Non-negative Matrix Factorization in Multimodality Data for Segmentation and Label Prediction

Zeynep Akata, Christian Thurau, Christian Bauckhage

16^{th} *Computer Vision Winter Workshop*
Andreas Wendel, Sabine Sternig, Martin Godec (eds.)
Mitterberg, Austria, February 2-4, 2011

Non-negative Matrix Factorization in Multimodality Data for Segmentation and Label Prediction

Zeynep Akata
Xerox Research Centre Europe
6, chemin de Maupertuis
38240 Meylan, France
zeynep.akata@xrce.xerox.com

Christian Thurau
Fraunhofer IAIS
Schloss Birlinghoven
53757 Sankt Augustin, Germany
christian.thurau@iais.fraunhofer.de

Christian Bauckhage
Fraunhofer IAIS
Schloss Birlinghoven
53757 Sankt Augustin, Germany
christian.bauckhage@iais.fraunhofer.de

Abstract. *With the increasing availability of annotated multimedia data on the Internet, techniques are in demand that allow for a principled joint processing of different types of data. Multiview learning and multiview clustering attempt to identify latent components in different features spaces in a simultaneous manner. The resulting basis vectors or centroids faithfully represent the different views on the data but are implicitly coupled and they were jointly estimated. This opens new avenues to problems such as label prediction, image retrieval, or semantic grouping. In this paper, we present a new model for multiview clustering that extends traditional non-negative matrix factorization to the joint factorization of different data matrices. Accordingly, the technique provides a new approach to the joint treatment of image parts and attributes. First experiments in image segmentation and multiview clustering of image features and image labels show promising results and indicate that the proposed method offers a common framework for image analysis on different levels of abstraction.*

1. Motivation and Background

The rise of the social web and the user generated content movement have turned the Internet into a virtually limitless repository of annotated and rated multimedia data. For example, as of this writing, there are more than 4.5 billion images available on `flickr` most of which are tagged, rated, categorized, and appraised by the community. This development offers tremendous possibilities for research on image understanding but also calls for methods that allow for an integrated processing of different types of data.

Our goal is a principled joint treatment of image features and image tags. We present a new technique for multiview clustering that simultaneously determines latent dimensions or centroid vectors in different feature spaces. In contrast to ad hoc methods such as, say, concatenating different types of features into a single descriptor, multiview clustering is faithful to the different characteristics of different descriptors. Since latent components or centroids are jointly estimated, multiview techniques allow for advanced inference. Since for every centroid in one feature space there is a corresponding centroid in another space, transitions between different views are straightforward. This offers auspicious new approaches to segmentation, automatic image tagging, or tag-based image retrieval.

Although they have a long and venerable tradition, there is a renewed interest in multiview learning and multiview clustering. The canonical example of a method that simultaneously uncovers latent components in different spaces is Hotelling's canonical correlation analysis (CCA) [12, 2] for which kernelized and probabilistic extension have been proposed as of late [7, 11, 3]. Other recent developments consider

extensions of spectral clustering to multiple graphs that encode different types of similarities [27, 21].

Our new approach to multiview clustering extends non-negative matrix factorization (NMF) [17, 16] to the joint factorization of several data matrices. It is motivated by the following considerations:

i) Similar to principal component analysis (PCA) [13] or singular value decomposition (SVD) [9] CCA does not necessarily do justice to purely non-negative data such as color histograms or term frequency vectors. Non-negative matrix factorization, however, typically yields results that can be seen as part-based representations and accommodate human perception.

ii) Methods based on spectral clustering of similarity matrices scale quadratically with the number of data and are therefore prohibitive in modern, large-scale data and image analysis problems.

iii) For NMF, on the other hand, there exist efficient algorithms that factorize matrices of billions of entries [23] which may apply to the multiview setting.

In the next section, we clarify the relation between matrix factorization and clustering. Then, in section 3, we briefly review NMF according to [17, 16] and extend this approach toward the joint factorization of different data matrices. In section 4, we present experiments on using multiview NMF in image segmentation, label prediction, and image retrieval. A conclusion will end this contribution.

2. Matrix Rank Reduction and Clustering

In this section, we briefly review how matrix rank reduction applies to the problem of clustering or vector quantization.

Consider a data matrix $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$ of rank $r \leq \min(m, n)$ whose column vectors $\boldsymbol{x}_i$ correspond to feature vectors obtained from some measurement process. Using the singular value decomposition (SVD) [9] any matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ can be written as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \sum_{j=1}^{r} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T \tag{1}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1 \dots \boldsymbol{u}_m] \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} = [\boldsymbol{v}_1 \dots \boldsymbol{v}_n] \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$. The SVD is a popular tool in data analysis because it is known that the optimal solution to the rank reduction problem

$$\min_{\operatorname{rank}(\tilde{\boldsymbol{X}})=k<r} \left\| \boldsymbol{X} - \tilde{\boldsymbol{X}} \right\|^2 \tag{2}$$

is given by

$$\tilde{\boldsymbol{X}} = \sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T. \tag{3}$$

Substituting $\boldsymbol{W} = \tilde{\boldsymbol{U}} \in \mathbb{R}^{m \times k}$ and $\boldsymbol{H} = \tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T \in \mathbb{R}^{k \times n}$, we recognize that $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$ is approximated as a product of a matrix of basis vectors and a matrix of coefficients. This allows for dimensionality reduction, since

$$\boldsymbol{x}_i \approx \sum_{j=1}^{k} \boldsymbol{w}_j h_{ji} \tag{4}$$

so that to every data vector $\boldsymbol{x}_i \in \mathbb{R}^m$ there is a coefficient vector $\boldsymbol{h}_i \in \mathbb{R}^k$ where $k < m$.

Depending on which constraints are imposed on $\boldsymbol{W}$ and $\boldsymbol{H}$, one obtains different dimensionality reduction schemes when solving the general matrix factorization problem

$$\min_{\boldsymbol{W},\boldsymbol{H}} \left\| \boldsymbol{X} - \boldsymbol{W}\boldsymbol{H} \right\|^2. \tag{5}$$

For instance, principal component analysis (PCA) [13] is recovered from

$$\begin{aligned} \min_{\boldsymbol{W},\boldsymbol{H}} & \left\| \boldsymbol{X} - \boldsymbol{W}\boldsymbol{H} \right\|^2 \\ \text{s.t. } & \boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}. \end{aligned} \tag{6}$$

Casting matrix factorization in a yet more general form reveals a connection to vector quantization and clustering. For example, running the k-means algorithm is tantamount to solving

$$\begin{aligned} \min_{\boldsymbol{G},\boldsymbol{H}} & \left\| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{G}\boldsymbol{H} \right\|^2 \\ \text{s.t. } & \boldsymbol{g}_j^T \mathbf{1} = 1 \\ & \boldsymbol{g}_j \succeq \mathbf{0} \\ & \boldsymbol{h}_i = [0 \dots 010 \dots 0]^T. \end{aligned} \tag{7}$$

Due to the convexity constraints on the columns of $\boldsymbol{G}$, the resulting basis vectors in $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{G}$ are convex combinations of certain data points in $\boldsymbol{X}$ and since the coefficient vectors in $\boldsymbol{H}$ are unitary vectors, every data point $\boldsymbol{x}_i$ in $\boldsymbol{X}$ will be represented by exactly one centroid $\boldsymbol{w}_j$ in $\boldsymbol{W}$.

3. NMF for Multiview Clustering

In this section, we first summarize non-negative matrix factorization (NMF) and then introduce our generalization of NMF toward multiview clustering.

3.1. Factorization of Data via NMF

Orthogonal basis vectors such as determined by PCA or SVD are not always the best choice for dimensionality reduction or clustering [17, 16, 25, 6, 15, 14]. In particular data that consist exclusively of non-negative measurements cannot be guaranteed to retain non-negativity after projection onto lower-dimensional subspaces that are spanned by its dominant eigenvectors. As an alternative that is true to the non-negative nature of certain data Lee and Seung popularized the idea of non-negative matrix factorization [17, 16]. In computer vision where image data typically consists of non-negative values, NMF was observed to yield superior results in segmentation, feature extraction, motion-, or pose estimation [26, 10, 4, 22].

Viewed as a constrained least squares optimization problem, NMF amounts to solving

$$\begin{aligned} &\min_{\boldsymbol{W},\boldsymbol{H}} \left\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\right\|^2 \\ &\text{s.t. } \boldsymbol{W},\boldsymbol{H} \succeq \boldsymbol{0}. \end{aligned} \tag{8}$$

Although (8) is convex in either $\boldsymbol{W}$ or $\boldsymbol{H}$, the simultaneous estimation of basis vectors and coefficients in (8) does not admit a closed form solution and is known to suffer from many local minima. A unique optimum provably exists [25], however, algorithms that are guaranteed to find it are not known to date (see the discussions in [25, 6, 15, 14]).

In the work presented here, we consider multiplicative fixed point iterations to find a solution to NMF because their extension to multiview clustering is immediate. In the following, $\boldsymbol{A} \odot \boldsymbol{B} \in \mathbb{R}^{m\times n}$ denotes the Hadamard product of two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ where $\big(\boldsymbol{A} \odot \boldsymbol{B}\big)_{ij} = a_{ij} \cdot b_{ij}$. The Hadamard division $\oslash$ is defined accordingly but for better readability we write $\boldsymbol{A} \oslash \boldsymbol{B} = \boldsymbol{A}/\boldsymbol{B}$.

Concerned with the problem in (8), Lee and Seung [17, 16] randomly initialize the matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. They derive the following update rules

$$\begin{aligned} \boldsymbol{W} &\leftarrow \boldsymbol{W} \odot \frac{\boldsymbol{X}\boldsymbol{H}^T}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T} \quad \text{and} \\ \boldsymbol{H} &\leftarrow \boldsymbol{H} \odot \frac{\boldsymbol{W}^T\boldsymbol{X}}{\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}} \end{aligned} \tag{9}$$

and prove their convergence using an expectation maximization argument. Next, we will extend this approach to multiview data.

3.2. Simultaneous Factorization of Multiview Data via NMF

Our main motivation behind the work presented in this paper is to cluster entities for which there are different types of data available. For instance, images retrieved from `flickr` can be characterized by means of different abstract image features but at the same time there are user generated tags or labels available that describe their content or formation. We hypothesize that simultaneous clustering of such different views on the data will yield more meaningful clusters and may provide a tool to fill in missing information. In particular, multiview clustering of image features and image tags may provide a way to predict a set of tags given an image or to retrieve relevant images from a database given a set of query tags.

Assuming a set of n different images, it can be characterized by an $m \times n$ image-feature matrix $\boldsymbol{X}$ as well as by an $l \times n$ term-by-image matrix $\boldsymbol{Y}$. Our basic idea is to uncover suitable bases $\boldsymbol{W}$ and $\boldsymbol{V}$ for the image- and text features, respectively, which are implicitly coupled via a common coefficient matrix $\boldsymbol{H}$. In other words, we aim at finding two low rank approximations

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} \quad \text{and} \quad \boldsymbol{Y} \approx \boldsymbol{V}\boldsymbol{H} \tag{10}$$

where $\boldsymbol{W} \in \mathbb{R}^{m\times k}$, $\boldsymbol{V} \in \mathbb{R}^{l\times k}$, and $\boldsymbol{H} \in \mathbb{R}^{k\times n}$.

Our solution is to formalize this idea as a convex combination of two constrained least squares problems

$$\begin{aligned} &\min_{\boldsymbol{W},\boldsymbol{V},\boldsymbol{H}} (1-\lambda)\left\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\right\|^2 + \lambda\left\|\boldsymbol{Y}-\boldsymbol{V}\boldsymbol{H}\right\|^2 \\ &\text{s.t. } \boldsymbol{W},\boldsymbol{V},\boldsymbol{H} \succeq \boldsymbol{0} \end{aligned} \tag{11}$$

where $\lambda \in [0, 1]$ is user specified constant that allows for expressing preferences for either of the two feature types. Just as with the original NMF problem in (8), the extended problem in (11) does not admit a closed form solution. We therefore adapt the Lee and Seung type fixed point iteration to our case. For the matrices of basis vectors $\boldsymbol{W}$ and $\boldsymbol{V}$, the update rules immediately carry through and read:

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{W} \odot \frac{\boldsymbol{X}\boldsymbol{H}^T}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T} \quad \text{and} \\ \boldsymbol{V} &= \boldsymbol{V} \odot \frac{\boldsymbol{Y}\boldsymbol{H}^T}{\boldsymbol{V}\boldsymbol{H}\boldsymbol{H}^T}. \end{aligned} \tag{12}$$

Since the coefficient matrix $\boldsymbol{H}$ now couples two bases, its update is slightly more involved. The simplified version of the fixed point iteration for the coefficients is:

$$\boldsymbol{H} = \boldsymbol{H} \odot \frac{(1-\lambda)\boldsymbol{W}^T\boldsymbol{X} + \lambda\boldsymbol{V}^T\boldsymbol{Y}}{\big((1-\lambda)\boldsymbol{W}^T\boldsymbol{W} + \lambda\boldsymbol{V}^T\boldsymbol{V}\big)\boldsymbol{H}}. \quad (13)$$

3.3. Discussion

Our choice of a convex combination of the individual optimization problems in (11) is not an arbitrary decision. There is a known close relation between non-negative matrix factorization and probabilistic latent semantic analysis [8, 5]. Assuming an appropriate normalization, NMF can be understood as learning the parameters of a joint probability distribution which is expressed as a product of marginal distributions. By choosing a convex combination of two NMF problems, this analogy may be lifted to the level of learning a distribution of distributions. This is akin to Latent Dirichlet Allocation [18, 1] but we will leave possible implications to future work.

We note that by setting $\lambda = 0$ or $\lambda = 1$ our model and its updates reduce to the original form of NMF. Moreover, the model is not confined to the case of two different types of views. Its extension to convex combinations of p different views is straightforward:

$$\begin{aligned} \min_{\boldsymbol{W}^i,\boldsymbol{H}} & \sum_{i=1}^{p} \lambda_i \big\|\boldsymbol{X}^i - \boldsymbol{W}^i\boldsymbol{H}\big\|^2 \\ \text{s.t. } & \boldsymbol{W}^i, \boldsymbol{H}, \boldsymbol{\lambda} \succeq \mathbf{0} \\ & \boldsymbol{\lambda}^T\mathbf{1} = 1 \end{aligned} \quad (14)$$

Finally, as with with all alternating least squares schemes, convergence of the extended update algorithm for multiview NMF is guaranteed. We omit the formal proof but sketch the argument: Given $\boldsymbol{H}$, none of the updates in (12) will increase either term in (11); given $\boldsymbol{W}$ and $\boldsymbol{V}$, the update in (13) cannot increase the expression in (11).

4. Experiments

In the following subsections we present first experimental results obtained from using multiview NMF for image segmentation, label prediction, and image retrieval. Note that, so far, these are preliminary experiments intended to validate the approach. We are currently working on extended experimental evaluations to compare the proposed approach to other methods in the literature.

4.1. Image Segmentation via Joint Non-negative Matrix Factorization

In a first series of experiments, we apply simultaneous NMF to the problem of image segmentation. We consider color images of natural scenes downloaded from `flickr`. We convert the RGB pixel values into the LUV color-space because of its alleged perceptual uniformity which ensures that equally distant colors in the color space would be also equidistant perceptually.

In order to segment an image into homogeneous regions, we sample 1000 pixels from each image and build two feature matrices, one containing 1000 three dimensional column vectors of color information and one containing 1000 two dimensional column vectors containing pixel coordinates. This way, we separate color from location and run simultaneous NMF to obtain centroid vectors $\boldsymbol{W}$ and $\boldsymbol{V}$ in the respective spaces that are coupled via the common coefficients $\boldsymbol{H}$.

We conduct several experiments where we vary the number of centroids $k = \{4, 10, 20\}$ and the weighting parameter $\lambda = \{0.1, 0.5, 0.9\}$. When λ is larger, more weight is given to the color descriptor of the pixels and when it is smaller more weight is given to the location of the pixels. After random initialization to positive values sampled from a Gaussian distribution, we run the update rules for the matrices $\boldsymbol{W}$, $\boldsymbol{V}$ and $\boldsymbol{H}$ until convergence but at most 100 times.

Given the results of the *training* phase, the *test* phase in these experiments consist in assigning every pixel $\boldsymbol{x}$ of an image to one of the k resulting cluster centroids. Given $\boldsymbol{W}$ and $\boldsymbol{V}$, we solve $\min(1-\lambda)\|\boldsymbol{x}-\boldsymbol{W}\boldsymbol{h}\|^2 + \lambda\|\boldsymbol{x}-\boldsymbol{V}\boldsymbol{h}\|^2$ for the coefficients $\boldsymbol{h}$ and determine the cluster index c according to

$$c = \operatorname*{argmax}_j h_j. \quad (15)$$

Figure 1 shows examples of images we considered in our segmentation experiments. The accuracy of the segmentation appears to improve with an increasing value of the weighting parameter λ. This corresponds to intuition because assigning more weight to color information should yield image segments grouped together based on color rather than on spatial proximity. However the result of segmentation seems best for $\lambda = 0.5$ where location and color values of the pixels contribute equally to the resulting matrix factors. This resembles the behavior of a bi-

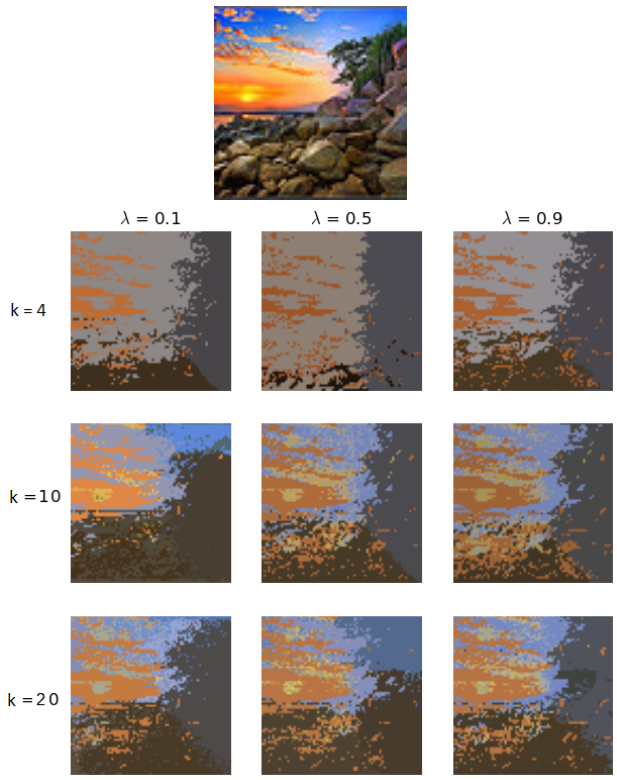

Figure 1. A sample image and its segmentation results obtained from computing cluster centroids using multiview NMF applied to pixel location- and color information. For a smaller λ, more weight will be assigned pixel location information, for a larger λ, more weight will be assigned to pixel color information. With larger weights on location information, small regions of rather homogenous color disappear in the segmentation process. For larger weights on color information, we observe a tendency towards over-segmentation and noisy segment boundaries. For the case where color and location information contribute equally, small regions are preserved and segment boundaries are smoother.

lateral filter [24] which also incorporates color- and location information and is known to yield smooth segment boundaries.

4.2. Label Prediction and Image Retrieval via Joint Factorization of Image- and Text-Features

This series of experiments aims at exploring whether or not multiview NMF is capable of filling in missing information. We considered a training set of natural images retrieved from the "most interesting" category at `flickr`. This set of training images contains 10 different classes (clouds, moonlight, beach, ship, bridge, mountain, forest, city, church, castle) of motives and we considered 300 images per class.

In these experiments, the feature vectors are calculated using local self similarity (SSIM) [20] feature extraction scheme. The feature vectors are then clustered into a visual vocabulary of $k = 750$ visual words. For each image in the dataset, a histogram of this vocabulary is created. The individual histograms of all the images in the dataset are then collected in an image-feature matrix $\boldsymbol{F} \in \mathbb{R}^{k \times n}$.

Textual descriptors for the tag list of the images are created by using the well known Bag of Features [19] approach. Firstly, the most frequent tags in the dataset are collected and the textual vocabulary or the dictionary is generated by filtering the irrelevant tags such as foreign names, `flickr` group names, and abbreviations. Secondly, all the tag lists corresponding to the respective images are compared with the dictionary and according to the presence (1) or absence (0) of the dictionary words in the tag list of an image, a binary text feature vector is formed. Finally, the feature vectors are stored in a matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ with n being the number of images in the dataset and $m = 1000$ being the size of the textual dictionary.

the matrices $\boldsymbol{W}$, $\boldsymbol{V}$ and $\boldsymbol{H}$ were initialized to random positive values sampled from a Gaussian and we ran the multiview NMF update algorithm until convergence but at most 100 times, to obtain coupled factorizations ($k = 10$, $\lambda = 0.5$) of the image- and text-feature matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. In the test phase of these experiments, we considered two different settings.

4.2.1 Label Prediction

Given an image that was not part of the training set, we compute its image-feature vector $\boldsymbol{x}$ and solve $\min\|\boldsymbol{x} - \boldsymbol{W}\boldsymbol{h}\|^2 \text{s.t.} \boldsymbol{h} \succeq \boldsymbol{0}$ for $\boldsymbol{h}$. Given $\boldsymbol{h}$, we plug it into $\boldsymbol{y} = \boldsymbol{V}\boldsymbol{h}$ to obtain a corresponding vector $\boldsymbol{y}$ in the text-feature space.

Given $\boldsymbol{y}$, we search for that column vector $\boldsymbol{y}_i$ of the training data matrix $\boldsymbol{Y}$ for which $\|\boldsymbol{y} - \boldsymbol{y}_i\|$ is minimal. We use $\boldsymbol{y}$ to predict a ranked list of tags. To this end, we determine and rank those words in the lexicon that correspond to the 20 basis vectors $\boldsymbol{t}_i$ in the original text-by-image space for which the projection $\boldsymbol{y}^T\boldsymbol{t}_i$ is maximal. The 10 highest ranked tags are selected to be the tag list of the test image. In Figure 1, the retrieved tags for some of the images are shown.

(a) bridge

(b) bridge + sea

(c) bridge + sea + sky

(d) bridge + sea + sky + building

Figure 2. The 3 most relevant images retrieved by querying with the word or the group of words below them. The retrieved images tend to be more specific with the increasing number of words used in the queries.

4.2.2 Image Retrieval

In this setting, we queried random words such as bridge, sea, sky individually or as a group to retrieve the best corresponding images. The text feature vector $\boldsymbol{y}$ of the random words are created the same way as training tag lists of the images. We then solve $\min\|\boldsymbol{y} - \boldsymbol{V}\boldsymbol{h}\|^2 \text{s.t.} \boldsymbol{h} \succeq \boldsymbol{0}$ for $\boldsymbol{h}$. Given $\boldsymbol{h}$, we plug it into $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{h}$ to obtain a corresponding vector $\boldsymbol{x}$ in the image-feature space.

Given $\boldsymbol{x}$, we search for that column vector $\boldsymbol{x}_i$ of the training data matrix $\boldsymbol{X}$ for which $\|\boldsymbol{x} - \boldsymbol{x}_i\|$ is minimal. The four most similiar images are shown in Table 2 that correspond to the words below.

high	blue	water
travel	water	sky
cruise	trees	clouds
holiday	bridge	waves
morning	reflections	rocks
cityscape	grass	sea
daybreak	yellow	ocean
tower	woods	seascape
sea	railing	raining
land	waterscape	waterscape

water	sky	nightscape
beach	clouds	blue
sunrise	blue	stars
outdoors	holiday	afterdark
nature	red	sky
reflection	castle	night
landscape	bluesky	landscape
raining	raining	moonlight
walking	disneyland	yellow
yellow	middleages	city

Table 1. Results of automatic image annotation. The taglist corresponds to the first ranked 10 tags retrieved by querying an unknown image.

5. Conclusion and Future Work

The work presented in this paper aims at the analysis of images for which there is additional information available. We introduced a new model for multiview clustering that extends the idea of non-negative matrix factorization (NMF) towards the joint analysis of different types of features. We cast multiview NMF as a convex combination of individual optimization problems and adopt the well known multiplicative fixed point algorithm for NMF to this case. The approach avoids ad hoc combinations of different types of features and thus stays true to the nature of different descriptors. The individual optimization problems in our multiview NMF formulation are coupled via a common coefficient matrix. Due to

this coupling, the resulting basis vectors or cluster centroids allow for inferring one type of descriptor (e.g. image labels) from another type of descriptor (e.g. image features).

In preliminary experiments we validated the applicability of the proposed approach in image segmentation, tag prediction, and tag-based image retrieval. Our first results suggest that multiview clustering can provide a framework for image analysis that applies to different levels of abstraction. Image parts could be identified by combining pixel-color and -location information in the principal manner that is provided by the multiview approach. Information as diverse as color histograms and text-by-image vectors were coupled using our framework and we found it to be capable to predict missing information from what data was available.

Currently, we are conducting more extensive experiments to provide a more quantitative analysis as well as to compare the proposed approach to other multiview methods such as (kernelized) canonical component analysis. In contrast to related methods from the literature, we expect that highly efficient implementations of multiview NMF will be possible. To this end, we are currently adopting techniques such as convex-hull NMF to our model. We will also further explore how multiview NMF relates to LDA and whether it offers an alternative approach to hierarchical latent topic models. Finally, we envision further applications of the proposed method, for instance in the area of hyperspectral imaging.

References

[1] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet Allocation. *J. of Machine Learning Research*, 3(Jan. 2003):993–1022, 2003. 4

[2] M. Borga, T. Landelius, and H. Knutsson. A Unified Approach to PCA, PLS, MLR and CCA. Technical Report LiTH-ISY-R-1992, ISY, Linköping University, 1997. 1

[3] K. Chaudhuri, S. Kakade, K. Liescu, and K. Shridharan. Multi-View Clustering via Canonical Correlation Analysis. In *Proc. ICML*, 2009. 1

[4] A. Cheriyadat and R. Radke. Non-Negative Matrix Factorization of Partial Track Data for Motion Segmentation. In *Proc. IEEE ICCV*, 2009. 3

[5] C. Ding, T. Li, and W. Peng. NMF and PLSI: Equivalence and a Hybrid Algorithm. In *Proc. ACM SIGIR*, 2006. 4

[6] D. Donoho and V. Stodden. When Does Non-negative Matrix Factorization Give a Correct Decomposition into Parts? In *Proc. NIPS*, 2004. 3

[7] K. Fukumizu, F. Bach, and A. Gretton. Statistical Consistency of Kernel Canonical Correlation Analysis. *J. of Machine Learning Research*, 8(2):361–383, 2007. 1

[8] E. Gaussier and C. Goutte. Relations between PLSA and NMF and Implications. In *Proc. ACM SIGIR*, 2005. 4

[9] G. Golub and J. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996. 2

[10] D. Guillamet, J. Vitria, and B. Schiele. Introducing a Weighted Non-negative Matrix Factorization for Image Classification. *Pattern Recognition Letters*, 24(14):2447–2454, 2003. 3

[11] D. Hardoon and J. Shaw-Taylor. Convergence Analysis of Kernel Canonical Correlation Analysis: Theory and Practice. *Machine Learning*, 74(1):23–38, 2009. 1

[12] H. Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3–4):321–377, 1936. 1

[13] I. Jolliffe. *Principal Component Analysis*. Springer, 1986. 2

[14] B. Klingenberg, J. Curry, and A. Dougherty. Non-negative Matrix Factorization: Ill-posedness and a Geometric Algorithm. *Pattern Recognition*, 42(5):918–928, 2008. 3

[15] A. Langville, C. Meyer, and R. Albright. Initializations for the Nonnegative Matrix Factorization. In *Proc. ACM KDD*, 2006. 3

[16] D. Lee and H. Seung. Algorithms for Non-negative Matrix Factorization. In *Proc. NIPS*, 2000. 2, 3

[17] D. D. Lee and H. S. Seung. Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature*, 401(6755):788–799, 1999. 2, 3

[18] D. MacKay and L. Peto. A Hierarchical Dirichlet Language Model. *Natural Language Engineering*, 1(3):1–19, 1995. 4

[19] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York u.a., 1983. 5

[20] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *IEEE Conference on Computer Vision and Pattern Recognition 2007 (CVPR'07)*, June 2007. 5

[21] W. Tang, Z. Lu, and I. Dhillon. Clustering with Multiple Graphs. In *Proc. IEEE ICDM*, 2009. 2

[22] C. Thurau and V. Hlavac. Pose Primitive Based Human Action Recognition in Videos or Still Images. In *Proc. IEEE CVPR*, 2008. 3

[23] C. Thurau, K. Kersting, and C. Bauckhage. Convex Non-Negative Matrix Factorization in the Wild. In *Proc. IEEE ICDM*, 2009. 2

[24] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. In *Proc. IEEE ICCV*, 1998. 5

[25] N. Vasiloglou, A. Gray, and D. Anderson. Non-Negative Matrix Factorization, Convexity and Isometry. In *Proc. SIAM DM*, 2009. 3

[26] R. Zass and A. Shashua. A Unifying Approach to Hard and Probabilistic Clustering. In *Proc. IEEE ICCV*, 2005. 3

[27] D. Zhou and C. Burges. Spectral Clustering and Transductive Learning with Multiple Views. In *Proc. ICML*, 2007. 2