

Dimensionality Reduction Through LDA and Bag-of-Features Applied to Image Retrieval

Ricardo E. Gonzalez Valenzuela & Helio Pedrini

Institute of Computing, University of Campinas, Brazil, 13083-852

William Robson Schwartz

Department of Computer Science, Universidade Federal de Minas Gerais, Brazil, 31270-010

ABSTRACT: Content-based image retrieval applications represent a relevant task in computer vision field. Many researches have been conducted to improve the retrieval process by means of discriminative descriptors, such as Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF), which describe high dimensional feature vectors. Problems involving high dimensional feature vectors usually require high computational time to process the data, high storage requirements to keep relevant information, and may suffer with low accuracy depending on the noise contained in the data. Most of the solutions addressing these problems propose a trade-off between accuracy, time and storage or even impose restrictions on the action of the application by compromising the distinctiveness of the feature descriptors. In this paper, we propose to apply linear dimensionality reduction kernels to reduce the dimensions of SIFT and SURF feature vectors and employ bag-of-features to create global features to maintain or even enhance the accuracy of the retrieval, while spending less storage and computational time requirements. The experiments compare the results achieved by the reduced feature vectors to the results obtained by the original features, demonstrating to gain in accuracy while reducing computational time and storage.

1 INTRODUCTION

Image retrieval is an important process that aims at successfully finding a desired image, based on its visual properties, in large data sets. Content-Based Image Retrieval (CBIR) (Maloo, Thepade, & Kekre 2010, Nilesingh & Koli 2010) applies computer vision techniques to analyze the contents of the images in the searching task and it is useful in various applications, such as medical diagnosis, art collections, photograph archives, geographical information systems, digital libraries, among others (Lew, Sebe, Djeraba, & Jain 2006).

Several different methods have been proposed to achieve the CBIR goals. Many of them employ SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006) descriptors and complement them with other image characteristics, such as texture and color. However, SIFT and SURF descriptors, as many others reported for high dimensional feature vectors, have to deal with computationally intensive processes and massive amounts of memory storage, that there have been experimented some compressions (Chandrasekhar, Makar, Takacs, Chen, Tsai, Cheung, Grzeszczuk, Reznik, & Girod 2010). Some

CBIR methods also attempt to reduce the SIFT descriptor by, for instance, considering rotation invariance (Williams & Ledwich 2004), which may limit the descriptor discriminative power preventing its successful application in the CBIR task. Other works have been proposed to reduce the feature vectors by applying dimensionality reduction techniques, where it has been demonstrated that the reduced feature vectors (to half or less than the original size) lead to a similar performance when compared to the original feature vectors (Gonzalez, Schwartz, & Pedrini 2012, Hua, Brown, & Winder 2007); furthermore, discriminant eigenfeatures have been used for image retrieval applications (Swets & Weng 1996).

Although the application of dimensionality reduction techniques is promising, the required computational time to calculate the matches between every image through the reduced feature vectors is still considerable. Therefore, the use of bag-of-features becomes a suitable solution since it preserves the individual influence of the feature vectors while adding generalization (Csurka, Dance, F., Willamowski, & Bray 2004) and has demonstrated to achieve good performance (Lazebnik, Schmid, & Ponce 2006, Zhang, Tian, Hua, Huang, & Li 2009, Jurie & Triggs 2005).

This paper aims at applying linear dimensionality reduction techniques and bag-of-features to reduce the descriptor dimensions and compute a global description for the images in order to improve accuracy of the retrieval process, while reducing the computational time and disk storage.

To validate our assumptions, we conduct a number of experiments. SIFT and SURF feature vectors are reduced to different dimensions (6, 12, 20 and 32), then this reduced feature vectors are quantized to be represented through several histogram sizes (50, 250, 500 and 1000 key feature sizes). After that, the image retrieval process is performed and some experiments are performed to compare the accuracy, computational time and storage requirements achieved through the original feature vectors and reduced feature vectors.

The remainder of the paper is organized as follows. Section 2 briefly presents SIFT and SURF descriptors, reviews some linear dimensionality reduction techniques and describes relevant aspects of the bag-of-feature process. The methodology used to validate our results is presented in Section 3. Section 4 presents experimental results obtained from an image retrieval problem using the Corel Dataset (Corel Dataset 2013), where accuracy, time and storage are compared. Finally, Section 5 concludes our work.

2 TECHNICAL BACKGROUND

In this section, we briefly review SIFT and SURF descriptors, as well as dimensionality reduction techniques and bag-of-features.

2.1 Interest Point Descriptors

Interest point descriptors such as SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006) are well known by the fact that they provide good performance, discriminability to describe images, invariance under scale, rotation, translation, and partial invariance under changes in illumination (Lowe 2004).

Both descriptor algorithms, SIFT and SURF, function in a similar way. Firstly, the interest points are located and the scale of each interest point is detected in this phase. Then, the feature vectors are described: for the SIFT descriptor, local image gradients are computed in the region around each interest point at the selected scale and weighted by a Gaussian window, whereas for the SURF descriptor, it is built a grid consisting of 4×4 square sub-regions centered at the interest point, a wavelet transform is computed over each sub-region and the responses (dx , $|dx|$, dy , $|dy|$) are stored in 2×2 subdivisions.

Accordingly, a SIFT feature vector is composed of a 128-dimensional vector (8 orientation bins for each 4×4 location bin), and a SURF feature vector of a 64-dimensional vector (4 region responses for each 4×4 location bin).

2.2 Dimensionality Reduction Techniques

Dimensionality reduction techniques have been proposed to achieve a compact representation of the data focusing on aspects such as discriminability. Among such techniques, it is possible to use linear dimensionality reduction techniques since they allow to create a kernel to project the desired features onto it.

Some of these techniques include Principal Component Analysis (PCA) (Pearson 1901, Jolliffe 2002), which calculates eigenvalues and eigenvectors of the data to create a kernel, then a covariance matrix contains the first n (where n represents the number of desired dimensions) eigenvectors with highest eigenvalues are selected; Independent Component Analysis (ICA) (Hyvärinen 1999, Hyvärinen & Oja 2000) that is frequently used to separate and reduce mixed information that comes from different sources; Random Projections (RP) (Bingham & Mannila 2001, Sulic, Perš, Kristan, & Kovacic 2010) that generates a kernel consisting of orthonormalized vectors and has been proven to perform efficiently; Partial Least Squares (PLS) (Wold, Ruhe, Wold, & Dunn 1984, Rosipal & Krämer 2006) which assumes that the observed data generated depends on a small number of non-observed data (which, in this case, will indicate the corresponding classes); and Linear Discriminant Analysis (LDA) (Balakrishnama & Ganapathiraju 1998) that classifies two or more classes while reducing dimensionality.

2.3 Bag-of-Features

Bag-of-features (Csurka, Dance, F., Willamowski, & Bray 2004, Lazebnik, Schmid, & Ponce 2006) is a method that characterizes the distribution of feature vectors present in the image using a histogram, that is, a sparse vector of occurrence counts of local image features.

To represent an image through bag-of-features, a set of keypoints must initially be detected. After feature detection, various local patches are used to represent the image as numerical vectors. Examples of descriptors that could be used are SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006), as described above. Another step to represent images using bag-of-features is the codebook (dictionary) generation, where a codeword is considered as a representative of similar patches. A common technique for generating the codebook is to apply the k -means clustering over all the vectors, such that the codewords are defined as the centers of the clusters. The codebook size corresponds to the number of clusters. Then, each patch in the image is mapped to a specific codeword and the image can be represented by the histogram of the codewords.

3 METHODOLOGY

The main purpose of this work is to maintain an approximate accuracy of the original descriptors while reducing the computational time and storage requirements by using dimensionality reduction techniques and bag-of-features.

The main steps of the proposed methodology are illustrated in Figure 1. Each stage is described in the following sections.

3.1 Training Stage

In order to reduce a feature vector, a projection kernel is usually created. This process may demand a high computational time, however, since every feature vector needs to be projected on the same kernel, this process must be executed on a training stage. To perform this phase, 75 images from each group in the Corel Dataset (Corel Dataset 2013) were selected. Feature vectors are densely detected and described by using SIFT or SURF algorithms over the 750 images.

Then, a dimensionality reduction technique (such as PCA, ICA, RP, PLS or LDA) is applied to the 250 images that are used to learn the projection kernel. If the projection kernel is learned with PLS or LDA, then each feature vector is associated with the image class identifier. Once the projection kernel is already computed, the feature vectors corresponding to the remaining 500 images are projected onto the kernel to obtain the reduced feature vectors. The reduced feature vectors of the 500 images are used to create the bag-of-features to obtain the dictionary through the k -means algorithm.

3.2 Testing Stage

Two elements are needed to perform the testing stage: the projection kernel and the bag-of-features.

The feature vectors for the 250 images belonging to this phase are densely detected and computed. Then, these features are projected onto the projection kernel to produce the 250 files of reduced feature vectors. After that, a histogram of n bins is created to have a global descriptor for each image, where n refers to the number of key features in the bag-of-features. Each bin is initialized with 0. To fill this histogram, it is calculated the Euclidean distance between each reduced feature vector and the key features in the bag-of-features.

A reduced feature vector belongs to some key feature when its Euclidean distance is the smallest one, then the bin corresponding to this key feature is increased by one. Subsequently, the histogram is normalized to one. This histogram will represent a unique feature vector describing one image. Finally, the number of similar images can be efficiently computed between these global descriptors.

3.3 Evaluation Metrics

During the testing stage, the histograms indicating the number of reduced feature vectors belonging to each key feature are computed for every image in the test group (25 images from each of the ten Corel Dataset groups). Then, each histogram is compared to the other 249 histograms, and a ranking of the first 24 images with histograms with which it achieves the maximum similarity measure is maintained, given by Equation (1), where H_x and H_y represent two histograms, n denotes the number of key features, $k_i^{H_x}$ and $k_i^{H_y}$ represent the respective number contained by the i -th bin of their histograms.

$$\text{sm}(H_x, H_y) = \sum_{i=1}^n \min(k_i^{H_x}, k_i^{H_y}) \quad (1)$$

The retrieved precision value is given by Equation (2). An image belonging to the class A will achieve 100% of retrieved precision if its top 24 matched images are composed of only images belonging to its class

$$\text{Retrieved Precision} = \frac{\text{Correct Retrieved Images}}{\text{Total Images to Retrieve}} \quad (2)$$

4 EXPERIMENTAL RESULTS

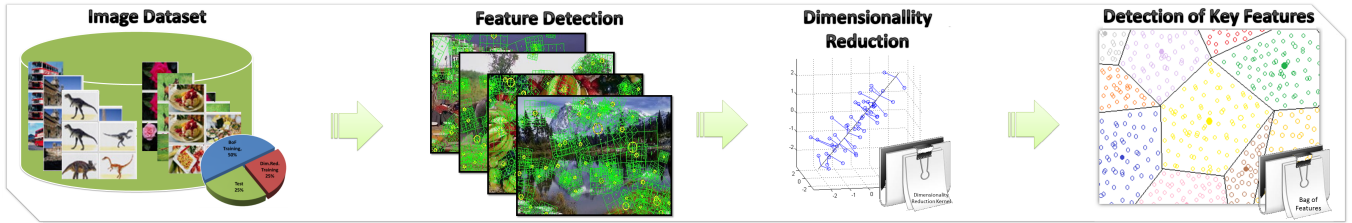
To validate our methodology, experiments with an image retrieval application over the Corel Dataset (Corel Dataset 2013) are conducted. This data set is composed of 10 groups of images, each one containing one hundred images. Dense SIFT and Dense SURF algorithms are executed to describe each image in the data set. After that, dimensionality reduction is applied to the descriptors.

The following sections compare the original and reduced feature descriptor in terms of accuracy, computational time and storage requirements.

4.1 Precision Evaluation

This first experiment compares the retrieved precision achieved by SIFT and SURF descriptors to the correspondent precision achieved by the reduced descriptors. PCA, PLS, ICA, LDA and RP techniques are used to reduce the feature vectors to 6, 12, 20 and 32 dimensions. Furthermore, the generation of several bag-of-features, consisting of 50, 250, 500 and 1000 key features, is performed.

In Figure 2, it is shown that the performance obtained by executing the experiments with SIFT (128 dimensions) and SURF (64 dimensions) and their respective 6 dimensional reduced feature vectors. It is noticed that even performing with these few dimensions, most of the reduced feature vectors achieved a similar precision in relation to the original feature vector (with a difference of at most 4% and 6% from



(a) Training Stage



(b) Testing Stage

Figure 1: Training and testing stages. (a) in the training stage, 25% images are separated from each group to create their feature vectors and then obtain a projection kernel, after that other 50% images from each group are used to create new reduced feature vectors to compute the bag-of-features; (b) in the testing stage, feature vectors are created and described over the last 25% images, then projected over the kernel to obtain the reduced feature vectors and, finally, the histogram of key features is computed to be compared to others.

the original), and it is worth mentioning that the feature vectors reduced with the LDA dimensionality reduction kernel achieved even superior results than the original feature vectors.

We assume that these favorable results are achieved by every dimensionality reduction kernel due to the fact that not only a small amount of information is removed from the original vectors but also possible noise. Furthermore, PLS and LDA dimensionality reduction techniques report better performance, which can be explained by their ability of marking a feature vector as belonging to a specific class, which aids to better identify the feature vectors of different classes while projecting them onto the kernel.

Figure 3 compares the precision retrieved by SIFT and SURF to their respective LDA reduced feature vectors. It is important to notice that, in both cases, the 6 dimensional reduced feature vectors produce superior results than any other reduced feature vector.

Finally, Figure 4 shows the retrieval precision obtained by each class in the Corel Dataset by SIFT and SURF feature vectors compared to their respective LDA reduced feature vectors using 6, 12, 20 and 32 dimensions and a bag-of-features composed of 500 key features, which corresponds to an appropriate balance between accuracy, computational time and storage. It is clear that LDA 6 dimensional features vectors present either better results than their original feature vectors or approximately similar.

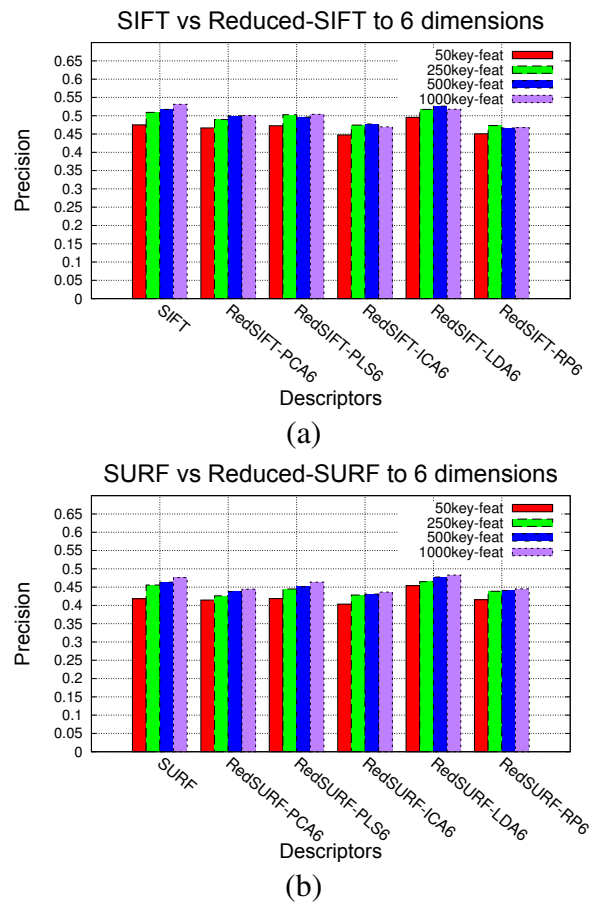


Figure 2: Comparison between the original feature vectors and the 6 dimensional reduced feature vectors using bag-of-features of 50, 250, 500 and 1000 key features (a) SIFT compared to RedSIFT-PCA, RedSIFT-PLS, RedSIFT-ICA, RedSIFT-LDA, RedSIFT-RP; (b) SURF compared to RedSURF-PCA, RedSURF-PLS, RedSURF-ICA, RedSURF-LDA, RedSURF-RP.

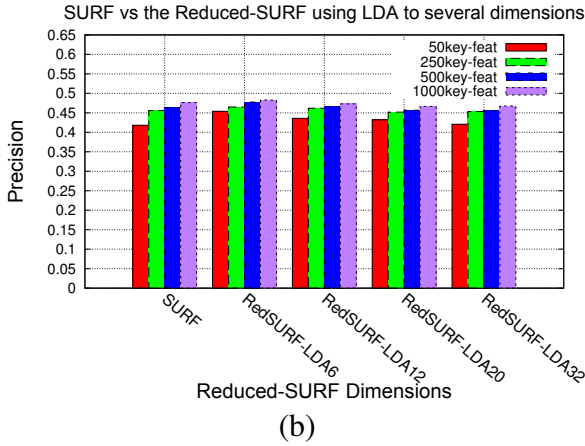
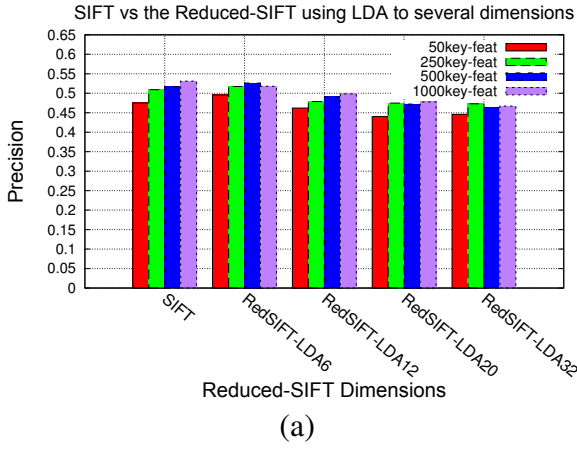


Figure 3: Comparison between the original feature vectors and LDA reduced feature vectors using 6, 12, 20 and 32 dimensions and bag-of-features of 50, 250, 500 and 1000 key features (a) SIFT compared to RedSIFT-LDA; (b) SURF compared to RedSURF-LDA.

4.2 Computational Time Evaluation

There are four phases to compute the testing stage: extracting the feature vector; reducing the original feature vector; computing the respective histogram (also called quantization phase); and calculating the retrieved precision value. The first and the last phases are the same for any feature vector or reduced feature vector, such that the reduction and quantization phases are taken into consideration to compute the difference in computational time.

The reduction phase, obviously, does not affect the SIFT or SURF descriptors; therefore, it does not represent an associated cost to them. However, as the quantization step is directly affected by the number of dimensions of the feature vectors. Figure 5 illustrates a comparison between the process performed with the original feature vector and its respective reduced feature vectors. clearly the time to compute the quantization phase is dominated by the number of dimensions of the feature vector, then the time spent in the reduction phase is totally remunerated by the time gained in the quantization phase.

4.3 Storage Evaluation

Content-based image retrieval applications usually store image descriptions to avoid computing them ev-

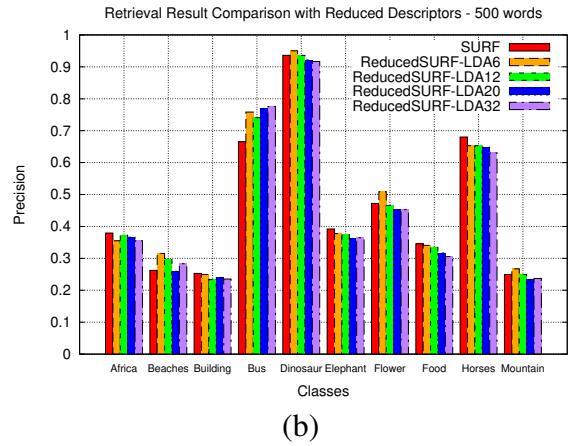
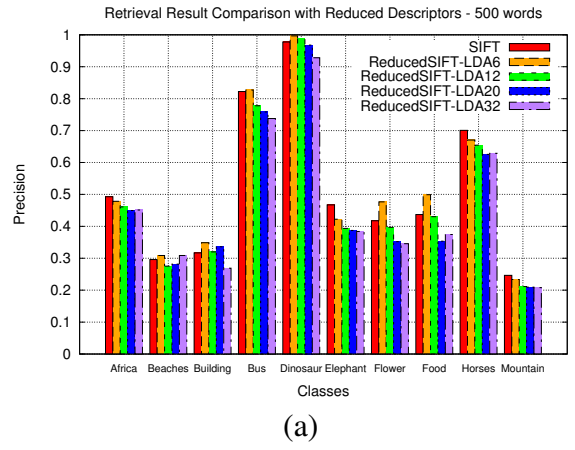


Figure 4: Detailed comparison between the original feature vectors and LDA reduced feature vectors using 6, 12, 20 and 32 dimensions and bag-of-features of 500 key features (a) SIFT compared to RedSIFT-LDA; (b) SURF compared to RedSURF-LDA.

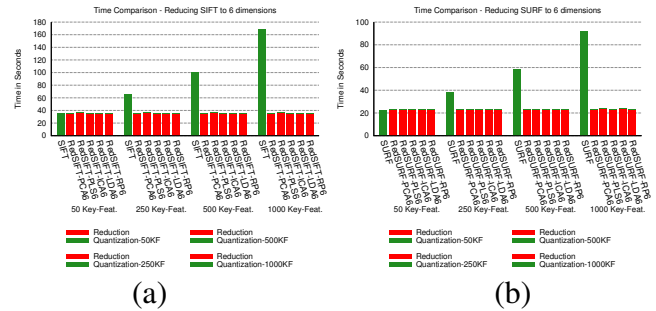


Figure 5: Computational time comparison (a) SIFT compared to RedSIFT-PCA, RedSIFT-PLS, RedSIFT-ICA, RedSIFT-LDA, RedSIFT-RP; (b) SURF compared to RedSURF-PCA, RedSURF-PLS, RedSURF-ICA, RedSURF-LDA, RedSURF-RP.

ery time. The actual features that need to be stored will be the image histograms, however, if it is desired to change the bag-of-features to compute them, then the feature vectors must also be stored.

Figure 6 shows the storage requirements needed to store a file corresponding to dense feature vectors compared to their respective reduced feature vectors using 6, 12, 20 and 32 dimensions. As the 6 dimensional feature vectors perform better than the original feature vectors, it is important to notice that they represent 5.86% and 15.93% of storage size required by the original SIFT and SURF descriptors, respectively.

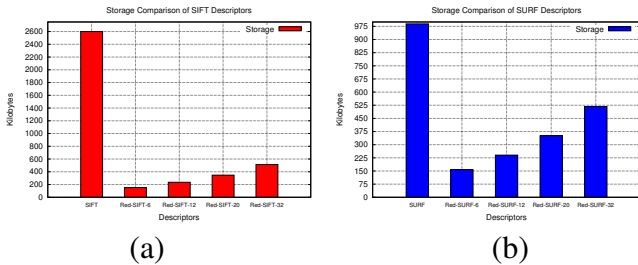


Figure 6: Storage comparison (a) storage for SIFT feature vectors compared to their respective reduced feature vectors; (b) storage for SURF feature vectors compared to their respective reduced feature vectors.

5 CONCLUSIONS

Content-based image retrieval applications usually compute image feature vectors once and employ them several times. Then, the reduction of feature vectors is desirable in terms of computational time and storage requirements. Since LDA and PLS are able to identify a feature vector as belonging to a particular class the performance they guarantee a better performance. Nevertheless, PCA, ICA and RP techniques also present an acceptable performance.

This work demonstrated the feasibility of reducing the dimensionality of descriptors, such as SIFT and SURF, while maintain a similar accuracy or even improve it when using less than 5% of the original feature vector dimensions. Computational time was not reduced only during the phase of calculating the reduced feature vectors but also when comparing the obtained histograms, as they represent an entire image through a unique vector.

6 ACKNOWLEDGMENTS

The authors are grateful to FAPESP, FAPEMIG, CNPq, and CAPES for the financial support.

REFERENCES

Balakrishnama, S. & A. Ganapathiraju (1998). Linear Discriminant Analysis - A Brief Tutorial. In *International Symposium on Information Processing*.

Bay, H., T. Tuytelaars, & L. Van Gool (2006, October). SURF: Speeded Up Robust Features. In *12th European Conference on Computer Vision*, Firenze, Italy, pp. 404–417.

Bingham, E. & H. Mannila (2001). Random Projection in Dimensionality Reduction: Applications to Image and Text Data. In *Knowledge Discovery and Data Mining*, pp. 245–250. ACM Press.

Chandrasekhar, V., M. Makar, G. Takacs, D. Chen, S. S. Tsai, N.-m. Cheung, R. Grzeszczuk, Y. Reznik, & B. Girod (2010, August). Survey of SIFT Compression Schemes. In *Second International Workshop on Mobile Multimedia Processing*, Istanbul, Turkey.

Corel Dataset (2013). <http://people.csail.mit.edu/torralba/research/bias/>.

Csurka, G., C. R. Dance, L. F. J. Willamowski, & C. Bray (2004). Visual Categorization with Bags of Keypoints. In *Workshop on Statistical Learning in Computer Vision*, pp. 1–22.

Gonzalez, R., W. R. Schwartz, & H. Pedrini (2012, August). Dimensionality Reduction Through PCA over SIFT and SURF Descriptors. In *Conference on Cybernetics Intelligent Systems*, Limerick, Ireland, pp. 58–63.

Hua, G., M. Brown, & S. Winder (2007). Discriminant Embedding for Local Image Descriptors. In *IEEE 11th International Conference on Computer Vision*, pp. 1–8.

Hyvärinen, A. (1999). Survey on Independent Component Analysis. *Neural Computing Surveys* 2, 94–128.

Hyvärinen, A. & E. Oja (2000). Independent Component Analysis: A Tutorial. *Neural Networks* 13(4-5), 411–430.

Jolliffe, I. T. (2002). *Principal Component Analysis* (Second ed.). Springer.

Jurie, F. & B. Triggs (2005, October). Creating Efficient Codebooks for Visual Recognition. In *IEEE International Conference on Computer Vision*, Volume 1, San Diego, CA, USA, pp. 604–610.

Lazebnik, S., C. Schmid, & J. Ponce (2006, June). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, New York, NY, USA, pp. 2169–2178.

Lew, M. S., N. Sebe, C. Djeraba, & R. Jain (2006, February). Content-based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications* 2(1), 1–19.

Lowe, D. G. (1999). Object Recognition from Local Scale-Invariant Features. In *IEEE International Conference on Computer Vision*, Washington, DC, USA, pp. 1150–1157. IEEE Computer Society.

Lowe, D. G. (2004, November). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110.

Maloo, A., S. D. Thepade, & H. Kekre (2010, July). CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences. *International Journal of Computer Applications* 3(12), 39–46.

Nileshsingh, V. & N. Koli (2010). An Overview of Feature Extraction Methods in CBIR. *International Journal of Computer Science and Applications* 3(2), 71–76.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2(6), 559–572.

Rosipal, R. & N. Krämer (2006). Overview and Recent Advances in Partial Least Squares. In *Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science*, pp. 34–51. Springer.

Sulic, V., J. Perš, M. Kristan, & S. Kovacic (2010, February). Efficient Dimensionality Reduction Using Random Projection. In *15th Computer Vision Winter Workshop*, Nové Hrad, Czech Republic, pp. 29–36.

Swets, D. & J. Weng (1996). Using Discriminant Eigenfeatures for Image Retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 18(8), 831–836.

Williams, L. & S. Ledwich (2004, December). Reduced SIFT Features For Image Retrieval and Indoor Localisation. In *Australasian Conference on Robotics and Automation*, Canberra, Australia.

Wold, S., A. Ruhe, H. Wold, & W. J. Dunn (1984). The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM Journal on Scientific and Statistical Computing* 5(3), 735–743.

Zhang, S., Q. Tian, G. Hua, Q. Huang, & S. Li (2009). Descriptive Visual Words and Visual Phrases for Image Applications. In *17th ACM International Conference on Multimedia*, Beijing, China, pp. 75–84. ACM.