Original Research Paper

Fragment-based Visual Tracking with Multiple Representations

[1]Junqiu Wang and [2]Yasushi Yagi

[1]*AVIC Intelligent Measurement, Aviation Industry Cooperation China, Beijing, China*
[2]*The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan*

Abstract: We present a fragment-based tracking algorithm that considers appearance information characterized by a non-parametric distribution and spatial information described by a parametric representation. We segment an input object into several fragments based on the appearance similarity and spatial distribution. Spatial distribution and appearance are important for distinguishing different fragments. We employee such information for separating an object from its background: Appearance information is described by nonparametric representation such as kernels; spatial information is characterized by Gaussians with spatial distribution of fragments. We integrate appearance and spatial information for target localization in images. The overall motion is estimated by the mean shift algorithm. This motion can deviate from the true position in the overall motion estimation because of the mean-shift drifting. We refine the estimated position based on the foreground probabilities. The proposed tracker gives better target localization results and better foreground probability images. Our experimental results demonstrate that the integration of appearance and spatial information by combining parametric and non-parametric representation is effective for tracking targets in difficult sequences.

Keywords: Visual Tracking, Fragment-Based Tracking, Parametric and Non-Parametric Representation, Appearance and Spatial Information

Introduction

Visual tracking is still a hard problem after the intensive investigation over the years. Adaptive tracking (Collins *et al*., 2005; Han and Davis, 2004; Wang and Yagi, 2013) is effective for improving the tracking accuracy of a tracking algorithm by choosing good features that distinguish the object against its background. Unfortunately, model drifts bring difficulties for adaptive tracking (Jepson *et al*., 2003; Collins *et al*., 2005). To make the target model adaptive to the appearance variations, the tracker has to classify the pixels in the region into foreground and background. The classification process can mistakenly classified background pixels as foreground. Such pixels are incorporated into the object model. Thus the model updating makes the object representation drift from the true representation. The misclassification of the pixels can lead to the failure of the trackers for adaptive model updating. This problem can be partially solved by using an effective representation of targets and their background. The objective of this work is to improve the pixel classification by explicitly considering spatial distribution and appearance representation. We also investigate on how to use effective representation in the localization process for a visual tracker.

Different trackers try to find good target and background description approaches and target localization methods. Object characterization and position estimation are the two important issues that need to be addressed in developing an effective tracker. A good object representation can describe the essence of the object that is representative for the object and sufficiently discriminative for distinguishing the object from the background. Moreover, the characterization needs adaptive ability for handling object changes due to illumination variations or viewpoints changes.

Histograms are simple and effective nonparametric representations. Other nonparametric forms such as kernel density estimation are proposed for better performance. A target can be described by a kernel

Science Publications

density distribution or histograms. Histograms robust against variations in pose and shape have found wide applications in various tracking algorithms. The advantage is not gotten without any cost. Actually, we lost the spatial information that is useful for segmenting foreground and background.

We find inspirations for designing a good tracking algorithm in human intelligence. Multiple cues acts an important role in human perception. One set of features is not sufficient for describing an object. A few different features can represent different aspects of an object and its background. The complementary features are adopted for object tracking and recognition. Color features are widely used in different systems due to its simplicity. However, color features are not always effective, especially when the illumination conditions are changed. Other features such are shape and texture features are more stable in such situations (Toyama and Blake, 2001). Color cue can be misleading when the background has similar colors with the foreground target. In addition, color information changes when the illumination condition of the scene is not constant. Shape cue might have difficulties in the variations due to viewpoint changes. Shape and texture features can overcome the difficulties faced by color features (Birchfield, 1998; Wang and Yagi, 2013).

We present a mechanism for choosing good features. We make a feature pool by representing an object by color and shape texture cues. Color cue are characterized by histograms that are nonparametric representations. Shape and texture cues are represented by gradient orientation histograms. We compute an object and its background histograms different color spaces. Three color spaces are used for the color histogram computation: The HSV spaces, in the RGB space and the normalized r-g space. We select the most distinctive features by measuring the distinguishing ability of different features. Wang *et al.* (2011) propose another superpixel-based tracking method. Their tracker is relatively slow due to the application of the clustering method with many iterations.

We combine multiple cues for describing the targets and background. Appearance information can be strengthened by spatial information. To integrate different features, we segment an object into a few fragments. In the segmentation process, we consider the appearance similarity and spatial distribution simultaneously. The proposed tracking algorithm describes appearance information by nonparametric representation (kernels). Our algorithm characterizes object spatial information by Gaussians distributions with spatial information. We estimate the object motion by using the mean shift algorithm. We calculate foreground probability images based on pixel classification using the fragment-based characterization. We refine the motion parameters by searching in the foreground probability images.

Related Work

Spatial information has been considered by Birchfield and Rangarajan (2005) in their tracking algorithm. They represent spatial information using spatiogram that consists of many bins. Each bin in a spatiogram contains the position value of pixel weighted by the mean and covariance. Since the target is presented by one histogram, the tracking is not reliable when occlusion exist.

A joint spatial-color space was proposed by Wang *et al.* (2006) for characterizing the appearance of objects based on mixture of Gaussians. The tracking is initialized by an Expectation and Maximization (EM) step. The initialization is computationally expensive. To conquer the difficulty brought by illumination variations, they use the normalized color space r-g and intensity for characterizing the appearance of the object. Although r-g is robust against illumination changes, they are not stable and effective due to the insufficient discriminative abilities. In addition to the above problem, they do not explicitly represent the object as a part collection. In contrast, we use parametric and nonparametric representations in different situations. We select good features from a feature pool. The selected feature should be discriminative against the background.

Adam *et al.* (2006) proposed a fragment-based tracking algorithm in which the target is manually segmented into many fragments. The segmentation is fixed as grid separation. The tracker provides higher localization accuracy compared to the basic mean shift tracker (Comaniciu *et al.*, 2003). However, their approach has several drawbacks. First, it is difficult to track articulated objects because the integral histogram only can be computed in rectangular regions. Second, the model updating is also relatively difficult. Model updating is important for tracking a target in a dynamic scene (Jepson *et al.*, 2003; Wang and Yagi, 2008). We try to improve the separation results of the foreground target and its background, which is useful for better model updating.

Wang and Yagi (2008) select discriminative features to achieve better tracking performance. However, the feature selection is performed by considering the whole target and its background. This approach works well when the foreground target does not have complicated appearance distributions. Sudden motion is handled by auxiliary particles in Wang and Yagi (2009). The sudden motion detection is important for the initialization of the auxiliary particles. The appearance of the foreground target should be relatively distinctive against the background. Similar to Wang and Yagi (2009; Kim and Jeon, 2013) propose a spatio-temporal auxiliary particle filtering method for robust visual tracking. The target template is matched with the candidates by l1 minimization, which has been used in Mei *et al.* (2011). Wang and Yagi (2013) match multiple correspondences

based on super-pixel's appearance information. The spatial information is employed for localization. However, the spatial information is not considered in the pixel discrimination against the background.

Avidan (2007) presents classification for tracking method. To classify pixels into foreground and background, he presents an ensemble of simple weak classifiers for separating foreground and background. Each weak classifier is trained online from a specific frame and the ensemble is collected from a predefined range of recent frames. Our work also relates with his work. The classification in Avidan (2007 is performed by learning a classifier for the whole target. It does not consider the separation of the target. Our work combines the discriminative ability of the fragments segmented by a clustering algorithm that makes the representation more effective. Bai et al. (2013) extend the ensemble learning method by introducing a randomized ensemble. Their work is more efficient than the original ensemble. However, the learning process depends on the classifiers that can be drift during the tracking process. Yin and Collins (2006) segment object foreground and background into several regions. The discriminative ability of different features are evaluated and the good features are selected. The segmentation process is arbitrary, which brings errors that can lead to tracking failures.

The paper is organized as follows. In section 3, we introduce the segmentation process, we also describe how to represent a target and its background in this section. Section 4 discusses the feature selection method. The adaptive tracking method of the target is described in section 5. We evaluate the performance of the proposed method in section 6. We summarize this work in section 7.

Generation and Representation of Object Fragment

We perform tracking initialization detecting and automatic segmentation. The detection algorithm provides a bounding box for the target. The GrabCut (Rother et al., 2004). is used for the automatic segmentation.

Fragment Generation

The target to be tracked is segmented using the GrabCut. We need a collection of fragments for better object representation. We use the k-means algorithm for separating the object into multiple fragments. The k-means algorithm has been applied in different color spaces such as HSV, RGB, r-g. The separation performances of the k-means using different color spaces are compared accordingly. We found that color information is not sufficient for the separation. To improve the segmentation, we embed spatial information into the k-means. We achieve the best separation results by using the HS-XY space. The HS-XY representation consists of H and S channels in the HSV color space; and the coordinates of the pixels as spatial information. This results prove that spatial information is important. The fragment generation is illustrated in Fig. 1.
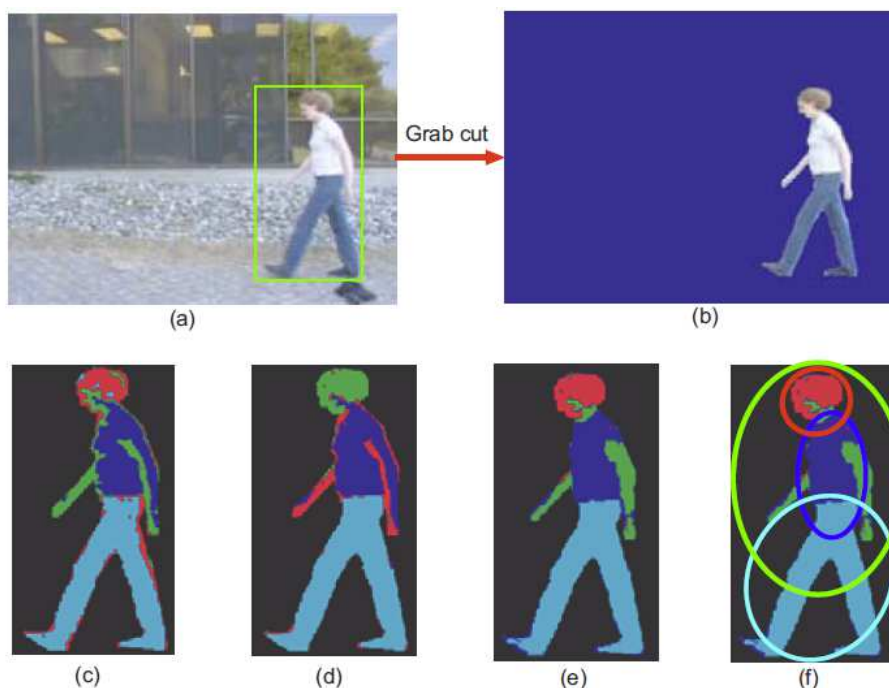

Fig. 1. (a) The input image of a woman. (b) Segmentation result using the GrabCut algorithm (Rother et al., 2004); Target separation using different spaces: (c) RGB; (d) HSV and (e) HS-XY. (f) Parametric representation of spatial information

The number of the fragments has to be defined before the decomposition. Different object might have different numbers of distinguish fragments. To handle this problem, we give a relatively large number setting in the beginning of the object fragment separation (e.g., 8 fragments). We run the k-means algorithm on an image and a few fragments are calculated. We evaluate the fragments based on their sizes. The small fragments are not discriminative. They are discarded in the separation process.

The results of object separation are a set of fragments. The pixels in each fragment have similar appearance and spatial adherence. We use G to denote a set of fragments sampled from the object region; and g indicate a single fragment in the fragment set. We describe the object using kernels that are nonparametric (Comaniciu *et al.*, 2003). Moreover, each fragment is characterized by both appearance and spatial information.

Object Representation

An object and each fragment is represented by their spatial and appearance information. We use nonparametric representation such as kernels (Comaniciu *et al.* (2003)). to describe object appearance information. Since spatial information can be better dealt with by parametric representations. We use spatial Gaussian of fragment g. Each spatial Gaussian is composed of its mean value μ_g and its covariance Σ_g. We show the spatial information of each fragment in Fig. 1.

Discriminative Selection for Good Features

Good features are helpful for tracking and recognition tasks. Here, we use feature selection aims find the best subset from the feature pool available. It has been proved useful in other tracking algorithms (Collins *et al.*, 2005). We can define criteria for our feature selection. Different criteria have been applied in feature selection, e.g., class separability measure (Nguyen and Smeulders, 2006), principal component analysis (Han and Davis, 2004), or variance ratio ranking. We found that variance ratio (Wang and Yagi, 2008) is a good measure for choosing discriminative features.

Probability Ratio Image Calculation

Variance ratio (Collins *et al.*, 2005) indicates the distinctiveness of an object with respect to its background. It is calculated based on probability ratios. Probability ratios project raw feature values nonlinearly into a new space. The pixels with certain appearance are found more on the object are transformed into positive values; and pixels with certain appearance that are frequently found in the background are transformed into negative values. We calculate log-likelihood ratios according to the histograms of an object and its background for a each feature. The values in different bins of a histogram represent the frequency of a certain feature. They are calculated by:

$$\zeta_F^{(bin)} = \frac{n_F^{(bin)}}{n_F} \tag{1}$$

And:

$$\zeta_B^{(bin)} = \frac{n_B^{(bin)}}{n_B} \tag{2}$$

Where:

n_F = The pixel number of the object region
n_B = The pixel number of the background region
$n_F^{(bin)}$ and $n_F^{(bin)}$ = The numbers of the foreground and background pixel number corresponding to a certain bin.

We calculate the log-likelihood ratio for a certain feature by Equation 3:

$$L = \max(-1, \min(1, \log \frac{\max(\zeta_F^{(bin)}, \delta_L)}{\max(\zeta_F^{(bin)}, \delta_L)})) \tag{3}$$

where, δ_L is a very tiny real number.

Discriminative Feature Selection

Variance ratio indicates the discriminative ability of a feature. A feature is discriminative when it is abundant in the target region but rare in the background region. We measure discriminative ability according to the log likelihood image calculated by histogram projection. We calculate the likelihood function of variance ratio by Equation 4:

$$v = \frac{\text{var}(B \cup F)}{\text{var}(F) + \text{var}(B)} = \frac{\text{var}(L; (p_F + p_B)/2)}{\text{var}(L; p_F) + \text{var}(L; p_B)} \tag{4}$$

We select 2 features from a feature set that consists of 6 color features and a shape-texture feature. The color features include R, G and B in RGB color space; H and S in HSV space; r and g in r-g space. The shape feature is described by the gradient orientation histogram of the target and its background. We calculate all the variance ratios for these features. The discriminative abilities of the features are ranked according to the variance ratios. The first two features with large variance ratio are considered as discriminative features.

Target Localization

We use the mean-shift algorithm to estimate the global motion of an object. The mean-shift searches for the mode in an efficient way. The searching steps are changed adaptively according to the distance to the mode. Large steps are taken when the searching is far from the

mode. The steps are adjusted to smaller ones when the searching is near to the mode. This strategy makes an efficient approach for mode seeking. The global motion estimation is estimated based on the good features with high discriminative ability from the feature pool.

Although the mean-shift algorithm is fast, the mode seeking result might not be the true position. For example, the mode seeking process can deviate from the true position when the probability distribution of the background is similar to the object. This happens especially when the pixel classification is not accurate. Since we have nonparametric representation for the appearance and parametric representation for the spatial information, we use these representations to give foreground or background labels to each pixels in the bounding box. The probabilities calculated based on these representations are integrated as the results for pixel classification. The position of the object is estimated based on the probabilistic image.

Probabilistic Formulation for Pixel Labeling

We make use of the following notation for the probabilistic formulation of pixel classification: Let R denote the ratio between the probability of the foreground and the background; S denote spatial information; A denote appearance information; and F and B denote foreground and background; x denote the position of the pixel. We use subscripts t to denote time. The tracking process can be represented by calculating the ratio between the foreground probability and the background probability given the spatial and appearance information Equation 5:

$$r_{\mathbf{x}} = \frac{p_{\mathbf{x}}(F \mid S, A)}{p_{\mathbf{x}}(B \mid S, A)} \approx \frac{p_{\mathbf{x}}(S, A \mid \theta_F) p_{\mathbf{x}}(F)}{p_{\mathbf{x}}(S, A \mid \theta_B) p_{\mathbf{x}}(B)} \qquad (5)$$

where, θ_F and θ_B denote the representation of the foreground target and the background, respectively.

We decompose the probability by Equation 6 and 7:

$$p_{\mathbf{x}}(S, A \mid \theta_F) = p_{\mathbf{x}}(A \mid S, \theta_F) p_{\mathbf{x}}(S \mid \theta_F) p_{\mathbf{x}}(\theta_F) \qquad (6)$$

And:

$$p_{\mathbf{x}}(S, A \mid \theta_F) = p_{\mathbf{x}}(A \mid S, \theta_B) p_{\mathbf{x}}(S \mid \theta_B) p_{\mathbf{x}}(\theta_B) \qquad (7)$$

We calculate the probability based on nonparametric appearance representation. The histograms are back-projected and the probability is achieved (Swain and Ballard, 1991; Wang and Yagi, 2008). We calculate the probability based the parameterized spatial information using a spatial Gaussian distribution:

$$p_{\mathbf{x}}(S \mid \theta_F, \theta_B) = \frac{\exp(-\frac{1}{2}(x - \mu^S)^{\mathrm{T}} \mid \Sigma^S \mid^{-1} (x - \mu^S))}{2\pi \mid \Sigma^S \mid^{\frac{1}{2}}} \qquad (8)$$

where, μ^S is the mean of the spatial distribution and Σ^S denotes the covariance.

We calculate a probabilistic image for each fragment. We combine the probabilistic images to obtain the foreground probability image.

Likelihood Ratios Integration

We select discriminative features for each fragment. The discriminative ability is selected according to the measure of certain areas characterized by the spatial Gaussians. Therefore, the contributions of the likelihood ratios obtained in different fragment are influenced by the discriminative abilities. The integration of different likelihood ratio images aims at including the discriminative abilities into the final results. The results are calculated by an interpolation Equation 9:

$$r_{\mathbf{x}}^g = \sum_g v_{\mathbf{x}}^g r^g(x) p_{\mathbf{x}}^g \qquad (9)$$

Where:
$v_{\mathbf{x}}^g$ = The variance ration score of fragment g
$p_{\mathbf{x}}^g$ = Calculated using the method in Equation 8

Target Location Refinement

The position estimated by the mean-shift algorithm can be refined based on the integrated log-likelihood images. The position refinement can provide a more accurate location estimation result of the object. In practice, we run another mean shift on the integrated likelihood ratio image (Bradski, 1998).

Experimental Results

We implemented the proposed approach. We run our algorithm on a few image sequences. We show the test results of three image sequences. These image sequences are captured by non-stationary cameras. The motion of the cameras brings difficulty to the tracking. The frames in the first two sequences have a size of 360×240 pixels. Likelihood images computation is not easy due to the low resolution of the frames. The articulated target structure, the dynamic background, the deformable property of the object are also the source of the difficulty. The images in the last sequence have a size of 640×480 pixels. This sequence has a dynamic background. We use this sequence to demonstrate the possibility for segmenting the target by using our tracking results.

We show the results of the image three sequences in Fig. 2a, b and Fig. 4, respectively. We show the likelihood ratio images obtained based on the global description in the middle columns of Fig. 2. The distinctive color of the upper body makes appearance information sufficient to calculate the likelihood ratio image. However, other

fragments of the person are not as distinctive as the upper body. They are not well addressed in the likelihood ratio image. In addition, the position indicated by the bounding box is not accurate. To solve this problem, the nonparametric and parametric representations of each fragment are used for likelihood ratio images estimation. Then, we integrate these likelihood ratios according to their discriminative abilities. We illustrate the integrated likelihood ratio image on the right in Fig. 2. The position of the bounding box is estimated according to the integrated likelihood ratio image. The position of bounding boxes are more accurate than the position by running a mean-shift algorithm in the middle column in Fig. 2.

We use several image sequences with ground truths to quantitatively evaluate the performance of the proposed algorithm. We label the images manually and get the ground truths of the sequences. The integrated the likelihood ratio images are converted into binary images by giving a threshold. We compare the results with the ground truth. We show the comparison results in Fig. 3. The integrated likelihood images give lower error rates than that of the direct back-projection of the single histogram representation in most cases. However, the integrated likelihood images are not as good as the direct back-projection in a few frames. We found that the bad performance is due to the drifting problem in the integration process. One feature can be discriminative in one frame. But it is not consistent in relatively long period. The target model is not updated accordingly. Therefore, this problem can be solved by considering the discriminative ability of a feature through a relatively stable period.

Fig. 2. The original image sequences and the foreground. The images in the first column are two frames in the input sequences. The second column shows the likelihood ratio images computed using the global description of the target. The original image sequences and the foreground. The images in the first column are two frames in the input sequences. The second column shows the likelihood ratio images computed using the global description of the target
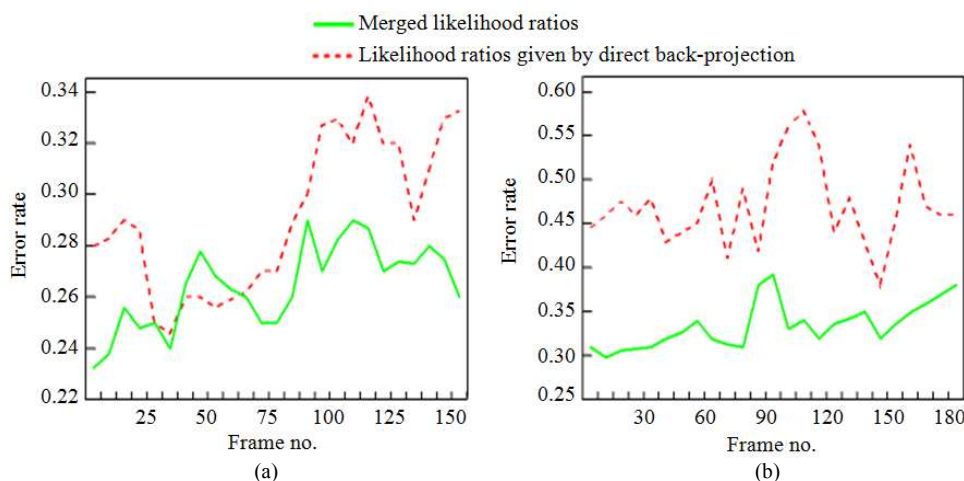

Fig. 3. The probabilistic images of different methods. The merged probabilistic image is better than the image gotten by direct projection
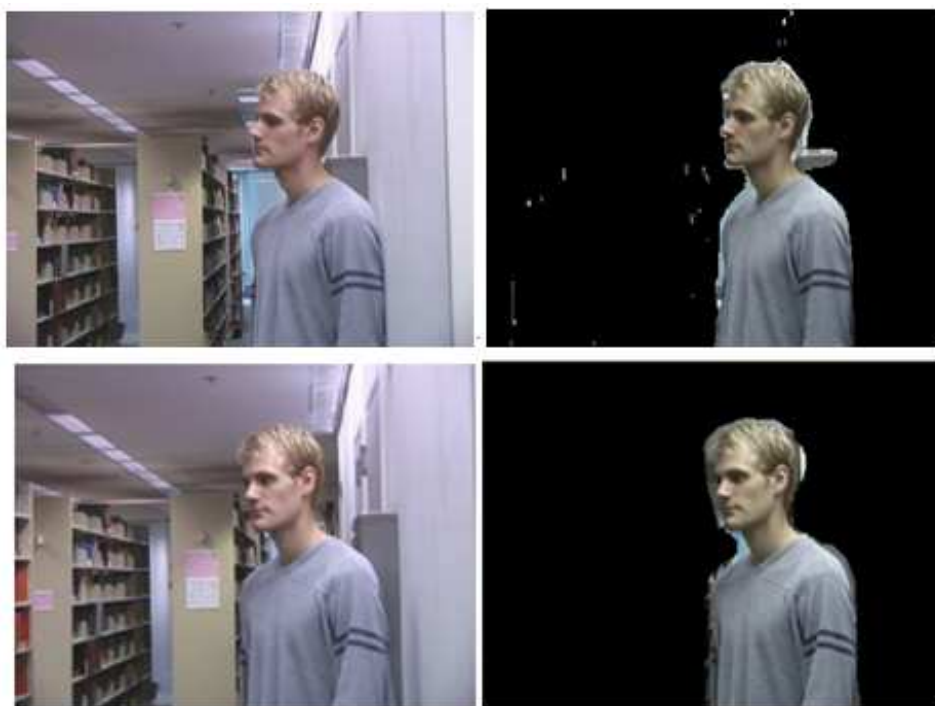
Fig. 4. The segmentation results of a video sequence. The images on the left are the input images; and the images on the right are the segmentation results by converting the likelihood image

In the experiments of the last sequence (Fig. 4), we employ appearance, spatial and motion information for the segmentation of the images. The segmentation results are gotten by converting the likelihood images by adding motion information. The results are useful for effectively updating the model.

Conclusion

We devised a fragment-based adaptive tracking algorithm. The target decomposition is effective for improving the performance of the tracking. The proposed method provides better target localization results. The foreground target is more distinctive in the likelihood images than the results calculated using the traditional methods. These results can be helpful for the segmentation in video sequences. It is also important in target model updating for avoiding the drifts.

Acknowledgement

Author's Contributions

Junqiu Wang: Designed this research, implemented the tracking system, perform the experiments, wrote the paper.

Yasushi Yagi: Co-designed the research.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Adam, A., E. Rivlin and I. Shimshoni, 2006. Robust fragments-based tracking using the integral histogram. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, pp: 798-805. DOI: 10.1109/CVPR.2006.256

Avidan, S., 2007. Ensemble tracking. IEEE Trans. Patt. Anal. Mach. Intell., 29: 261-271.
DOI: 10.1109/TPAMI.2007.35

Bai, Q., Z. Wu, S. Sclaroff, M. Betke and C. Monnier, 2013. Randomized ensemble tracking. Proceedings of the IEEE International Conference on Computer Vision (ICCV), Dec. 1-8, IEEE Xplore Press, Sydney, NSW., pp: 2040-2047. DOI: 10.1109/ICCV.2013.255

Birchfield, S., 1998. Elliptical head tracking using intensity gradients and color histograms. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 23-25, IEEE Xplore Press, Santa Barbara, CA., pp: 232-237. DOI: 10.1109/CVPR.1998.698614

Birchfield, S.T. and S. Rangarajan, 2005. Spatiograms versus histograms for region-based tracking. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 20-25, IEEE Xplore Press, pp: 1158-1163. DOI: 10.1109/CVPR.2005.330

Bradski, G.R., 1998. Real time face and object tracking as a component of a perceptual user interface. Proceedings of 4th IEEE Workshop Applications of Computer Vision, Oct. 19-21, IEEE Xplore Press, Princeton, NJ., pp: 214-219.
DOI: 10.1109/ACV.1998.732882

Collins, R.T., Y. Liu and M. Leordeanu, 2005. On-line selection of discriminative tracking features. IEEE Trans. Patt. Anal. Mach. Intell., 27: 1631-1643.
DOI: 10.1109/TPAMI.2005.205

Comaniciu, D., V. Ramesh and P. Meer, 2003. Kernel-based object tracking. IEEE Trans. Patt. Anal. Mach. Intell., 25: 564-577.
DOI: 10.1109/TPAMI.2003.1195991

Han, B. and L. Davis, 2004. Object tracking by adaptive feature extraction. Proceeding of the International Conference on Image Processing, Oct. 24-27, IEEE Xplore Press, pp: 1501-1504.
DOI: 10.1109/ICIP.2004.1421349

Jepson, A.D., D.J. Fleet and T.F. El-Maraghi, 2003. Robust online appearance models for visual tracking. IEEE Trans. Patt. Anal. Mach. Intell., 25: 1296-1311. DOI: 10.1109/TPAMI.2003.1233903

Kim, D.Y. and M. Jeon, 2013. Spatio-temporal auxiliary particle filtering with l1-norm-based appearance model learning for robust visual tracking. IEEE Trans. Image Process., 22: 511-522.
DOI: 10.1109/TIP.2012.2218824

Nguyen, H. and A. Smeulders, 2006. Robust tracking using foreground-background texture discrimination. Int. J. Comput. Vis., 69: 277-293.
DOI: 10.1007/s11263-006-7067-x

Mei, X., L. Ling, Y. Wu and E.P. Blasch, 2011. Minimum error bounded efficient l1 tracker with occlusion detection. IEEE Trans. Image Process., 22: 2661-2675. DOI: 10.1109/TIP.2013.2255301

Rother, C., V. Kolmogorov and A. Blake, 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. ACM Trans. Graph., 23: 309-314.
DOI: 10.1145/1015706.1015720

Swain, M. and D. Ballard, 1991. Color indexing. Int. J. Comput. Vis., 7: 11-32. DOI: 10.1007/BF00130487

Toyama, K. and A. Blake, 2001. Probabilistic tracking with exemplars in a metric space. Int. J. Comput. Vis., 48: 9-19. DOI: 10.1023/A:1014899027014

Wang, H., D. Suter and K. Schindler, 2006. Effective appearance model and similarity measure for particle filtering and visual tracking. Proceedings of the European Conference on Computer Vision, May 7-13, Springer Press, Austria, pp: 606-618.
DOI: 10.1007/11744078_47

Wang, J. and Y. Yagi, 2008. Integrating color and shape-texture features for adaptive real-time object tracking. IEEE Trans. Image Process., 17: 235-240.
DOI: 10.1109/TIP.2007.914150

Wang, J. and Y. Yagi, 2009. Adaptive mean-shift tracking with auxiliary particles. IEEE Trans. Syst. Man Cybernet., 39: 1578-1589.
DOI: 10.1109/TSMCB.2009.2021482

Wang, J. and Y. Yagi, 2013. Many-to-many superpixel matching for robust tracking. IEEE Trans. Syst. Man Cybernet., 44: 1237-1248.
DOI: 10.1109/TCYB.2013.2296511

Wang, S., H. Lu, F. Yang and M.H. Yang, 2011. Superpixel tracking. Proceedings of the International Conference on Computer Vision, Nov. 6-13, IEEE Xplore Press, Barcelona, pp: 1323-1330.
DOI: 10.1109/ICCV.2011.6126385

Yin, Z. and R. Collins, 2006. Spatial divide and conquer with motion cues for tracking through clutter. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, pp: 570-577. DOI: 10.1109/CVPR.2006.286