

A General Framework for Approximate Nearest Subspace Search

Ronen Basri*	Tal Hassner**	Lihi Zelnik-Manor
Weizmann Institute of Science Rehovot, 76100, Israel	The Open University of Israel Raanana, 43107, Israel	Technion - Israel Inst. of Technology Haifa, 32000, Israel

Abstract. Subspaces offer convenient means of representing information in many Pattern Recognition, Machine Vision, and Statistical Learning applications. Contrary to the growing popularity of subspace representations, the problem of efficiently searching through large subspace databases has received little attention in the past. In this paper we present a general solution to the Approximate Nearest Subspace search problem. Our solution uniformly handles cases where both query and database elements may differ in dimensionality, where the database contains subspaces of different dimensions, and where the queries themselves may be subspaces. To this end we present a simple mapping from subspaces to points, thus reducing the problem to the well studied Approximate Nearest Neighbor problem on points. We provide theoretical proofs of correctness and error bounds of our construction and demonstrate its capabilities on synthetic and real data. Our experiments indicate that an approximate nearest subspace can be located significantly faster than the nearest subspace, with little loss of accuracy.

1 Introduction

Although the use of subspace representations has increased considerably over the years, one fundamental question related to their use has so far received little attention: How does one efficiently search through a database of subspaces? There are two main reasons why we believe this question to be paramount. The first is the demonstrated utility of subspaces as a (sometimes only) means for conveniently representing varying information. The second is the ever-growing volume of information routinely collected and searched through as part of Computer Vision and Pattern Recognition systems, information often represented by subspaces. The goal of this paper is to address this question by presenting a general framework for efficient subspace search.

In a recent paper [4] Basri et al. have presented a method for sub-linear *approximate nearest subspace* (ANS) search. Their solution, however, was limited to a particular scenario where the queries are high dimensional *points*. They thus ignore cases where the query itself may be a subspace. Moreover, their method cannot handle databases of subspaces with different dimensions, which may be the case if variable amounts of information are available when the database is produced or when object representations allow for different degrees of freedom. In this paper we extend their work and provide the following contributions.

Author names are ordered alphabetically due to equal contribution.

*Part of the work was done while at the Toyota Technical Institute.

**Part of the work was done while at the Weizmann Institute of Science.