usually requires initializing the contour manually, and it only handles small non-rigid motion. Various color-based methods have been proposed in the literature to solve the tracking problem. Many of them use only one color histogram model per object. In [3], authors propose to use an appearance model using the temporal color feature, but the proposed approach is not invariant to arbitrary rigid motion. Multiple color models and their relative localization should be considered for an efficient use of the color in object tracking. In [1], a multiple color model approach was proposed for human detection, but it required a segmentation of detected blob into the head, torso, and legs.

Changes of appearance are expected while tracking moving people in the scene. Indeed, limbs motion will create localized shape variations and self-occlusions. Therefore, object appearance models have to be continuous in the sense that a small localized change of the object color and shape should create a small localized variation in its signature. The object description should also be invariant to 2D rigid transformation and scale change in order to accommodate for change of perspective.

We propose an appearance descriptor that is invariant to 2D rigid transformation and scale change over wide range of transformation within a large resolution. This model, defined by multiple polar distributions, provides a description of object's colors and shape properties. This 2D distribution model is used for measuring the similarity of tracked objects and for reacquiring objects after occlusions of long duration.

The rest of the paper is organized as follows. Section 2 introduces the object's appearance descriptor and its invariance to 2D scaling and rigid transformation. Section 3 presents several similarity measures for object reacquisition, and compares their performance. Section 4 presents several experimental results on real video sequences. Finally, in Section 5, we discuss our future work.

2. Invariant appearance descriptor

The color distribution model is obtained by mapping the blob into multiple polar representations. Several shape or color distribution models using a polar representation have been proposed [4][8][7]. In [4], the proposed approach is

focused on the object's shape description (edge) instead of their appearance (color), and it is only limited to representing local shape properties. In [8] the proposed model measures color distribution using a similar polar representation, but focuses on characterizing a global appearance signature of the object. The model is not 2D rotation-invariant and we propose here to use the shape description model proposed in [7] for guaranteeing invariance to 2D rigid transformation and scale change.

Given a detected moving blob, we set a reference circle C_R defined by the smallest circle containing the blob. This circle is uniformly sampled into a set of control points P_i. For each control point P_i a set of concentric circles of various radii are used for defining the bins of the appearance model. Inside each bin, a Gaussian color model is computed for modeling the color properties of the overlapping pixels of the detected blob. Therefore, for a given control point P_i we have a one-dimensional distribution $\gamma_i(P_i)$. The normalized combination of the distributions obtained from each control point P_i defines the appearance model of the detected blob: $\Lambda = \sum \gamma_i(P_i)$.
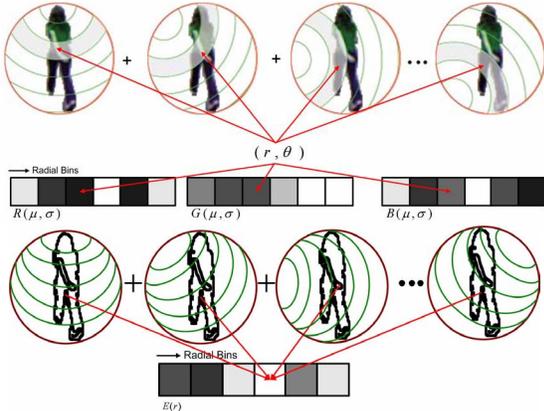
Figure 1. *Computation of the color and shape based appearance model of detected moving blobs*

An illustration of the definition of the appearance model is shown in Figure 1 where we sampled the reference circle with 8 control points. The defined model is translation invariant. Rotation invariance is also guaranteed, since a rotation of the blob in the 2D image is equivalent to a permutation of the control points. This is achieved by taking a larger number of control points along the reference circle. Finally, normalizing the reference circle to unit circle guarantees invariance to scale.

The shape-based description of the blob is obtained similarly by counting in each bin the number of edge pixels belonging to the moving blob. The 2D shape description is obtained by collecting and normalizing corresponding edge points for each bin as follows.

$$E(j) = \frac{\sum_i E_j(P_i)}{\max_j \left(\sum_i E_j(P_i) \right)} \quad (1)$$

where, $E(j)$ is edge distribution for j^{th} radial bin, and $E_j(P_i)$ is the number of edge points for the jth radial bin defined by the ith control point P_i.

3. Probabilistic appearance model

The proposed appearance descriptor is used for deriving an appearance probability model for defining an object reacquisition scheme. The appearance probability model is defined as a similarity measure among detected blobs in successive frames. Appearance of detected blobs is described thru a distribution function. We employ three types of similarity functions (e.g. cross-correlation, Bhattacharyya distance, and Kullback-Leibler distance) for measuring the similarity of the computed appearance models.

3.1. Correlation-based similarity

The first probability model is directly derived from the Gaussian approximation of the color properties in each bin using cross correlation. The appearance probability model is defined as follows:

$$P_{red} = \frac{N \sum_i \mu^i_{(r,t)} \mu^i_{(r,t+1)} - (\sum_i \mu^i_{(r,t)})(\sum_i \mu^i_{(r,t+1)})}{\sqrt{(N \sum_i (\mu^i_{(r,t)})^2 - (\sum_i \mu^i_{(r,t)})^2)(N \sum_i (\mu^i_{(r,t+1)})^2 - (\sum_i \mu^i_{(r,t+1)})^2)}} \quad (2)$$

where, N is the total number of bins (angular bins * radial bins), $\mu^i_{(r,t)}$ is the mean of the red component of the bin i, $P_{red}, P_{greed}, P_{blue}$ are the probability (likelihood estimation) of each color component.

$$P_{edge} = \frac{N \sum_j E_t(j) E_{t+1}(j) - (\sum_j E_t(j))(\sum_j E_{t+1}(j))}{\sqrt{(N \sum_j (E_t(j))^2 - (\sum_j E_t(j))^2)(N \sum_j (E_{t+1}(j))^2 - (\sum_j E_{t+1}(j))^2)}} \quad (3)$$

reflects the similarity of detected blobs shape. The color and shape appearance are combined as follow to define the appearance probability model:

$$P_{App\_Corr} = \frac{P_{red} + P_{green} + P_{blue} + P_{edge}}{4} \quad (4)$$

3.2. Bhattacharyya distance

The second probability model is obtained using the Bhattacharyya distance. Due to the different distribution models, Gaussian distribution for the color model and uniform distribution for the shape model, the similarity measurements are computed separately.

The similarity function of the color model is expressed in terms of the mean and variance of the Gaussian model (μ_t, σ_t). The obtained similarity is given by equation (5).

$$Dist_{B\_Color} = \sum_{N_{rgb}} \left\{ \frac{1}{8}(\mu_{t+1} - \mu_t)^T \cdot \left[\frac{\sigma_t + \sigma_{t+1}}{2} \right]^{-1} \cdot (\mu_{t+1} - \mu_t) + \frac{1}{2} \log \left[\frac{\frac{\sigma_t + \sigma_{t+1}}{2}}{\sqrt{|\sigma_t| \cdot |\sigma_{t+1}|}} \right] \right\} \quad (5)$$

where, μ_t and σ_t are respectively the mean and the variance of the considered color component at time t. N_{rgb} is total number of bins of the color component.

Shape similarity is derived using the Bhattacharya coefficient as follows:

$$Dist_{B\_Shape} = -\log \sum_{j} \sqrt{E(j)_t, E(j)_{t+1}} \qquad (6)$$

where, $E(r)_t$ is the shape probability distribution of each bin. The probability of the appearance model using Bhattacharyya distance is obtained by equation (7).

$$P_{App\_Bhatta} = \frac{1}{1 + \sqrt{(Dist_{B\_Color})^2 + (Dist_{B\_Shape})^2}} \qquad (7)$$

3.3. Kullback-Leibler Distance

The last probability model is defined using Kullback-Leibler distance. The calculation of the probability model follows the same convention as the probability model derived using the Bhattacharyya distance.

The similarity measurement of the color model is derived in terms of mean and variance of the Gaussians in each bin and is given by the likelihood ratio:

$$\Lambda_{KL\_Color} = \frac{1}{2 N_{rgb}} \sum_{N_{rgb}} \left\{ (\mu_t - \mu_{t+1})^2 \cdot \left(\frac{1}{\sigma_{t+1}^2} + \frac{1}{\sigma_t^2} \right) + \frac{\sigma_t^2}{\sigma_{t+1}^2} + \frac{\sigma_{t+1}^2}{\sigma_t^2} \right\} \qquad (8)$$

The similarity measurement of the shape model is obtained by equation (9).

$$Dist_{KL\_Shape} = \frac{1}{2} \sum_{j} (E(j)_t - E(j)_{t+1}) \log \frac{E(j)_t}{E(j)_{t+1}} \qquad (9)$$

The probability of the appearance model using the Kullback-Leibler distance is then defined by the following equation:

$$P_{App\_KL} = \frac{1}{\sqrt{(\Lambda_{KL\_Color})^2 + (Dist_{KL\_Shape})^2}} \qquad (10)$$

4. Experimental results

In the following we illustrate the contribution of color and edge information for inferring a good similarity measure.
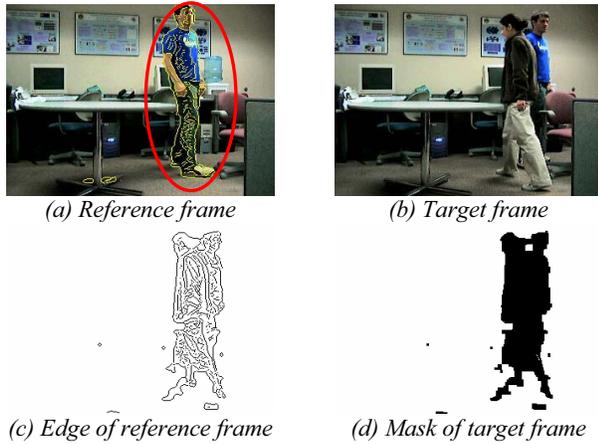

(a) Reference frame   *(b) Target frame*

(c) Edge of reference frame   *(d) Mask of target frame*

Figure 2. *Frames use for the experiment (red ellipse: target object)*

Figure 2 shows two frames considered for this experiment. In the following we will focus only on the moving object outlined by the red ellipse in Figure 2.a. We also highlighted the edge of the moving object in yellow. The proposed similarity measures are compared and evaluated using the target frame presented in Figure 2.b by searching for the optimal location in the image maximizing the various similarity measures. Figure 2.c and d present edge map and the masked target frame focusing on moving blob. The target frame considered contains all detected moving blobs. Note that the detected moving blobs in the target frame contain multiple persons merged into a single blob.

In Figure 3, we illustrate the contribution of the edge information to the appearance model. The proposed correlation-based similarity locates the reference moving object amongst the detected moving blobs as illustrated by Figure 3.a and b. Detection of the reference blob in the target frame using similarity functions based on Bhattacharyya distance and Kullback-Leibler distance provide more accurate and very similar results as expected (see Figure 3.c, d).
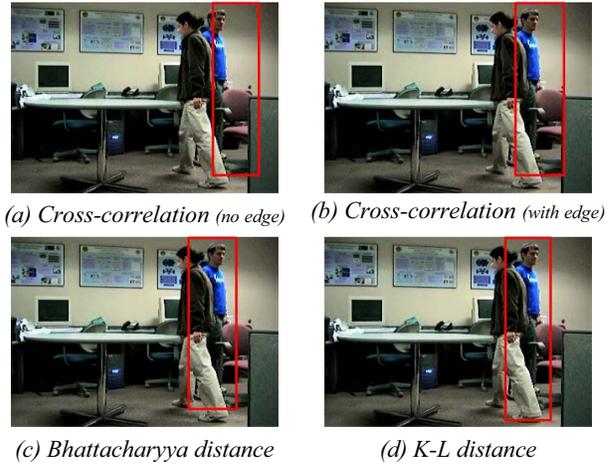

(a) Cross-correlation (no edge)   *(b) Cross-correlation (with edge)*

(c) Bhattacharyya distance   *(d) K-L distance*

Figure 3. *Comparison of the various similarity functions*

In the following we present some results obtained on real sequences for illustrating the object reacquisition using the proposed appearance model. In Figure 4, we illustrate object reacquisition after the tracked object was temporarily outside the FOV. The consistent labelling of the tracked objects is depicted by the continuous bounding box colors of the tracked objects. The proposed appearance model has also the capability of distinguishing objects. In Figure 5 we show a sequence of frames illustrating this. As one can observe in this sequence, the first tracked object (red box) hides behind the large pole and stays there for a while. Then, another person appears from behind the pole and resumes moving along the same direction as the first detected object. These two moving objects are distinguished by a very low appearance probability.

In Figure 6 we illustrate the continuous tracking of people while they interact and occlude each other. The figure depicts the tracking of two people during a fight. The people jump, run, fight… As one can observe from the figure the tracked objects are consistently reacquired by the proposed appearance descriptor.

5. Conclusion

We have presented a novel approach for tracking multiple

objects using an object's appearance model invariant to 2D rigid transformation and scaling. As depicted in the experimental results, moving objects are consistently reacquired frame by frame. For each moving object, the combination of color and shape represented by the multi-polar representation is presented in Section 2. The appearance model encodes both color and edge information of the detected blob. It is used to accurately measure the appearance's similarity regardless of the blobs' rigid motion. The probabilistic models derived from the invariant appearance descriptor are presented in Section 3. The appearance model proposed reacquires the moving objects in various situations such as: objects that leave the FOV temporarily. Currently, we evaluate the performance of proposed approach by observing the color and location of the bounding box of each object manually, and we can observe that moving objects are correctly reacquired and located.

Other issues remain to be addressed such as: the integration of motion information and modeling of object's kinematics. Also, the propagation of temporal correlation needs to be addressed for improving the performance of the proposed model.

Acknowledgements

References

[1] A. Elgammal and L. S. Davis, "Probabilistic Framework for Segmenting People Under Occlusion", *In Proc. of IEEE ICCV*, 2001.

[2] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-time tracking of the human body", *In IEEE Trans. on PAMI*, 1997.

[3] H. Roh, S. Kang and S. Lee, "Multiple People Tracking Using an Appearance Model Based on Temporal Color", *In Proc. of IEEE ICPR*, pp. 643-646, 2000.

[4] H. Zhang and J. Malik, "Learning a discriminative classifier using shape context distance", *In Proc. of IEEE. CVPR*, 2003.

[5] I. Haritaoglu, D. Harwood and L. S. Davis, "W4: Who? when? where? what? a real time system for detecting and tracking people", *In Proc. of International Conference on Face and Gesture Recognition*, 1998.

[6] I. Haritaoglu, D. Harwood and L. S. Davis, "Hydra: Multiple people detection and tracking using silhouettes", *In Proc. of 2nd IEEE Workshop on Visual Surveillance*, 1999.

[7] I. Cohen and H. Li, "Inference of Human Postures by Classification of 3D Human Body Shape", *IEEE IWAMFG*, 2003.

[8] J. Kang, I. Cohen and G. Medioni, "Continuous Tracking Within and Across Camera Streams", *IEEE CVPR*, 2003.

[9] T. Darrell, G. Gordon, M. Harville and J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection", *In Proc. of IEEE CVPR*, 1998.

[10] Y. Rui and Y. Chen, "Better Proposal Distributions: Object Tracking Using Unscented Particle Filter", *In Proc. of IEEE CVRP*, 2001.
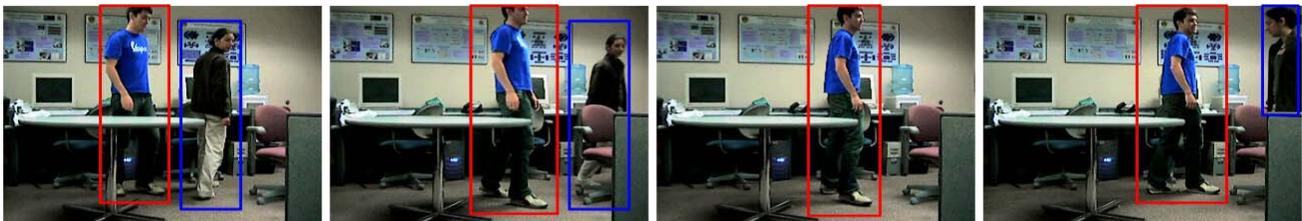
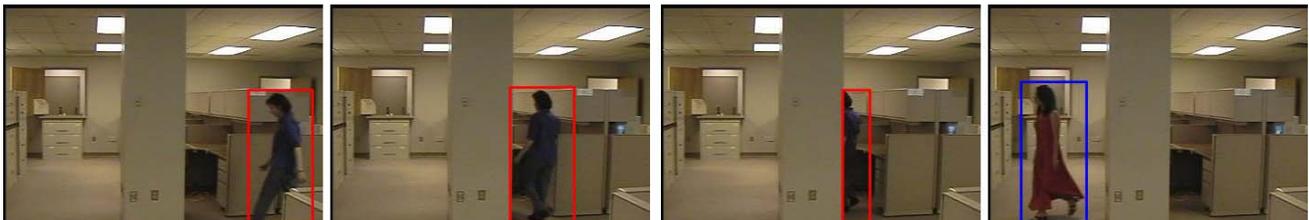**Figure 4.** *Continuous objects reacquisition when the tracked object is temporary out of FOV*

Figure 5. *Distinguishing tracked objects after occlusions. Object re-appearing behind the pole is not the same as the one previously tracked.*

Figure 6. *Object tracking with occlusions and large non-rigid deformations*