Chapter 1

Introduction

Motion classification is an important task in image sequence processing and computer vision applications. There is a high demand for an effective method for the classification of image sequences. For instance, successful classification of objects from background can lead to an object-based coding such as MPEG4 [1], [6-7]. A number of algorithms have been proposed in the past and each of which has its particular features and applications.

Image motion is one of the most important information that can be used for image classification. Knowing the direction and velocity of moving objects, it is easy to detect the position of objects at any moment. To achieve the better classification, it is necessary to compute dense motion vectors in the sequence.

For real time implementation, gradient based optical flow techniques are widely used for estimating motion vector fields. An appropriate choice of a technique for computing dense motion vectors requires the consideration of both its accuracy and computational cost. Gradient based methods are used to obtain dense motion vectors and the method is effective when the displacement between images over time is small [14].

In many applications such as surveillance systems, the most common techniques for detection of moving objects are to calculate the difference image between two consecutive frames. Then a pixel is considered to belong to moving area if the temporal difference exceeds a pre-determined threshold value. Determination of this threshold is based on the estimation of noise in image sequence and may differ from one image to another. Background subtraction is another common method in detection of objects in surveillance systems. However, one needs to consider of changes due to noise, illumination variation or background changes.

Classification of moving objects is one important task in segmentation process. A number of algorithms have been proposed in the past. Neural network techniques are successful and efficient for object classification. There are three types of neural networks: supervised, unsupervised, and a combination of the two. For the image segmentation, unsupervised learning is preferred to supervised learning because the later requires a set of training samples which may not always be available. In unsupervised, self organizing map is based on competitive learning and it is widely used in image segmentation applications.

In this study, we are going to perform the accurate and automatic motion estimation in image sequences and visualization of motion patterns for real time applications. We are going to choose one of motion estimation techniques which give accurate and less computational cost. For visualization of motion vectors, we are going to choose a neural network technique which is promising for real time data.

1.1 Moving Object Segmentation

1.1.1 Motion estimation techniques

Motion estimation is normally done by using two main methods which are optical flow and block matching. In both approaches, motion information is extracted through detecting the change of pixel intensities between consecutive frames in image sequence. However, optical flow estimation method is often chosen for achieving accurate motions in image sequence segmentation because it allows motion detection at pixel level and ensures better object's boundaries than what block matching method can accomplish. Moreover, optical flow estimation is less computational cost than block matching method [4].

Gradient information of the image serves several purposes. It can relate the structure of objects in an object, identify features of interest for segmentation/ classification directly or provide the basis of further processing for various computer vision tasks.

Gradient representation can be categorized into two techniques:

1. Gradient Method (Spatio-temopral gradient method)
2. Gradient Structure Tensor based method (GSTM)

Both methods are applicable to three dimensional data, such as a spatio-temporal volume (x, y, t) or medical imaging input (x, y, z).

The spatio-temporal gradient method (often simply referred to as the gradient method or differential method) is a conventional choice for estimating motion in an image sequence [3]. However, it is a known fact that the motion vectors estimated by the gradient method (GM) are not sufficiently accurate, especially when an image is noisy or full of fine textural patterns. In addition, the range of motion that GM can estimate is small, typically a few pixels only. These problems have restricted the use of motion information for image segmentation.

Both methods are based on the well-known optical flow constraint equation (OFCE) [3]. GM estimates motion vectors by solving the OFCE using the least-squares method, while the gradient structure tensor method (GSTM) solves the OFCE using the total least-squares approach [7]. GSTM works far more accurately and robustly than GM, and also it can estimate larger motion much better than GM [11].

In 2003, 3-D structure tensor is introduced in Hai-Yun Wand and Kai-Kuang Ma paper about automatic video object segmentation via 3-D structure tensor. Their work showed the new scheme by generating the spatial-constrained motion masks without computing dense motion fields. They have used tensor as a correlation matrix and locates the edges of moving objects. [5].

In 2004, R. Strzodka and C. Garbe implemented the real-time motion estimation and visualization on graphic cards by eigenvector analysis of the spatio-temporal structure tensor at every pixel location. For the visualization method, they have used coloring, blending and fading methods [13].

In 2005, R. Pless and J. Wright presented an analysis of persistent motion patterns. 3-D structure tensor is used in background modeling as a convenient tool for representing the joint distribution of x, y and t derivatives at each pixel. Their algorithm runs in real – time and can adapt to both slow drift and abrupt changes in motion patterns. Since structure tensor method considers covariance rather than intensity itself, their algorithm is robust to lighting changes in the scene [12].

Therefore, to improve the accuracy and reliability of motion estimation, we employ the gradient structure tensor method (also referred to as gradient square tensor method or 3-D structure tensor method).

1.1.2 Motion classification and visualization technique

Self-organizing map (SOM) is a data visualization technique invented by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing neural networks. The problem that data visualization attempts to solve is that humans simply cannot visualize high dimensional data as it is, so techniques are created to help us understand this high dimensional data. SOM reduces the dimensions by producing a map of usually 1 or 2 dimensions which plot the similarities of the data by grouping similar data items together. Thus, SOMs can accomplish two things; they reduce dimensions and display similarities.

In 1996, Suchendra M. Bhandarkar, Jean Koh and Minsoo Suk used extended SOM neural networks for image segmentation. The extended network, hierarchical self organizing feature map (HSOFM) was applied to intensity and range images and it was shown that HSOM can be used in any application that calls for clustering of vectors in feature space[15].

J. Kim and T. Chen used block matching technique for estimating motion vectors and extract multiple features such as motion, texture and color information on the pixel basis by using SOM. For segmentation, they have proved that clustering algorithms such as K-means and fuzzy C-means are iterative and difficult to implement a practical system, however, neural network clustering, SOFM is promising, simple and 22 percent faster than the conventional K-means clustering algorithm [6],[7].

SOM is mostly used in color based image segmentation since it is more efficient and robust than the methods based on edge or region. In 2003, X. Wu and L. Xu hand detection paper states that SOM is fast and robust in dynamic gesture segmentation. In this paper, they used block matching technique to detect the hand motion [18].

SOM can cluster input data into groups without external supervision. SOM can be classified motion vectors into several groups, which leads to the segmentation of an image sequence.

1.1.3 Moving object extraction techniques

A number of algorithms have been proposed to extract moving objects from stationary background. The most common techniques for extracting moving objects are to calculate the difference image for two consecutive frames. Then, a pixel is considered to belong to changed area if the value of the corresponding pixel in the difference image exceeds a pre-determined threshold. Accumulative difference images (ADIs) technique is normally used for moving object extraction because of the demand of computational efficiency.

In 2004, Desa. S.M and Salih. Q.A have proposed a technique of extracting moving object from static irrelevant background by background subtracting and consecutive frame differencing [3]. In the same year, J. Yuan and Z. Shi also used frame differencing method. After dividing the difference image into blocks of equal size, they used mean and standard deviation of each block to separate background and foreground of the image [9].

1.2 Motivations and Goals of Thesis

Successful image sequence segmentation leads to object-based video coding, such as MPEG-4, in which objects and its background in an image are encoded separately, achieving higher compression rates. Image sequence segmentation is important not only for video coding, but is also useful for video indexing in order that we can automatically find particular scenes from numerous video collections. Provided segmentation is performed at the video rate, a certain object can be located and tracked, which finds challenging applications such as human face tracking for surveillance cameras and hand tracking for gesture recognition and so on. Therefore, the goals of this thesis are :

1) To study characteristics of surveillance image sequences
2) To study algorithms of moving object extraction such as
    - Accumulative Difference Images (AADI)
    - Background Subtraction by Temporal Median Filtering
3) To study gradient representation of images such as
    - Gradient Method (GM)
    - Gradient Structure Tensor Method (GST)
4) To study effective segmentation algorithm such as Self-Organizing Feature Map (SOM)
5) After finding the suitable algorithm, we will present the effective technique for image sequence segmentation. We will test our technique using MATLAB (version R2006a).
6) For the final step, we will improve the performance of segmentation and apply to the real-time image sequences.

1.3 Thesis Organizations

Chapter 1: Introduction.

- 1.1) Moving Object Segmentation
  - 1.1.1) Motion estimation techniques
  - 1.1.2) Motion classification and visualization technique
  - 1.1.3) Moving object extraction techniques
- 1.2) Motivations and Goals of this thesis
- 1.3) Thesis Organizations

Chapter 2: Motion Estimation and Classification

- 2.1) Motion Estimation in Image Sequences
  - 2.1.1) Gradient Structure Tensor Method (GSTM)
- 2.2) Motion Classification and Visualization
  - 2.2.1) Self-Organizing Feature Map (SOM)

Chapter 3: Application to Surveillance Camera

- 3.1) Moving Object Extraction/Segmentation
  - 3.1.1) Accumulative Difference Images (ADIs)
  - 3.1.2) Background Subtraction using Median Filter
- 3.2) Motion Visualization within a Moving Object

Chapter 4: Results and Discussion

- 4.1) Motion estimation results using GSTM
- 4.2) Classification of motions using SOM
- 4.3) Moving object extraction results from stationary background
- 4.4) Applications to other surveillance systems

Chapter 5: Conclusions

References: