Generic models for adaptive robust feature extraction in video

Dimitri Van Cauwelaert

Supervisor(s): Wilfried Philips, Peter Veelaert

Abstract— **In computer vision applications, it is important to extract and describe robust and characteristic features from objects in images. Many recognition and tracking algorithms rely on this information. One of the main problems with current state-of-the-art algorithms, such as SIFT [1] , GLOH [2] , SURF [3], is that the retrieval of features at the border of an object is sensitive to changes in the background. This work presents a possible solution using background/foreground segmentation**

Keywords— **robust anisotropic feature descriptor, SIFT, GLOH, SURF**

I. Introduction

The recent advances in computer hardware have led to a significant reduction in execution time of computational demanding algorithms. One of the fields of research highly benefiting of these achievements is computer vision which strongly relies on the availability of computational power and memory to process the typically large amounts of data available in video sequences. This research focuses on finding generic algorithms which are able to extract features from video data in a robust way.

Features are objects in images carrying similar properties, eg. regions of a certain color, corners, lines, parabola, ... Such features can than be used in many computer vision techniques, including object or scene recognition, solving for 3D structure from multiple images, stereo correspondence and motion tracking. One of the main problems with many feature extraction algorithms is that features located at the border of an object are not very robust since the varying background significantly influences their mathematical description. Having more robust features significantly improves the recognition algorithms.

D. Van Cauwelaert is with the Electronics Department of the Faculty of Applied Engineering, University College Hogeschool Gent (Hogent), Gent, Belgium and the department of Telecommunication and Information technology (IPI/TELIN/IBBT) of the Faculty of Engineering Science, Ghent University (Ugent), Gent, Belgium. E-mail: dimitri.vancauwelaert@hogent.be .

II. Scale invariant Feature Transform

The Scale Invariant Feature Transform (SIFT) [1] extracts point features from the image and creates a high dimensional description vector (descriptor) for the local image content. The description of the feature is then used to look for matching features in a model image and a test image. The extraction and description step are both invariant to rotational, scaling and illumination effects, as well as the addition of noise. The technique also works for a substantial range of affine distortion and change in viewpoint.

The features are strong extremal points in a Difference of Gaussians pyramid (scaling invariance). After extraction, a relative coordinate system (rotational invariance) is assigned to each feature, based on local gradient information extracted at the scale at which the feature point is found. The descriptor is then computed based on local gradient information aligned with the new coordinate system.

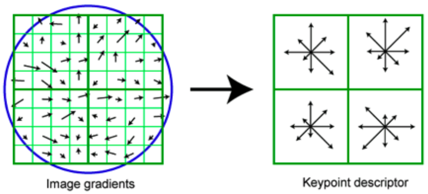

Figure 1. Construction of the SIFT descriptor. The circle indicates its isotropic nature.

III. Limitations of the current descriptors

A main disadvantage of the SIFT-descriptor and other descriptors used in state-of-the-art algorithms such as GLOH [2] (a variant of SIFT, which is more distinctive for the same dimensionality) and SURF [3] (a recent technique based on a Hessian matrix-based measure for the detector and using a distribution based descriptor) is the use of isotropic local information to describe feature points. This limits the robustness of features located at the border of an object against a varying background, since background changes can significantly influence the descriptor. It also increases the euclidian distance between corresponding features and by doing so, weakens the robustness of these features. This is a major disadvantage since a lot of features can be expected to be detected at the corners of an object. Improving the robustness of these border features might significantly improve current recognition and tracking algorithms for video data.

IV. Improving the robustness of the SIFT descriptor

This doctoral research focuses on designing anisotropic descriptors which limit the description to the foreground of the object in order to improve the robustness of the algorithm. In order to accomplish this, we will divide the region around the feature in sections and indicate whether this section belongs to the foreground of the image or not. This basically brings the first step down to a foreground/background segmentation. When designing the descriptor more weight must be given to components of the foreground, so that the influence of the background is less prominent. In a further stage, color information can be included in the descriptor to make it more discriminative.

V. Performance evaluation

Recall vs. 1-precision graphs are used to obtain a quantitative indication of the performance of the algorithm. Recall is defined as the amount of correct matches to the number of correspondences (maximum number of correct matches) and 1-precision as the number of false matches to the total number of matches. Using model-test image pairs with a known transformation (eg. an artificially applied homography), we know exactly where every match should be. Using a measure on the correspondency between the observed and the expected match, eg. a threshold on the location mismatch, we retrieve the number of correct and false matches.

VI. Conclusion

Many modern feature descriptors suffer from limitations, as they use isotropic local information for video sequences. We propose to use an anisotropic descriptor based on a foreground/background segmentation to improve the robustness of features.

References

[1] DG Lowe, "Distinctive image features from scale-invariant keypoints," *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 60, no. 2, pp. 91–110, NOV 2004.

[2] K Mikolajczyk and C Schmid, "A performance evaluation of local descriptors," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 27, no. 10, pp. 1615–1630, OCT 2005.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "SURF: Speeded up robust features," *COMPUTER VISION - ECCV 2006 , PT 1, PROCEEDINGS*, vol. 3951, no. Part 1, pp. 404–417, 2006.