VIDEO FEATURE EXTRACTION

ABSTRACT

The Video is collection of frames which are refreshing at a rate at which human eyes are not able to recognize them as separate. So Extracting the Video Features in abstract form means extracting features from the frames which when collectively runs create a Video. If we have to understand we have to work on two steps mainly, the local feature of a particular frame, like question 1 and 2 which is extracting the local features of the frame, the second is extracting the relative features with respects to frames. The features which are dependent on the frames which are in vicinity of that Frame, like question 3 of this project that is Extracting Motion Vectors.

These local and relative features are then being used to Represent information which can be further used for Searching and many other applications. Most of the compression techniques are based on the extraction of these features.

Group NO 21

Name of the Team Members:

Karan Uppal, Meha Shah, Prashant Garg, Priyanka Dubey, Sweta Singhal

Keywords:

Color Histogram, Sift Vectors, Sift descriptors, Motion Vectors.

INTRODUCTION

Color Histogram: - Color Histogram is numbering the occurrence of a color with respect to its color representation. The numbering unit is mainly considered as pixel in terms of an image. [1]

SIFT (Scale-invariant feature transform): - The algorithm was published by David Lowe. This algorithm finds the local features in an image, mainly the point of intersection of curves, which become the property of that object in an image. If that object is used in any scaled form, sift can be used to identify it, but it should not have been rotated axis other than out of the plane of image axis. [2]

Bins: - Bins size are some value ranges in terms of histogram to make histogram count a Value range as one entity. By default, for a grayscale image bin size is 1 and are 256 in number. [3]

Motion vector: - A Motion vector is a feature which is extracted by comparing adjacent images. The motion vector compares an image with its adjacent image to find out where that particular region has shifted to another section in the adjacent image.[4]

It has some attributes like: -

x_i – srcx: absolute source position of row

y_i – srcy: absolute source position of column

x_d – dstx: absolute destination position of row

y_d – dsty: absolute destination position of column

Scaling: - The zooming in and zooming out of an object is scaling that object. It is relative increase of the object with respect to its length such that its centre of mass point doesn't changes. [5]

Orientation: - The rotation of an object about some arbitrary axis, in Two-Dimensions Images, we mainly consider rotation about Z axis (the plane which points towards or outwards of the Object's Image).[6]

Goal Description

The Goal of the project is to learn the Video Feature Extraction. The Project Phase is divided into 3 phases. First Phase focus on the conversion of Video into set of Frames and then converting them into grayscale for counting the intensity of the pixel Values of a sub section of an image called Cells. Second Phase Focus on the extracting of local features based on the change in intensity of curves, or mathematically finding the point of intersection of two independent curves. Third Phase focus on finding relative feature extraction MOTION Vectors which is relative motion of a section of an image with respect to a reference frame. In all these three phases, we are focusing on learning the Video Feature Extraction process. The goal is to learn these methods in this part of project.

Assumptions

Part 1: -

1. The assumptions made in in first question is that the extension of the Video file is mp4, as the code itself load all .mp4 files from the specified directory.
2. The second assumption made is That user will provide an integer input for r, which represents the value which will be used to convert the image into r*r cells, the value of r lies between 0 and min(size(image,1), size(image,2)).

   0 < r <= min((row ,column)) *, where row and column are x and y dimension of image*.
3. Third assumption is keeping the bin no as 256, in histogram (though it can be kept dynamic but it was not the purpose of the project).
4. The output is a file which is formatted with an assumption that the editor you are using understands the "\n" as the new line character.
5. The divisions of image into no of cells where the cells cannot be of equal size as by the limitation of image resolution is handled by me through my own algorithm, where it takes the floor value of the no which we get from the division of that row/column by the r and preserving the dot value until it becomes the complete integer, which then is considered as a complete pixel in next cell. The division if of the form as shown in the following table 1.1. The numbering of the cells in the output is done on the basis of the criteria shown in table 1.2.

Assume this to be an image of 120 x 160, where r value is 13.
The no in the cell is max value of row, column in the image

9,12	9,24	9,36	9,49	9,61	9,73	9,86	9,98	9,111	9,123	9,135	9,147	9,160
18,12												
27,12												
36,12												
46,12												
55,12												
64,12												
73,12												
83,12												
92,12												
101,12												
110,12												
120,12												

Table 1.1

Assuming the Value of r to be 4
The following shows the numbering of the cells for the output

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

Table 1.2

Part 2: -

1. The assumptions 1,2,4,5 made in this “PART 1” are same for this part.
2. Assumption that the output from sift required is frames and descriptors.
3. The sift vectors are first extracted from the image and then they are divided into cells by the same criteria defined in the above part.

Part 3: -

1. The assumptions 2,4,5 made in this “PART 1” are same for this part.
2. Assuming that the Motion Vectors are in correctly generated by the code given.
3. Assuming Cell division on the basis of the dist_x and dist_y values and R =6.
4. The assumption made in this path is that each video file path will be provided by the user in command line as an argument to the .exe. file

Program.exe “PATH TO VIDEO1” “PATH TO VIDEO2” “PATH TO ….” …..

Description of the Proposed Solution

1. For Part one, I am iterating each frame of each video one by one, then during that iteration, I am first converting that image into grayscale. Then dividing that image into cells by taking r as input from user, then finding intensity histogram using imhist() on that cell and saving the result in the file.
2. For Part two, I am iterating each frame of each video one by one, then finding the sift vectors in each frame using sift function as given as library of Matlab and then dividing them into cells and saving the output into file.
3. For third Part, I am passing all the video files path as command line argument and then finding the motion vectors in each frame, then saving the output in the final output file.

INTERFACE SPECIFICATIONS

1. For Part One: - The Matlab code contains three script, and as input from user, it demands the Value of r (the Division into r*r cells), the video file for processing are required which are assumed as in .mp4 format, output is the file which requires an editor which can read “\n” as new line character.
2. For Part two: - The Matlab code contains three script, and as input from user, it demands the Value of r (the Division into r*r cells), output is the file which requires an editor which can read “\n” as new line character.

3. For Part three: - The C++ code given requires you to pass the path to all video files as command line argument and the value of r, When the code by Visual Studio is done, the output file is created.

System Requirements

1. For Part 1 and 2, any system which have Matlab is good to go, independent of operating system with minimum 32-bit Processor, execution time can vary according to processor speed. Requires a space of around 400 MB. Matlab requires to have 2 GB Plus Ram for proper functionality.
2. For Part 3, you require a C++ compiler, and my solution requires WINDOWS as OS The solution is of 32 bit, so 32bit or 64bit processor will work, with requirement of around 200MB space.

Execution instructions: -

For Part One

1. Open the code from /Code/Part1/MAIN_Q1.m in Matlab and then set the current path to video files folder in the Matlab and then append the path of your workspace of the 3 code file in your code.
2. Then run the code, it will ask r (r *r cells) value from you, then it will create a text file Phase1Q1.txt in your pwd, which is the output file.

For Part Two

1. Open the code from /Code/Part2/MAIN_Q2.m in Matlab and then set the current path to video files folder in the Matlab and then append the path of your workspace of the 3 code file in your code.
2. Then run the code, it will ask r (r *r cells) value from you, then it will create a text file Phase1Q2.txt in your pwd, which is the output file.

For Part Three

1. Open command prompt, set pwd to /Code/Part3/debug, where .exe file is kept. Then run the command as shown
   \ffmpeg.exe “path to video file 1 “ “path to video file no 2” ……..
2. This will generate a file names as Phase1Q3.txt .

RELATED WORK

1. David Lowe, Demo Software: SIFT Keypoint, http://www.cs.ubc.ca/~lowe/keypoints/, version 4, July 2005

2. David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

3. David G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157

Conclusions

This phase of the project we learn about the feature extraction process, reading video as set of frames and then fetching useful features are very important, this Phase of the project taught us how local features of an image can be extracted, how image can be divided into sub parts. We also learn about how the features which are dependent upon the other frames can also be extracted. By doing this Phase, we have learnt a good amount of How Video feature extraction process take place.

Bibliography

The following site are refered for reading:-

[1] https://en.wikipedia.org/wiki/Color_histogram

[2] https://en.wikipedia.org/wiki/Scale-invariant_feature_transform

[3] https://www.quora.com/What-are-bins-in-histograms

[4] https://en.wikipedia.org/wiki/Motion_vector

[5] https://en.wikipedia.org/wiki/Image_scaling

[6] http://nsosp.nso.edu/node/62