10. 6. 2008

School of Computer and Communication Sciences
COMPUTER VISION LAB (CVLab)

EPFL IC CVLab
Dr. Vincent LEPETIT
BC 308 (Building BC)
Station 14
CH -1015 Lausanne

Tel. +41 21 693 75 19
Fax +41 21 693 75 20
E-mail : vincent.lepetit@epfl.ch
Web site : http://cvlab.epfl.ch/~lepetit

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, June 1st 2008

TO WHOM IT MAY CONCERN

REPORT ON THE DISSERTATION THESIS OF KAREL ZIMMERMANN

The Dissertation Thesis of Karel Zimmermann tackles the important problem of object tracking in video sequences. The general approach falls into the direct approach category, in which the target object is tracked by matching a template representing the object against the image.

The author first divides the many existing approaches to tracking into two categories, optimization-based tracking and regression-based tracking. He recalls that the two are related but that the regression-based methods are less prone to "loss-of-lock" that is to tracking failure. The thesis work is firmly on the regression-based side, and extends what are called *learned linear predictors* (LLiP). These are linear regressors that provide the object motion given the image intensities at the previous locations, and they are learned from a training set made of motion and intensities couples.

The first contribution is a rigorous method to learn an optimal sequence of LLiP (Chapter 4). It is first noted that a single predictor has a limited power, in that it is only able to provide the object motion over a limited range, and up to a limited accuracy. It means that in practice the predicted position does not correspond exactly to the correct object position, but only falls into what is called an "uncertainty region". Also predictors can have different complexities, directly related to the number of image locations used to predict the motion.

The general idea is therefore to build a sequence (called SLLiP) of LLiPs such that the range of a predictor in the sequence corresponds to the uncertainty region of the previous predictor. By successively applying the predictors, one can retrieve an accurate position for the object, even for large motions. The user can choose the level of accuracy. However, to be optimal, the sequence must minimize the total complexity of the sequence, and a first method (called minimax algorithm) based on the Dijkstra algorithm is provided to find the optimal sequence. A second method for "anytime learning" (meaning that the solution is continuously improved) based on the Branch-and-Bound algorithm is also provided. Finally a greedy algorithm is given to search for image locations that minimize the prediction errors.

Chapter 5 shows how to use SLLiPs to track a 3D object with known geometry. The 2D translations of several object parts can be obtained using one SLLiP for each part, then the object pose can be robustly estimated using RANSAC. The part locations are automatically selected to optimize a weighted some of a coverage measure and a complexity measure. The resulting tracking method is called NoSLLiP. The use of RANSAC allows to handle partial occlusions and to tolerate failures of some of the local SLLiPs, making NoSLLiP particularly robust. The formulas to compute the optimal number of RANSAC iterations and the number of tracked parts given the time the user is willing to dedicate to tracking are derived.

Chapter 6 presents several experimental evaluations of the previously introduced methods:
First, different objects of different natures (a mousepad, a t-shirt, the rear of a car…) are tracked using NoSLLiP under various and difficult conditions (outdoor or blurred sequences, or poorly textured objects). The two algorithms proposed to build the SLLiPs, the minimax algorithm and the anytime learning one, are compared using real sequences for which ground truth is known. It is found that the minimax algorithm yields often more robust tracking at a much higher computational cost.
The SLLiP method alone is then compared to other possible methods, namely a simple method based on SIFT, the Lucas-Kanade tracker, and the Jurie and Dhome method. The latter one is mostly a single LLiP, that is, no sequence is used. The method proposed in the thesis outperforms the other methods and achieves remarkable robustness, with a computational cost actually lower than the Jurie and Dhome method, probably thanks to the multiple optimizations introduced at the different steps of the method.
Finally several empirical evaluations are presented: the distribution of the LLiP complexities depending on their place in the sequence; the total complexity of SLLiPs for different patches, the influence of the parameters over the minimax and anytime learning algorithms, and the efficiency of the support selection greedy algorithm.

Then come what is probably the most interesting part of the thesis (Chapter 7), in which the previous method is improved to adapt to appearance changes of the tracked object. First a new regressor suitable for varying appearances is defined as a weighted sum of linear regressors, where the weights are the appearance parameters. The appearance parameters are taken to be linear in the object image intensities, and so a linear regressor that maps the object image and the appearance parameters exists. However, the nature of the appearance parameters are not explicitly chosen, and let to be decided automatically by an optimization process. A matrix with random coefficients is first used to initialize the regressor from the object image to the appearance parameters. Given this appearance regressor and a training set, the regressor for motion estimation can be estimated in closed-form. In return, given the motion regressor matrices, the appearance regressor can be estimated in closed-form. These two steps are iterated as an alternate optimization. It is shown experimentally that this process converges, and that the retrieved appearance parameters seem to capture meaningful properties of the appearance changes.

Finally, a method to automatically update over a video sequence the optimal time spent in the SLLiP and its initial range is provided. It uses a Kalman filter in a quite unusual way: Such a filter is used to predict the covariance of the probability distribution of success depending on the time and range. From this distribution, the optimal time and ranges can be computed, given an upper bound chosen by the user on the probability of failure. It is shown experimentally that this updating procedure can adapt to the nature of the object motion, yielding to more accuracy when the object moves slowly, and more robustness during saccadic motions.

Relevance to current needs of the scientific community
Tracking is one of the main topics of Computer Vision and has many practical applications. The thesis aims to bring more robustness and accuracy to these applications, while keeping the computational aspect in mind. The relevance of this work makes therefore no doubt.

Fulfillment of the main objectives
The proposed methods are demonstrated on difficult video sequences in which different objects, and the results are convincing and impressive. They are also shown to outperform existing methods using quantitative evaluations.

Methods
I was impressed by the rigor present throughout the thesis. Every statement is proven theoretically, or when a proof was not possible, it is clearly discussed and justified empirically. Each proposed method is demonstrated empirically on difficult video sequences, and evaluated using ground truth data.

Main results and contributions
The thesis brings many contributions, including a rigorous method to build optimal sequences of linear predictor for object tracking, a not less rigorous extension to appearance changing objects, and a method for automatically adapting the parameters during tracking. This work has been submitted and accepted to the Transactions on Pattern Analysis and Machine Intelligence, one of the most prestigious journals in Computer Vision, which shows the importance of the contributions.

Importance for the further development of science
Many of the introduced methods, in particular the SLLiPs and the LLiP with appearance changes, can be used as fundamental bricks into future methods, and therefore constitute a great help for further developments.

Creativity
I particularly appreciated the originality and elegance of the method for tracking under motion and appearance changes, and the adaptive parameter optimization method. This proves the scientific creativity of the author.

The author of the thesis proved to have an ability to perform research and to achieve scientific results. I do recommend the thesis for presentation with the aim of receiving the Degree of Ph.D.

Dr. Vincent Lepetit