

Region of Interest Tracking In Video Sequences

T.Johnncy Rani
PG Student
Anna University Tirunelveli
Tirunelveli, India

S.Suja Priyadharsini
Lecturer
Anna University Tirunelveli
Tirunelveli, India

ABSTRACT

Tracking a region-of-interest (ROI) in a video is still a challenging task. Various high level applications rely on tracking. e.g. motion picture indexing, object recognition, video surveillance, audiovisual postproduction etc. Initially ROI is defined in a reference frame and the purpose is to determine the ROI in subsequent target frames in video sequences. The region was detected by determining the similarity measures between the reference and the target frames. Similarity measures between the frames are determined using two classical methods like sum of squared differences(SSD) and sum of absolute differences(SAD). This paper deals with the method of ROI tracking in video sequences by estimating the colour and geometric features between the frames and the similarity measures was determined using the Kullback- Leibler Divergence. The increase of description features improves the accuracy. Their combination leads to high dimensional PDFs. Tracking experiments were performed on several standard video sequences and its efficiency was proved.

Keywords

Region-of-interest(ROI), Similarity Measures, Colour and geometric features, Probability density function, Kullback-Leibler divergence.

1. INTRODUCTION

Region-of-interest tracking is a very depth level research area. ROI tracking is an important problem in the field of object based on video processing. When a physical object appears in several consecutive frames, it is necessary to identify its appearance in different frames for the purpose of processing.

The colour and geometric features between the reference and the target frames are the two aspects of similarity measures which is used for ROI tracking. The colour histogram of the object was made in the colour based probabilistic ROI tracking[5]. This model compute the colour similarity between the frames. Sum of squared differences(SSD) and sum of absolute differences(SAD) are the two classical methods which is used to find the motion compensation error(MSE) between two consecutive frames[2]. These methods have strict constraint, since the underlying residual is computed with deterministic pixel to pixel correspondence between the reference and target regions. Therefore it is not adapted to complex motions.

An alternative way for real time tracking of non-rigid objects seen from a moving camera based on the mean shift(MS) iterations and finds the most probable target position in the current frame[3]. The dissimilarity between the frames is expressed by a metric derived from Bhattacharyya coefficient. Some of the object tracking approaches combine two tracking

methods by overcoming their respective disadvantages. The rapid model change in SSD tracker is overcome by the MS tracker module, while the inability of MS tracker to handle large displacement is circumvented by the SSD module[7]. Different features were integrated into the ROI PDF model using Gabor or wavelet filters[10]. In 2-D Gabor wavelets, 2-D mesh and a 2-D golden section algorithm based on object tracking method, the GWT coefficient used to represent the local colour features and it is able to track objects in a video sequence with large deformations.

The increase of description features improves accuracy their combination leads to high-dimensional PDFs. There exist efficient method to estimate multivariate PDFs using Parzen windowing[6]. Due to fixed cardinality of the data set, a limitation known as the curse of dimensionality appears: as the dimension of the domain of definition of the PDFs gets higher the domain sampling gets sparser. Hence the resulting PDF is oversmoothed. The combination of colour and geometric features improves the tracking accuracy[4]. This paper deals with the colour feature extraction using the colour histogram model, and the geometric feature extraction using the moment invariant analysis. Moment invariants are evaluated as feature space for pattern recognition in terms of discrimination power and noise tolerance[8]-[9]. Geometric moment invariant produce a set of feature vectors that are invariant under shifting, scaling and rotation which is widely used to extract the global features for pattern recognition due to its discriminative power and robustness.

Similarity measures based on the KNN PDF estimator is only adapted to high dimensions the resulting entropy estimator appears to be accurate in both low and high dimensions[1]-[13]. We use an estimator for Kullback-Leibler divergence efficient for high dimensional PDFs. The distance is expressed from the samples using the Kth nearest neighbour framework(KNN).

2. PROPOSED WORK

The flow chart of the ROI tracking algorithm is shown in Fig.1. The reference frame and target frame from the video frames is selected first. Then the particular region-of-interest is selected from the reference frame and the (Y,U,V) colour feature was extracted. The colour and the geometric features are extracted from the reference and the target frame. Finally, the Kullback leibler Divergence is used for finding the similarity measure between the reference and the target frames.

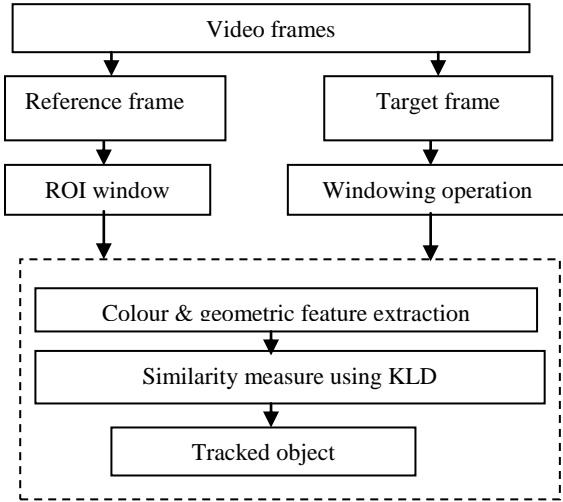


Fig 1. Flow diagram for tracking

2.1 Colour feature extraction

Normally all the video sequences are in (R,G,B) format. They are converted into (Y,U,V) format. The components of the feature vectors were normalized as follows: Y,U, and V were rescaled into the interval [0,1] and the coordinates (x,y) were rescaled into [-1,1], both the ROI and the candidate regions, the origin being located at the centre of the bounding box region. Tracking was performed with IR being set to I1 while IT was successively equal to It = 2,3,4... When searching for the ROI in frame It, the search area was empirically limited to ± 12 pixels horizontally and vertically around the position of the center of the ROI computed in frame It-1.

A well-known density estimator is the histogram., the (one-dimensional) histogram is defined as,

$$f(x) = \frac{1}{nh} \times (\text{number of } X \text{ in the same bin as } x) \quad (1)$$

where n is the number of pixels with value X in the image, h is the bin width and x is the range of data. Two choices have to be made when constructing a histogram. First, the bin-width parameter needs to be chosen. Second, the position of the bin edges needs to be established. Both choices affect the resulting estimation.

2.2 Geometric feature extraction

The geometric features are extracted using the moment invariant analysis. In image processing, computer vision and related fields, an image moment is a certain particular weighted average (moment) of the image pixel's intensities, or a function of such moments, usually chosen to have attractive property or interpretation. Image moments are useful to describe objects after segmentation. There are two types of moment invariants. They are region based and contour based. Here contour based moment was applied. This moment invariants are used for computer vision applications, remote sensing, medical imaging, trademark identification, insect identification and pattern recognition. Normally the moment invariants are described as follows:

Geometric moment m_{pq} of image $f(x,y)$ where p,q are non-negative integers and (p+q) is called the order of the moment, is defined as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy \quad (2)$$

Corresponding central moment and normalized moment are denoted as

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) dx dy \quad (3)$$

Where,

$$\bar{x} = \frac{m_{10}}{m_{00}} \text{ and } \bar{y} = \frac{m_{01}}{m_{00}} \quad (4)$$

The normalized central moment of order (p+q) is defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (5)$$

For p+q=0,1,2,...where,

$$\gamma = [(p+q)/2] + 1 \quad (6)$$

A set of seven 2-D moment invariants that are insensitive to translation, scale change, mirroring, and rotation can be derived from these equations. They are

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\ &\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) \\ &\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned}$$

The moments invariants are determined for the selected ROI and the scaled ROI images. The scaling factor was normally taken between 0.98 to 1.02.

Table 1. 2-D moment invariants

moments	Sample video 1		Sample video 2	
	Original ROI	Scaled ROI	Original ROI	Scaled ROI
ϕ_1	0.0013261	0.0013315	0.0013235	0.0013264
ϕ_2	2.059e-09	1.87e-09	1.33e-08	1.349e-08
ϕ_3	1.05e-10	1.057e-10	6.314e-11	6.269e-11
ϕ_4	1.20e-10	1.217e-10	7.51e-11	7.437e-11
ϕ_5	1.09e-20	1.10e-20	4.05e-21	4.083e-21
ϕ_6	-1.15e-10	-1.16e-10	-7.36e-11	-7.31e-11
ϕ_7	8.07e-21	8.307e-21	3.216e-21	3.018e-21

2.3. Similarity measures

Let IR be the reference frame in which the ROI is defined and IT is the target frame in which the ROI is determined using the similarity measures. The amount of similarity is given by the geometric transformation

$$\phi = \arg \min D_{KL}(I_R, I_T) \quad (7)$$

Where DKL is the Kullback Leibler divergence between the reference and the target frames. Normally similarity measures are determined in two ways. They are colour based and geometry based. The colour based method checks if the regions have similar colours and the geometry checks if these colours appear at the same location.

2.3.1. Colour based similarity measures

One of the colour based similarity measure used for tracking is Bhattacharya distance.

$$D_{BHA}(I_T, I_R) = \int \sqrt{f_R(s)f_T(s)} ds \quad (8)$$

Another widely used similarity measure is the Kullback-Leibler distance

$$D_{KL}(I_T, I_R) = \int f_T(s) \log \frac{f_T(s)}{f_R(s)} ds \quad (9)$$

This equation can be decomposed as follows

$$\begin{aligned} D_{KL}(I_T, I_R) &= \int f_T(s) \log f_T(s) ds \\ &\quad - \int f_T(s) \log f_R(s) ds \\ &= -H(I_T) + H_{\times}(I_T, I_R) \end{aligned} \quad (10)$$

Where H is the differential entropy and HX is the cross entropy, also called relative entropy or likelihood.

2.3.2. Geometry based similarity measures

Geometry can be added by using transformation ϕ , which represents a motion model, to compute the pointwise residual

between the reference and the target frames. A function of residual can serve as a similarity measure, classically in the discrete framework, SSD

$$D_{SSD}(I_T, I_R) = \sum_x (T(x) - R(x))^2 \quad (11)$$

Or function used in robust estimation such as SAD or differentiable approximation of SAD

$$D_{SAD}(I_T, I_R) = \sum_x \phi(T(x) - R(x)) \quad (12)$$

3. EXPERIMENTAL RESULTS

This algorithm had been tested on several video datasets. The appearance of the object and its background keep changing during almost all the time. The exciting results are obtained in most experiments. Here three video sequences are used to test the ability of the algorithm to track object deformations and shifts. Only the first frame in every 30 frames is used to force the motion and deformations to be large. In the first frame, a rectangle is interactively selected by the user. Although the user specifies a rectangle to cover the object to find the feature point selection, all the feature points are located inside the object. The experiments also prove that the tracking algorithm is very efficient.

3.1 Experiment-I

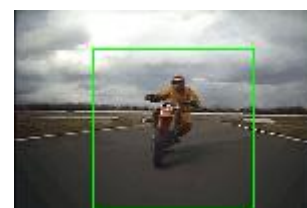
This video sequence consists of 174 frames and 29.966 frames per second. Here the window size is changed to 80X80. This will result more tracking accuracy than the previous video. Fig.2. shows the tracking results of the particular ROI.



Frame 1



Frame 30



Frame 60



Frame 90



Frame 120

Fig.3. ROI tracked in frame 1,30,60,90,120.

3.2 Experiment-II

This video sequence consists of 103 frames and 29.96 frames per second. Here the window size is changed to 80X80. Fig.4. shows the tracking results of the particular ROI.



Frame 1



Frame 30



Frame 60

Fig.3. ROI tracked in frame 1,30,60.

Table 2. Comparison between colour based and colour+Geometric feature tracking

Window position	Frames	Colour based tracking		Colour + Geometry based tracking	
		Row	Column	Row	Column
Sample video1	Frame1	9	46	10	46
	Frame30	20	46	20	56
	Frame60	20	39	20	40
Sample video2	Frame1	14	58	15	55
	Frame30	20	33	22	35
	Frame60	6	15	18	23

The tracking accuracy improvement in the video samples was compared using the Kullback Leibler divergence between the frames. Minimum values of divergence denote better tracking in the video frames.

Table 3. Kullback Leibler divergence

Kullback Leibler Divergence			
Colour tracking based	Sample video1		0.0469
	Sample video2		0.0481
Colour + Geometry based tracking	Sample video1		0.0438
	Sample video2		0.0467

4. CONCLUSION

This paper presented an approach for tracking the region-of-interest (ROI) used by the combination of (Y,U,V) colour features with the moment invariant geometric features. Here the similarity measures are determined by using Kullback-Leibler divergence. Experiments proved that ROI was perfectly tracked with high accuracy. In future tracking will be done using Bregman divergence [11] and the geometric features are extracted using Gabor wavelets. This work can be further extended to track more than one object in video sequences.

5. REFERENCES

- [1] Sylvain Boltz, Eric Debreuve, & Michel Barlaud, 2009 High-Dimensional statistical measure for region-of-interest tracking, *IEEE Transactions on image processing*, Vol.18,No.6,ppl 1266-1283.
- [2] S. Boltz, A. Herbulot, E. Debreue, M. Barlaud, & G. Aubert, 2008.Motion and appearance nonparametric joint entropy for video segmentation *Int.J. comput. Vis*, vol 80,no.2, 242-259.
- [3] D. Comaniciu, V. Ramesh, & P.Meer, 2000.Real-time tracking of non-rigid objects using mean shift, presented at *Int. conf. computer vision and pattern recognition*.

- [4] T. Brox, M. Rousson, R. Deriche, & J. Weckert, 2003. Unsupervised segmentation incorporating colour, texture, and motion, presented at the computer analysis of Images and Patterns, Gronings, the Netherlands.
- [5] P. Perez, C. Hue, J. Vermaak, & M. Gangnet, 2002. colour-based probabilistic tracking”, Presented at the European conf. computer vision, Copenhagen, Denmark.
- [6] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, & L. Pronzato, 2006. A minimum-entropy procedure for robust motion estimation”, presented at the Int. conf. Image Processing, Atlanta GA.
- [7] R.V. Babu, P. Perez, & P. Bouthemy, 2007. Robust tracking with motion estimation and local kernel-based colour modeling, *Image Vis.comput.* vol.25.no.8, pp11205-1216.
- [8] Yaser S. Abu-Mostafa & Demetri Psaltis, 1984. Recognitive aspects of moment invariants, *IEEE Transactions on pattern analysis and machine intelligence*, Vol.PAMI-6, No. 6, pp1 698-706.
- [9] Mohamed Rizon, Haniza Yazid, Mohd Rozailan Mamat, 2006. Object detection using geometric invariant moment, *American Journal of Applied Sciences* 2(6): 1876-1878.
- [10] Chao He, Yuan F. Zheng, and Stanley C. Ahalt, 2002. Object tracking using the Gabor wavelet transform and the golden selection algorithm, *IEEE Transaction on Multimedia* Vol 4.No.4, 528-538.
- [11] A. Banerjee, S. Merugu, I. Dhillon, & J. Ghosh, 2005. Clustering with bregman divergence”, *J. Mach. Learn. Res* vol. 6, 1705-1749.
- [12] A. Bugeau and P. Perez, 2007. Detection and segmentation of moving objects in highly dynamic scenes”, presented at the Int. Conf. Computer vision and pattern recognition, Minneapolis, MN.
- [13] Paolo Piro, Sandrine Anthoine, Eric Debreuve, Michel Barlaud, 2008. Image retrieval via kullback leibler divergence of patches of multiscale coefficients in the KNN framework”, published in CSMI 2008, proceeding of the International workshop on content-based multimedia indexing.