KALMAN FILTERING FOR POSE-INVARIANT FACE RECOGNITION

Horst Eidenberger

Vienna University of Technology, Interactive Media Systems Group

ABSTRACT

We propose a novel algorithm for the identification of faces from image samples. The algorithm uses the Kalman filter to identify significant facial traits. *Kalmanfaces* are compact visual models that represent the invariant proportions of face classes. We employ the Kalmanfaces approach on the UMIST database, a collection of face images that were recorded under varying camera angles. Kalmanfaces show robustness against invisible facial traits and outperform the classic Eigenfaces approach in terms of identification performance and algorithm speed. The paper discusses Kalmanfaces extraction, application, tunable parameters, experimental results and related work on Kalman filter application in face recognition.

Index Terms–Face recognition, Kalman filtering.

1. INTRODUCTION

Face recognition is one of the classic areas of pattern recognition [6]. Applications are manifold ranging from video surveillance to content-based retrieval. Face recognition research focuses on two problem areas: detection of faces in visual media streams and identification of detected faces. Face identification comprises of holistic approaches (e.g. Eigenfaces [3]), classification approaches (Linear Discriminant Analysis, Support Vector Machines, etc.) and regression approaches (e.g. Neural Networks).

We propose a novel holistic approach for face identification from image samples that uses a simplified Kalman filter [2, 4] to detect invariant face features. The Kalman filter is frequently employed in face recognition for face detection in video sequences. However, the author is not aware of proposals to use the powerful linear data processing capabilities of the Kalman filter for the extraction of face features. We use the Kalman filter on a dataset of face images that were recorded under varying camera angles. These data are distinguished by a high variance of location and partial invisibility of the typically considered face features (eyes, mouth, etc.). Viewpoint-variant recognition is one of the hardest problems in face recognition (and object recognition in general). The authors of [6] stress that "generalization even from one profile to another is poor" (page 5, first paragraph). The results for classic Eigenfaces (Section 3) confirm this judgment. The Kalman filter considers variances in the data analysis process. Hence, we expect it to be an advantageous element of a model for robust face identification under such circumstances.

The paper is organized as follows. Section 2 explains the Kalmanfaces extraction process. Section 3 discusses experiments and results. Section 4 sketches related work in face recognition and Kalman filter application.

2. KALMANFACES APPROACH

2.1. Kalmanfaces Extraction

The Kalmanfaces approach identifies the most likely face class for an image by feature similarity. It expects every face class (person) to be represented by a sequence of image samples. The number of inputs should not be smaller than 3-5 for reasonable application of the Kalman filter. Each face class is represented by a single feature vector that is extracted as follows:

1. *Image normalization*. All face images are transformed to luminance (greyscale) matrices of the same size (for example, just three by three pixels). In the experiments we employ a simple nearest neighbour selection. Every pixel represents one face *region*.
2. *Averaging*. An average face is computed from the normalized images by a Kalman filter ("Kalmanface").
3. *Feature extraction*. Only those regions of the Kalmanface are considered as features that are sufficiently invariant. The luminance variance of a region must not exceed a certain threshold.

The Kalman filter is applied in the second step to compute a class average that represents the facial traits adequately. Figure 1 shows an example. The face class consists of twenty images like the ten examples on the left (*a*). Element *b* shows the mean image, element *c* shows the Kalman-averaged image. As can be seen, the Kalman-averaged image contains a large amount of the high-frequency information present in the examples. The properties of the Kalman filter cause that variances in the most prominent positions (portrait, side-face) are preserved and in-between images are (to a certain extent) absorbed. The average image *b* is not able to capture relevant facial traits properly. It consists of self-similar pixel neighbourhoods.

Kalmanface averaging is performed as follows. We assume the class samples to be a temporal sequence and compute the Kalman estimate *for each pixel*:

$$x_t = x_{t-1} + k_t\left(x_{t-1} - l_t\right) \tag{1}$$

x_t is the estimate of the pixel average at time t (the tth example image), l_t is the luminance value and k_t is the Kalman weighting factor (depending on the luminance variances at times t and $t-1$):

$$k_t = \frac{\sigma_{t-1}}{\sigma_{t-1} + \sigma_t} \tag{2}$$

The *Kalmanface* of a face class is the spatial aggregation of the pixel averages at time T ($[x_T]$). Weight k_t is the crucial factor in this simplified application of the Kalman filter that does not consider complex noise models and weights for the measurements l_t. σ_t is the standard deviation of the considered face region at time t. k_t approaches zero if the variance increases, i.e. if the luminance of a pixel changes from sample to sample. In this case, the Kalman

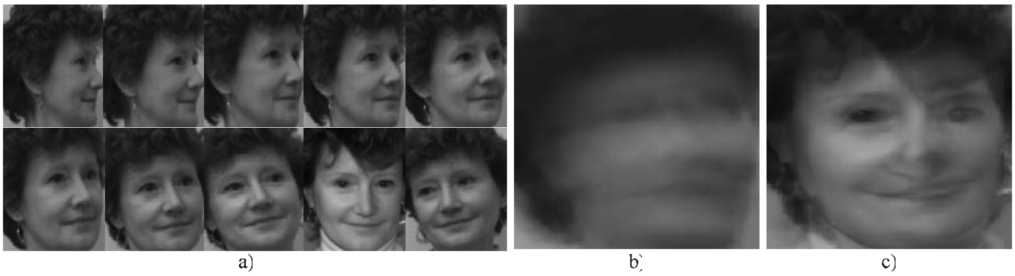

Fig. 1. Kalmanface example. Kalmanface *c* is constructed from face class *a* (*b* is the mean image).

filter trusts on the earlier estimate and disregards l_t. k_t approaches *1* if the variance decreases. In this case, the Kalman filter trusts on the luminance. The short time behaviour of the Kalman filter is to eliminate variances. In the long term Kalman filtering results in an average that preserves the properties of the input sequence (see Figure 1). It processes all information that is provided [2].

In the third step, features are extracted from the Kalman-averaged face. We select those pixels as face features that have a luminance variance σ_t below a certain threshold. Hence, the face feature vector consists only of those traits that are relatively invariant over the samples. The threshold is an endogenous variable (Section 3 discusses approximations).

2.2. Similarity Measurement

Kalmanfaces querying is a straightforward application of the vector space model. We assume an Euclidean feature space. The query example is normalized to the same number of pixels as the face classes and compared to each of the Kalmanfaces. That is, one distance measurement operation per individual in the database has to be performed. We suggest a first order Minkowski distance (city block distance) for dissimilarity measurement.

$$d(f,c) = \frac{\sum_{i}^{n_c} |f_i - c_i|}{n_c} \tag{3}$$

The dissimilarity of a face *f* to a face class c_i (represented by a Kalmanface) depends on the first order distance normalized by the feature vector size n_c of the face class. Only those pixels are considered for distance measurement that satisfy the variance condition stated above. This feature selection may (and usually will) change from face class to face class.

2.3. Discussion

We propose the Kalmanfaces approach as a solution for face detection in environments with high variance (e.g. varying camera angles, varying lighting). In particular, Kalmanfaces have the following advantages:

1. Face class information is easily extensible. One further iteration of the Kalman filtering process is sufficient to add a new face image of an already registered person.
2. The application of the variance condition leads to short feature vectors. Distance measurement of short vectors by a linear function allows for fast querying.
3. The length of the feature vectors is generally independent of the number of individuals in the database. Effective discrimination will require longer feature vectors for larger databases. However, database size and feature vector length are not as closely linked as, for example, in the Eigenfaces approach. In classic Eigenfaces the number of weights equals the number of face classes [3].

The Kalmanfaces approach does not try to influence the order of the images that are employed in the filtering process. If k_t is assumed constant, then the Kalman filter tends to lay a higher weight on the last samples than on the first. However, in practical application k_t alone determines the extent to which samples are represented in Kalmanfaces. By quantitative analysis we have found that the order of examples does not influence the face retrieval performance (this study is currently under review).

The Kalman filter weight (equation 2) and the variance threshold cause – to a certain extent – opposite effects. The application of k_t leads to a maximum of entropy in the Kalmanfaces. However, only those pixels are picked by the variance condition that are sufficiently invariant ("trustworthy"). Hence, Kalmanfaces feature vectors represent face classes by a maximum of information at a controllable level of trust.

3. EXPERIMENTS AND RESULTS

Below, we compare the performance of Kalmanfaces to the Eigenfaces approach [3] on the UMIST dataset [5]. Every Eigenfaces class Ω_i is averaged over *all* class members (highest quality). The UMIST dataset comprises of photographs showing individuals that were recorded with the same expression from various camera angles under constant lighting conditions. The nearest neighbour function is used for image resizing. Kalmanfaces and Eigenfaces are implemented as Matlab functions and can be downloaded from [1]. Kalmanface averaging is employed with the following approximation for all pixels of the first two faces (otherwise, k_1 would always be zero):

$$x_1 = \frac{l_0 + l_1}{2} \tag{4}$$

3.1. Face Retrieval Performance

Figure 2 summarizes the face identification performance depending on the feature size, i.e. the frame length in pixels of the quadratic face images. Kalmanfaces that use all feature elements (variance threshold of 100%, depicted as "v=100%") perform comparable to Eigenfaces. Interestingly, Eigenfaces and Kalmanfaces with a variance threshold of 50% or more (all features with at most 50% of the maximum variance are used) fail in reaching 100% face identification performance independently of feature size. Obviously, these parameterizations lead to feature

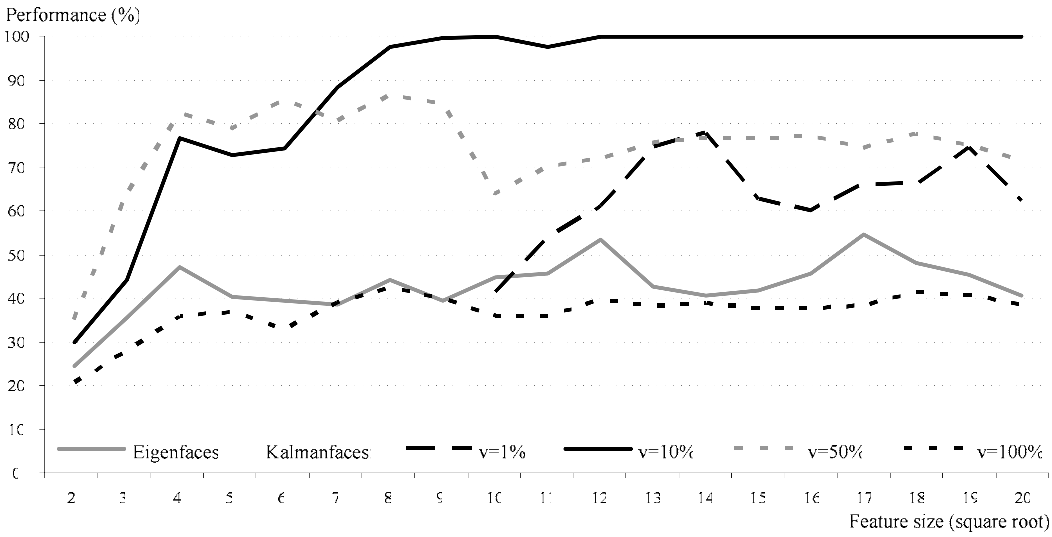

Fig. 2. Performance of Eigenfaces and Kalmanfaces at different variance threshold levels (not smoothed). "v=50%" means that the Kalmanfaces variance threshold is set to 50%.

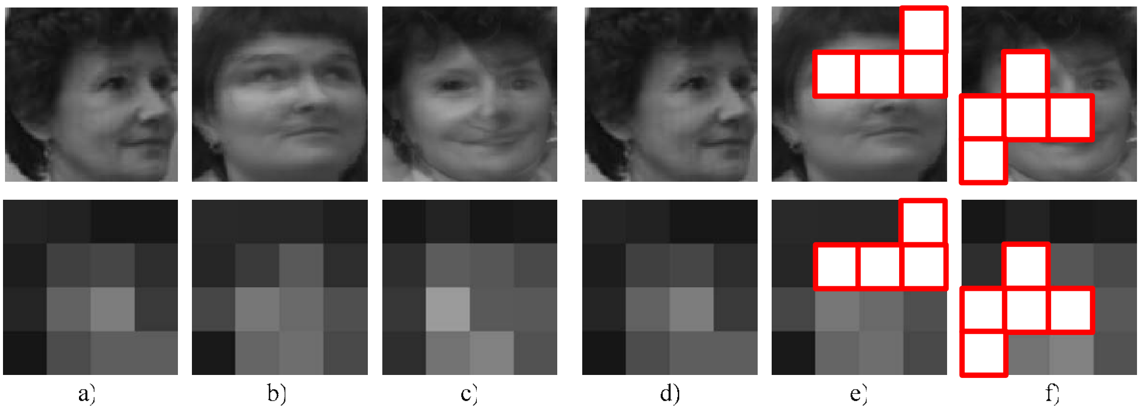

Fig. 3. Misclassification example. If all features (bottom line, *a-c*) are considered, then face *a* is misclassified as person *b* instead as person *c*. If only features with a variance of at most 50% of the maximum are considered (*d-f*), then the face is classified correctly.

vectors that contain partially misleading elements (those with high variance). Kalmanfaces with a variance threshold of 50% perform already significantly better than Kalmanfaces with a threshold of 100% (a gap of about 30-50% face identification performance).

However, the best performance can be observed for Kalmanfaces with a variance threshold of 10%. At this level the relationship of entropy and invariance in the feature data leads to optimal results. The parameterization reaches a face classification performance of 100% for features of 8 by 8 pixels. Even for very short feature vectors it is already superior over Eigenfaces. (A Kalmanface of 4 by 4 pixels and a variance threshold of 10% lead to a feature vector of approximately 8 (4^2*0,5) elements.) At a feature size of about 50 elements (10^2*0,5) the face identification performance reaches the ceiling. For more accurate Kalmanfaces performance remains constant at the optimal recognition level.

The results reflect the seriousness of the investigated recognition problem in the weak performance of the Eigenfaces. We have decided to compare Kalmanfaces to this approach, because they are structurally similar. Firstly, both are holistic approaches. That is, they derive face similarity from the entire image data and do not try to extract particular facial features. Secondly, both methods neglect semantic knowledge. In fact, both methods could be applied to arbitrary object recognition problems. Eventually, neither Kalmanfaces nor Eigenfaces require a training process for sample-based classification. Without doubt, a feature-based approach that makes use of kernel-based learning (e.g. a Support Vector Machine) would be able to outperform Eigenfaces (especially, on a small but well-defined scientific database). However, it would be difficult to find a relation of such results to the performance gain achieved by Kalmanfaces with increasing feature size.

Clearly, the variance threshold is the decisive parameter in the application of Kalmanfaces. At a threshold of 50% Kalmanfaces outperform Eigenfaces, at 10% retrieval performance is soon optimal. At 1% (approximately 10% of the most invariant features are chosen for similarity measurement) performance drops under the level of a 50% variance threshold. Precise judgement of the variance threshold is obviously crucial for retrieval performance.

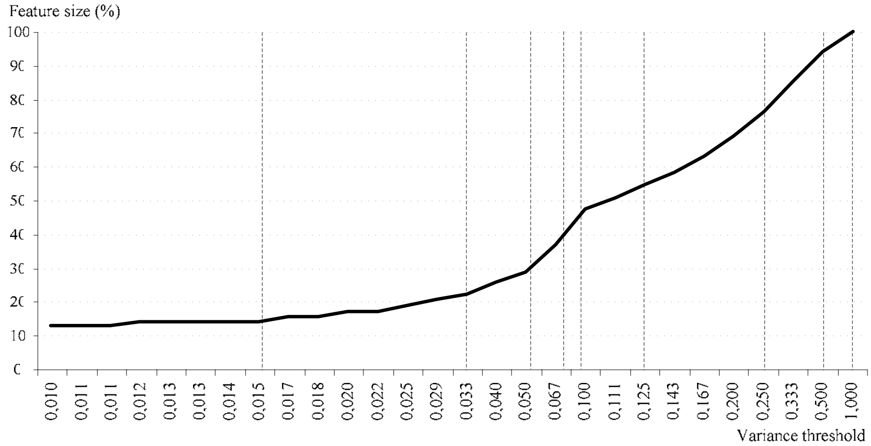

Fig. 4. Variance and feature size. Feature size depends on the threshold. The graph uses two log scale at the intervals [0, 0.1] and [0.1, 1].

The next subsection discusses this parameter in detail.

3.2. Relevance of the Variance Condition

The variance threshold determines which features are used for Kalmanfaces similarity measurement. Only those features of a face class are selected that have a variance σ_t below the threshold (given in percent of the maximum variance in the face class). That is, only sufficiently invariant features are considered. The application of variance thresholds causes no computational overhead, since the variances have to be calculated for Kalmanface averaging anyway.

Figure 3 shows a typical misclassification example (feature vector edge lengths of four pixels). If a threshold of 100% is used, face *a* is misclassified as member of face class *b* instead of *c*. If a threshold of 50% is employed, fewer regions are considered (depending on the face classes). Then, face *d* is correctly classified as a member of Kalmanface class *f*.

However, it has to be mentioned that the reduction of the feature vector to just eleven elements in Figure 3f is not typical. Figure 4 gives the average relationship of variance threshold value and feature vector size (in percent to the Kalmanface class sizes). In average, a variance threshold of 50% leads to the elimination of just 3-5% of the features with highest variance.

We would like to close this section with a few considerations on the relationship of performance and feature vector length. A variance threshold of 10% means that approximately 50% of the features of each Kalmanface class are considered for similarity measurement. Hence, the best-performing Kalmanfaces in Figure 2 (v=10%) use feature vectors with 50% of the total feature size. The feature vectors of Eigenfaces are proportional to the number of face classes (UMIST dataset: 20). A feature vector length of 20 elements is reached by Kalmanfaces of edge length 6 and v=10% (6^2*0.5). At this level, Kalmanfaces with a variance threshold of 10% outperform Eigenfaces by 35%.

4. RELATED WORK

Face detection and face recognition have applications in a large number of domains (visual retrieval and video surveillance, to name a few). Hence, it is not surprising that hundreds of new approaches are suggested every year (see, for example, [6]). The Kalman filter [2] is employed in a number of approaches to identify face locations in video streams. However, its beneficial properties are hardly exploited for face class description. This is surprising, since a large number of approaches depend on face class averaging (Principal Component Analysis, Linear Discriminant Analysis, Machine Learning approaches, etc.). Most of these approaches rely on the statistical mean, though the mean is for two reasons disadvantageous for this task. Firstly, it does not conform structurally to the original population. The mean can only under certain assumptions be interpreted with respect to the underlying data. Secondly, the application of the mean function has a blurring effect. Fragile high-frequency information (as the facial traits important for identification) is lost.

5. CONCLUSIONS

We propose a novel approach for face identification that uses the Kalman filter for face class averaging. Experimental evaluation shows that Kalmanfaces perform excellently on face images that were recorded under varying camera angles. Classic Eigenfaces are outperformed by up to 50%. The Kalmanfaces approach scales well with increasing numbers of individuals and face examples. In future work, we will investigate its sensitivity to more/less-variant face classes and its performance on other face identification problems (e.g. aging) and databases (e.g. FERET).

7. REFERENCES

[1] Kalmanfaces procedures, www.ims.tuwien.ac.at/~hme/tr/Kalmanfaces.zip, last visited 2006-04-25.

[2] Maybeck, P.S., *Stochastic models, estimation, and control*, Academic Press, Burlington MA, 1979.

[3] M. Turk, and A. Pentland, "Eigenfaces for recognition," Cogn. Neurosciences 3, MIT Press, Cambridge MA, 1991, pp. 71-86.

[4] G. Welch, and G. Bishop, "An Introduction to the Kalman Filter," Technical Report, University of North Carolina at Chapel Hill, www.cs.unc.edu/~welch/kalman/kalmanIntro.html, 2004, last visited 2006-04-25.

[5] UMIST face database, images.ee.umist.ac.uk/danny/database.html, last visited 2006-04-25.

[6] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," ACM Computing Surveys, ACM Press, New York NY, 2003, pp. 399-458.