Title Title Title Title Title

AUTHOR AND INSTITUTE

Motivation

Here is the motivation.

Contributions

Here is our contributions.

Method

The cost is interpreted as a directed acyclic graph with weights on the nodes and edges. The nodes encode candidate positions, and the edges the transition costs between candidates. Additional edges (dashed) allow occlusion transitions which skip frames.

The optimal track is found with a modification of Dijkstra's shortest path search. The search was speed up by lower bounding the cost, and lazily evaluating the accurate cost only where necessary to find the global optimum.

Source Code

The source code and compiled executables with an interactive interface are available at http://math.ecnu.edu.cn

Results

Between one and three user clicks were needed to achieve accurate tracking for the head sequence. Note the correct handling of the occluded ear, which required only a single click.

The eye of the running giraffe required eight user interactions, of which three marked occlusions.

Background Model

We incorporate a background model, such that a click tells us not only 'this is how the landmark looks like', but also 'this is how the landmark does *not* look like' for all other patches in that frame.

A Future Direction

We incorporated a background model, where a click informs us not only that 'this is how the patch looks like', but also for the rest of the frame, 'this is how the patch does not look like'.

Can we also *efficiently* use a background tracks model, allowing us to reason, 'this would be a good track, but part of it can be better explained by tracking another point'.

Comments

Between one and three user clicks were needed to achieve accurate tracking for the head sequence. Note the correct handling of the occluded ear, which required only a single click.

The eye of the running giraffe required eight user interactions, of which three marked occlusions.

Speed

Speed is achieved by preprocessing the video with an adaptive filter bank as in [1, 2]. Preprocessing was sped up significantly, but is still slower than realtime.

This encodes the video into 16 byte per pixel feature vectors. We implemented an efficient search for similar patches using the SIMD hardware of modern processors, and only evaluate the cost on these candidate patches. (Typically 200 patches per frame). The Graph-Structure focuses the evaluations on the most important areas, and makes candidate search and reasoning highly efficient, such that the system runs at interactive speed.

Note that the preprocessing is not specific to the interestpoints tracked later. A single preprocessed video can therefore be used in many annotation sessions.

References

References

[1] G. H. Golub, History of numerical linear algebra: A personal view, Stanford, 2007.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The 4th Editon, The Johns Hopkins University Press, Baltimore, MD, 2013.