Annotated Sketches for Intuitive Video Retrieval

EPSRC Engineering and Physical Sciences Research Council

Stuart James and John Collomosse

Centre for Vision Speech and Signal Processing (CVSSP)
University of Surrey, the United Kingdom

University of Surrey

Introduction

Free-hand sketch provides a natural and expressive modality for interaction with computers. This project explores methods to intuitively search video databases using sketches. Although video search is typically performed using keywords that specify content, text is cumbersome for describing scene appearance. Rather, a sketched depiction of a scene represents an orthogonal channel to constrain search. Although sketch based image retrieval (SBIR) has received much attention, the related problem of video retrieval (SBVR) is only sparsely researched – especially the fusion of text and sketch.

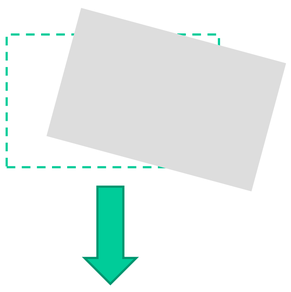
Building upon prior SBIR [1] and SBVR [2,3] we describe intermediate results from such a hybrid system. Our sketches describe objects based on their semantics (e.g. horse), a sketch of their motion trajectory, and their colour; collectively representing a natural interface for conveying multiple facets of events.

Annotated Sketch to Video Relationship

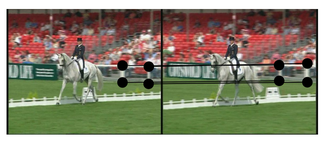
Sketch based retrieval presents a difficult problem of how to relate the sketch to the videos in the dataset. We relate the query canvas, to the video panorama. This allows for an intuitive indexing method as well as a clear relationship being formed. This relationship is able to describe colour and trajectory of a specific class.

Video Indexing

Camera Compensation



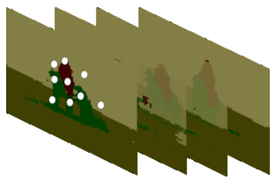
We calculate the camera motion of frames, doing so allows us to be able to create a panorama of the motion in the video. As mentioned before the panorama acts as a metaphor for the canvas. We use standard feature based techniques to calculate the camera motion.



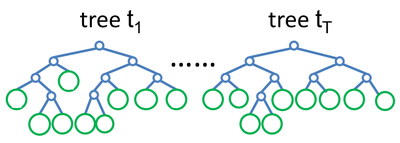
Feature Detection

To try to identify moving objects we detect sparse points on all frames in the video. We reduce the points to only identify moving points using camera compensated optical flow, this results in small magnitude for objects moving with the background.

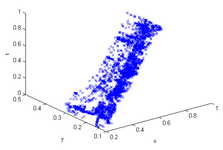
Semantic Segmentation



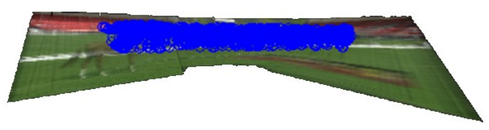
To take advantage of the semantic tags, we segment the video using a semantic segmentation put forward by Shotton [5], this approach allows fast segmentation of frames using extremely randomised decision forests.



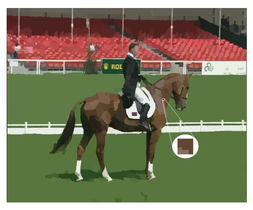
Spatial-Temporal Descriptor



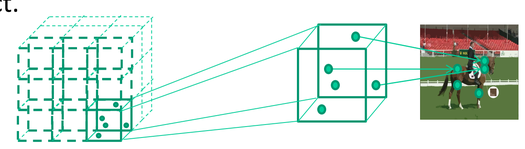
For each class in the dataset we normalise the points in panorama space and in time, using the limits of the panorama and the number of frames. This normalised space can then be quantised to form a descriptor describing the trajectory.



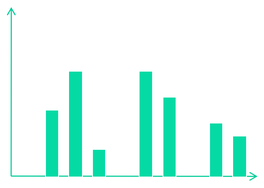
Colour Descriptor



To expand the system and use the colour information on the objects. This is done by mapping points in each cell of the quantised space to a meanshift segmented version of the original frame that each point came from. This allows for us to be able to compensate for the different variance of colour that occur across the object.



Improved Retrieval



To reduce dimensionality of the descriptor we perform principal component analysis. This reduces the trajectory descriptor and colour descriptor independently to allow faster online search performance.

Semantic Sketch Interface

We provide an intuitive interface for sketch retrieval, allowing users to select the class of object they are searching for a colour palette for retrieval.

The system allows the user to draw a free hand arrow of the trajectory, that is based on a survey conducted in research from [4]. Results are displayed in a standard video search layout.

Video Search

To retrieve relevant videos we use the descriptors of the videos generated previously. We create a descriptor from the query by projecting the query object depiction along the trajectory line. This space is then normalised according to the region of the canvas space used within the query. We then follow the steps for Spatial-Temporal descriptor creation and the colour descriptor expansion.

When ranking the dataset we use a modified version of histogram intersection. This approach is modified to work over the different classes in the video dataset.

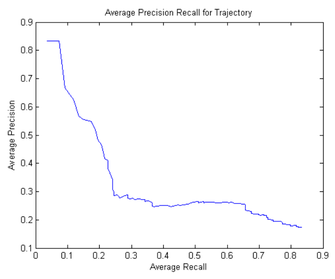
$$d(H^Q, H^I) = \sum_{c=1}^{n} \frac{\sum_{i=1}^{n} min(H^Q(c,i), H^I(c,i))}{min(\sum H^Q, \sum H^I)}$$

At current all items are searched linearly in the dataset, using histogram intersection, within the small dataset evaluated on this style of search is acceptable in larger datasets, clustering approaches would be used to improve retrieval performance.

Evaluation

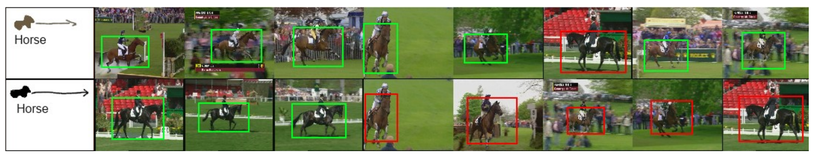
We demonstrate the results in two ways, first using trajectory results alone then with the combination of trajectory and colour. This work was evaluated over a subset of the TSF dataset, this dataset has been used to evaluate two sketch based retrieval systems in recent years. This dataset is composed of horse and snowboarding footage, and is based on the original VideoQ dataset[6]. At current there is no dataset available for semantic sketch based video retrieval, therefore we marked up the dataset for these intermediate results.

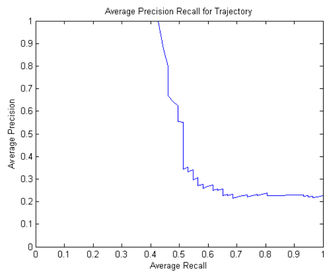
Annotated Class + Trajectory



The queries demonstrated above for trajectory and annotated class, these results have a high precision over the first two results, this result is demonstrated in the precision recall graph.

Annotated Class + Trajectory + Colour





We also evaluate our system over the proposed colour expansion these results demonstrate and improved ability of colour, within the horse class, within the person class there was not enough variation in the dataset to successfully differentiate between people.

Multiple Objects Classes

We also demonstrate the ability of our system to be able to identify multiple classes within the top ten results over both class, colour and trajectory we achieve three exact matches.

References

[1] R. Hu, M. Barnard, J. Collomosse. "Gradient Field Descriptor for Sketch based Retrieval and Localization". Proc. Intl. Conf. Image Processing ICIP 2010.

[2] J. Collomosse, G. McNeill, Y. Qian. "Storyboard Sketches for Video Retrieval". ICCV 2009.

[3] R. Hu, J. Collomosse. "Motion-sketch based video retrieval using a Trellis Levenshtein Distance". Proc. Intl. Conf. Pattern Recognition ICPR (Oral) 2010.

[4] J. Collomosse, G. McNeill, L. Watts. "Free-hand sketch grouping for video retrieval". Proc. Intl. Conf. Pattern Recognition ICPR (Poster). 2008.

[5] J. Shotton, M. Johnson, J. Cipolla. "Semantic texton forests for image categorization and segmentation". Proc. Intl. Conf. Pattern Recognition ICPR . 2008.

[6] S. Chang, H. Meng, D. Zhong. "VideoQ : An Automated Content Based Video Search System Using Visual Cues". Proc. Intl. Conf. On Multimedia ICM. 1997.