지역적 영역 컬러 특징 정보를 이용한 이동물체 추적

주문원, 진성아
성결대학교 멀티미디어학부

Object Tracking in Video Sequences using Local Block Features

Moon Won Choo, Seongah Chin
Division of Multimedia, Sungkyul Univ.
E-mail : mchoo@sungkyul.edu

요 약

본 논문에서는 칼라 동영상에서 물체 이동에 의하여 형성된 동작영역을 확인하고, 이동방향을 추적하는 시스템을 제안한다. 비디오 동영상에서 포착된 물체의 영역을 color invariance 의 분석을 통해 추출하고, 추출된 영역에서 radial homogeneity 정도를 영역의 특징값을 추출하여 대응되는 물체 영역을 추적함으로써 물체의 궤적을 확인한다.

In this paper, we propose an object tracking system which extracts moving areas+ shaped on objects in video sequences and decides tracks of moving objects. Color invariances are exploited to extract the plausible object blocks and the degree of radial homogeneity is utilized as local block feature to find out the block correspondences.

1. introduction

Tracking the motion of objects in video sequences is becoming important as related hardware and software technology gets more mature and the needs for applications where the activity of objects should be analyzed and monitored are increasing[3][4][5][11][12]. In such applications lots of information can be obtained from trajectories that give the spatio-temporal coordinates of each objects in the environment. Information that can be obtained from such trajectories includes a dynamic count of the number of object within the monitored area, time spent by objects in an area and traffic flow patterns in an environment. The tracking of moving object is challenging in any conditions, since image formations in video stream is very sensitive to changes of conditions of environment such as illumination, moving speed and directions of objects, the number and sizes of objects, and background. Therefore the scope of researches are usually confined to specific application domains and the processes of capturing video streams are also carefully controlled . In this paper, general moving objects are considered and color information is utilized as main source for extracting features for tracking since gray-level images may lose much of information available in color space such as combined and synthesized features derived from separate color channels. We suggest a system for obtaining such spatio-temporal tracks of objects in video sequences. Camera in static position produces video sequences which are analyzed in real time to obtain trajectories. In each frame of video stream, segmentation technique such as differencing gray-level intensities embedded in inter-frame images could well work in real time and yield regions of interest for blocking quickly. An important step towards the track of objects is the definition of a proper set of features which could reliably identify the corresponding objects in adjacent frames. Most of the motion tracking researches focus on the features generated from the gray-level images which generally derived from the RGB space. Color image can be assumed to contain richer information for image processing than its gray-level image. Also separate color channel could be applied to different problem domains. In this paper, color invariances are exploited to extract the reliable features for tracking moving objects.

2. Related Research

Jakub and Sarma[7] developed a system for realtime tracking of people in video sequences. They use a model-based approach to object tracking, identifying feature points like local curvature extrema in each video frame. Their system has an advantage of handling occlusion problems, but disadvantage of unreliable extraction of extrema of curvature from object contours. William[13] suggests motion tracking by deriving velocity vectors from point-to-point correspondence relations. Relaxation and optical flow are very attractive methodologies to detect the trajectories of objects[10]. Those researches are based on the analysis of velocity vectors of each pixel or group of pixels between two neighboring frames. This approach requires heavy computation for calculating optical flow vectors. Another method infers the moving information by computing the statistical features, difference images and edge features for complementary information to estimate plausible moving tracks[1][2][11][16]. This method may be very sensitive to illumination and noise imposed on video stream. The other method adopts the model-based approach, which has disadvantage of extracting the previously trained objects only[14]. A flow chart showing the main steps is given in figure 1.

3. The moving object tracking system

3.1 Block detection module

This module plays a very important role in object tracking applications. Many blocking methods assume that the lighting in the scene considered would be constant. The accuracy of these methods decreases significantly when they are applied to real scenes. A multiple-level blocking method based on thresholded differences and a morphological focus of attention able to reduce the effects of noise and of changes in lighting is suggested. In this paper, Kubelka-Munk theory which models the reflected spectrum of a colored body based on a material-dependent scattering and absorption function, under assumption that light is isotropically scattered within the material[8][9]. This module receives a pair of gray-level image frames, $I_{k-1}(x,y)$ and $I_k(x,y)$, acquired at successive time instants t_{k-1} and t_k

respectively. Then a list of minimum bounding rectangle-shaped blocks of image areas where significant changes (related to possible moving objects) is produced. The proposed module consists of three steps.

(1) computing the difference $D_k(x,y)$ between the two input images $I_{k-1}(x,y)$ and $I_k(x,y)$, $D_k(x,y) = |I_k(x,y) - I_{k-1}(x,y)|$.

(2) Establishing whether each point $(x,y) \in D_k(x,y)$ is a background point or a moving-object point, then generating the binary block image B. The function used for point labeling is a spatial hysteresis function designed to improve the algorithm robustness[4]. If the state is background and $D_k(x,y) > \theta_{in}$, the state of the point is switched to object. If the state is object and $D_k(x,y) > \theta_{out}$, the state of the point is switched to background. The selection of threshold values, θ_{in} and θ_{out} are strictly dependent on the geometry of the image frame considered such as dimensions of a moving object and distance between moving objects and camera, etc..

(3) Noise filtering and searching for the minimum bounding rectangular shaped blocks are performed by means of simple morphological operation and searching extremal points using progressive projection. Progressive projection is very effective and fast method to isolate the region of interest from the binary image.

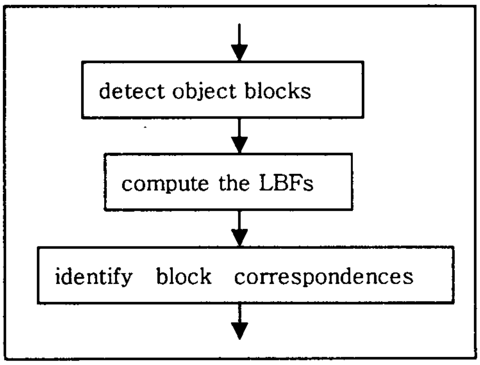

Fig. 1. Main steps in the moving object tracking system.

3.2 LBF(Local Block Feature) extraction module

The presence of object detected with its bounding rectangular block is associated with the RGB values imposed on corresponding pixels. Within the Kubelka-Munk model, assuming dichromatic reflection and equal energy illumination, $H = E_\lambda / E_{\lambda\lambda}$ is an object reflectance property independent of viewpoint, surface orientation, illumination direction, illumination intensity and Fresnel reflectance coefficient. This invariance value is adopted for calculating the LBF of each detected object region. To get this spectral differential quotients, the following implementation of Gaussian color model in RGB terms is used (for details, see [8]).

$$\begin{bmatrix} E \\ E_\lambda \\ E_{\lambda\lambda} \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

After obtaining the color invariance image H, the LBFs of each block are computed utilizing this image matrix. Since the H determines the ability to find unique interframe block correspondences, some transform[16] by considering the sources of H within a local image block that may contains an occlusion boundary and background. This should ignore the undesirable effects of other objects and background, but be sensitive to color features embedded in detected objects. The color features around central regions of objects should be captured fully both in magnitude and sign by transform. The transform captures the color invariance values closest to the center of the block, and attenuates all else. Hence, it is comprised of a central color invariances and a local neighborhood of this attributes. The neighborhood computes the local invariance relative to the central attribute attenuated to discount background influence.

Formally, given a block image $B(x,y) \in I(x,y)$, a LBF of B is comprised of two terms, a central value, and a neighborhood function:

$$L_B(x,y) = \{C_B(x,y), N_B(x,y,r,\theta)\}.$$

The central value is the color invariance value averaged within a small radius of the given image location:

$$C_B(x,y) = \frac{1}{\pi R^2} \sum_{r,\theta}^{r \le R} H(x + r\cos\theta, y + r\sin\theta)$$

The radius R is determined depending on the image condition and the requirement specifications of a particular applications. However, the minimum dimension of moving objects to be detected could be set without any difficulties. In this paper, the R from 5 to 10 pixels (the image dimension is 480 * 640), gives good results.

The neighborhood function N is proportional to the likelihood that underlying image attribute is unchanged along a ray from the center of the block to that pixel. To compute the local image invariance energy N, which is simply the MSE between the central point at which the transform is being defined (x,y) and nearby points at a given radical offset (r,θ).

$$E_B(x,y,r,\theta) = \alpha \| (C_B(x,y) - H(x + r\cos\theta, y + r\sin\theta)) \|^2$$

,where α is a color invariance sensitivity coefficient. In this paper, this value is set to 1. The intergral of E along a ray from the central point, and take the negative exponential to obtain N.

$$N_B(x,y,r,\theta) = \exp\{-\int_{\rho<r} E_B(x,y,\rho,\theta)d\rho\}.$$

N_B and E_B are defined over block coordinates $r \le \min(W_x, W_y)$, where W_x, W_y are the block dimensions along x-axis and y-axis respectively, and $0 \le \theta \le 2\pi$.

3. 3 Tracking module

The goal of this module is to determine the 3-D positions and the motion parameters of objects recognized by the blocking module, at every time instant. Each detected object block may be assumed to contain only one moving object. But this constraint can be alleviated, since the LBF could be jointly utilized to differentiate multiple objects possibly occluded.

To establish the correspondence relations of blocks between sequential frames, the displacement of a LBF at time t are compared with the LBFs at time $t+1$.

$$(x', y') = \arg\min_{l} D_\lambda(L_B^t(x,y), L_B^{t+1}(x',y')),$$

where l is the number of detected blocks in I_{t+1}. The LBF distance D_λ is defined by computing the weighted L_2 error of the transformed data using a combination of neighborhood difference and central value difference terms.

$$D_\lambda(L_B^t(x,y), L_B^{t+1}(x',y')) = (1-\lambda)\Delta N + \lambda \Delta C.$$

The neighborhood difference ΔN is defined as the MSE between $N_B(x,y,r,\theta)$ and $N_B(x',y',r,\theta)$ computed over r,θ. ΔC is the MSE between $C$and C'. The bias term λ expresses a trade-off between the contribution of the central attribute error and the neighborhood function error. Generally, the neighborhood error is the most important, since it captures the spatial structure at the given point. However, in certain cases of spatial ambiguity, the central attribute value is critical for making the correct match unambiguous. In real situation, this bias term could be adjusted dynamically if a priori knowledge about objects is available.

4. Experimental results

Several video clips are generated using static SONY digital video camera according to the kinds of objects and their speed variations under natural illumination conditions. Also the variety of different values of parameters associated with the morphological operator, differential operators applied to projection table, and the bias for LBF are tested.

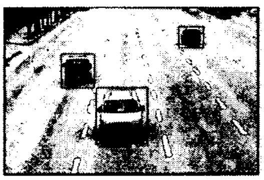

Fig.2 the video frame tested and detected blocks.

Fig.2 show the one of the video clips tested. This frame illustrates the results of well-defined block detection module. But the objects which move relatively slower generate somewhat ambiguous trajectories(see Fig.3).

Fig. 3 the trajectories detected.

Fig. 3 shows the detected tracks of three objects after processing 20 frames running 30fps. Figi.4 shows the difference image to be used to detect the blocks of interest. Several morphological operators to eliminate the noises. In this case, the simple dilatation is applied with 2*3 structure element.

Fig. 5 the difference image after applying the morphological operator.

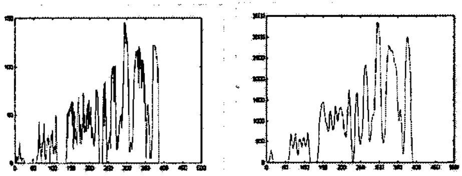

Fig. 6 (a) the projection of the difference image along the x axis. (b) after applying Gaussian filter

Fig.6 shows the result of projection along the x axis. To compute the start and ending locations of each block, the false alarm should be detected early by applying the Gaussian filters. Progressive filterings are applied using scalable Gaussian kernels [tony].

5. Conclusions and future research

A system for tracking objects using color invariance features is presented. The system outputs tracks that give spatio-temporal coordinates of objects as they move within the field of view of a camera. We try to solve the occluded objects problem, which may be ignored in many researches intentionally by utilizing radial cumulative similarity measures imposed on color invariance features. This system is very fast to be used in real time applications. But several parameters should be adjusted adaptively in order to be used in more general settings. The occlusion issues are not solved properly. Future research needs to address these kinds of issues and tracking objects across moving cameras.

References

[1] A. Latif, et. al," An efficient method for texture defect detection: sub-band domain co-occurrence matrices," Image and Vision Computing 18, pp. 543-553, 2000.

[2] B. Chanda, B.B. Chaudhuri, D. Dutta Majumder "On image enhancement and thread selection using the greylevel co-occurrence matrix", PR Lett. Vol.3, no.4, pp. 243-251, 1985

[3] D. Beymer and K. Konolige, " Real-Time Tracking of Multiple People using Stereo, " Proc. IEEE Frame Rate Workshop, 1999

[4] Gian Luca Foresti, et. al,"Vehicle Recognition and Tracking from Road Image Sequences," IEEE Trans. On Vehicular Tech., vol. 48, NO. 1, Jan. 1999.

[5] Gonzalez, Woods "Digital Image Processing" Addison Wesley 1992

[6] Ismail Haritaoglu, "David Harwood, Larry S. Davis, " W^4: Real time Surveillance of people and their activities", IEEE Trans. on pattern analysis and machine intelligence, Vol. 22, No. 8, Aug. 2000

[7] Jakub Segen and Sarma Pingali,"A Camera-Based System for Tracking People in Real Time," IEEE Proceedings of ICPR '96, 1996

[8] Jan-Mark Geuseboek, et. Al,"Color Invariance," IEEE Trans. PAMI, vol 23, no.12, Dec. 2001

[9] Kristen Hoffman,"Applications of the Kubelka-Munk Color Model to Xerographic Images," http://www.cis.rit.edu/research/thesis/bs/1998/hoffman/contents.html

[10] Milan Sonka, Vaclav Hiavac, Rogger Boyle " Image processing Analysis and machine vision," International Thomson Publishing Co., 1999, 2nd edition

[11] Rita Cucchira, Massimo Piccardi, Paola Mello "Image analysis and rule based reasoning for a traffic monitoring system" IEEE, Intelligent transportation system, pp119-pp130VOL.1, No.2, June 2000

[12] Robert M, Haralick, Linda G, Shapario " Computer and Robot vision Vol I ", Addision-wesley, USA pp 318-321, 1992

[13] Ross Culter, Larry S. Davis " Robust Real-Time Periodic Motion Detection, and analysis and Application", IEEE Pattern analysis and machine Intelligence Vol22 No 8, pp781-795 August 2000

[14] Seongah-Chin "Image text segmentation", Ph.D dissertation 1999, SIT

[15] Tony Lindelberg," Feature Detection with Automatic Scale Selection," Int. J. of Computer Vision, vol 30, number 2, 1998.

[16] Trevor Darrell and Michele Covel " Correspondence with cumulative similiarity transforms" IEEE pattern analysis and machine intelligence, pp 222-227 Vol.23 No2 February 2001