TITLE: COMPRESSED HISTOGRAM OF GRADEIENTS A LOW BITRATE DESCRIPTOR

ABSTRACT: Establishing visual correspondences is an essential component of many computer vision problems, which is often done with local feature-descriptors. Transmission and storage of these descriptors are of critical importance in the context of mobile visual search applications. We propose a framework for computing low bitrate feature descriptors with a 20× reduction in bit rate compared to state-of-the-art descriptors. The framework offers low complexity and has significant speed-up in the matching stage. We show how to efficiently compute distances between descriptors in the compressed domain eliminating the need for decoding. We perform a comprehensive performance comparison with SIFT, SURF, BRIEF, MPEG-7 image signatures and other low bit-rate descriptors and show that our proposed chog descriptor outperforms existing schemes significantly over a wide range of bitrates. We implement the descriptor in a mobile image retrieval system and for a database of 1 million CD, DVD and book covers, we achieve 96% retrieval accuracy using only 4 KB of data per query image.

TITLE: ROBUST REAL-TIME FACE DETECTION

ABSTRACT: This paper describes a face detection framework that is capable of processing images extremely rapidly while achieving high detection rates. There are three key contributions. The first is the introduction of a new image representation called the "Integral Image" which allows the features used by our detector to be computed very quickly. The second is a simple and efficient classifier which is built using the adaboost learning algorithm (Freund and Schapire, 1995) to select a small number of critical visual features from a very large set of potential features. The third contribution is a method for combining classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions. A set of experiments in the domain of

face detection is presented. The system yields face detection performance comparable to the best previous systems (Sung and Poggio, 1998; Rowley et al., 1998; Schneiderman and Kanade, 2000; Roth et al., 2000). Implemented on a conventional desktop, face detection proceeds at 15 frames per second.

TRACKING LEARNING DETECTION

ABSTRACT: This paper investigates long-term tracking of unknown objects in a video stream. The object is defined by its location and extent in a single frame. In every frame that follows, the task is to determine the object's location and extent or indicate that the object is not present. We propose a novel tracking framework (TLD) that explicitly decomposes the long-term tracking task into tracking, learning, and detection. The tracker follows the object from frame to frame. The detector localizes all appearances that have been observed so far and corrects the tracker if necessary. The learning estimates the detector's errors and updates it to avoid these errors in the future. We study how to identify the detector's errors and learn from them. We develop a novel learning method (P-N learning) which estimates the errors by a pair of “experts”: (1) P-expert estimates missed detections, and (2) N-expert estimates false alarms. The learning process is modeled as a discrete dynamical system and the conditions under which the learning guarantees improvement are found. We describe our real-time implementation of the TLD framework and the P-N learning. We carry out an extensive quantitative evaluation which shows a significant improvement over state-of-the-art approaches.

TITLE: FACE RECOGNITION USING EIGEN FACES

ABSTRACT: An approach to the detection and identification of human faces is presented, and a working, near-real-time face recognition system which tracks a subject's head and then recognizes the person by comparing characteristics of the face to those of known individuals is described. This approach treats face recognition as a two-dimensional recognition problem, taking advantage of the fact that faces

are normally upright and thus may be described by a small set of 2-D characteristic views. Face images are projected onto a feature space ('face space') that best encodes the variation among known face images. The face space is defined by the 'eigenfaces', which are the eigenvectors of the set of faces; they do not necessarily correspond to isolated features such as eyes, ears, and noses.

TITLE: EIGEN VS FISHERFACES: RECOGNITION USING CLASS SPECIFIC LINEAR PROJECTION

ABSTRACT: We develop a face recognition algorithm which is insensitive to large variation in lighting direction and facial expression. Taking a pattern classification approach, we consider each pixel in an image as a coordinate in a high-dimensional space. We take advantage of the observation that the images of a particular face, under varying illumination but fixed pose, lie in a 3D linear subspace of the high dimensional image space if the face is a Lambertian surface without shadowing. However, since faces are not truly Lambertian surfaces and do indeed produce self-shadowing, images will deviate from this linear subspace. Rather than explicitly modeling this deviation, we linearly project the image into a subspace in a manner which discounts those regions of the face with large deviation. Our projection method is based on Fisher's Linear Discriminant and produces well separated classes in a low-dimensional subspace, even under severe variation in lighting and facial expressions. The Eigenface technique, another method based on linearly projecting the image space to a low dimensional subspace, has similar computational requirements. Yet, extensive experimental results demonstrate that the proposed "Fisherface" method has error rates that are lower than those of the Eigenface technique for tests on the Harvard and Yale Face Databases.

TITLE: SHAPE MATCHING AND OBJECT RECONITION USING SHAPE CONTEXTS

ABSTRACT: We present a novel approach to measuring similarity between shapes and exploit it for object recognition. In our framework, the measurement of similarity is preceded by: (1) solving for correspondences between points on the two shapes; (2) using the correspondences to estimate an aligning transform. In order to solve the correspondence problem, we attach a descriptor, the shape context, to each point. The shape context at a reference point captures the distribution of the remaining points relative to it, thus offering a globally discriminative characterization. Corresponding points on two similar shapes will have similar shape contexts, enabling us to solve for correspondences as an optimal assignment problem. Given the point correspondences, we estimate the transformation that best aligns the two shapes; regularized thin-plate splines provide a flexible class of transformation maps for this purpose. The dissimilarity between the two shapes is computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. We treat recognition in a nearest-neighbor classification framework as the problem of finding the stored prototype shape that is maximally similar to that in the image. Results are presented for silhouettes, trademarks, handwritten digits, and the COIL data set.