Détection de visages par caractéristiques génériques calculées à partir des images de luminance

Face detection by robust generic features computed from luminance

Daniela Hall and James L. Crowley
PRIMA, GRAVIR-IMAG
INRIA Rhône-Alpes, 38349 St. Ismier, France
Daniela.Hall@inrialpes.fr

Résumé

Dans cet article nous proposons une methode pour l'apprentissage automatique de caractéristiques robustes et générales à partir d'images de luminance de visages. Les détecteurs ainsi obtenus sont robustes aux changements d'illumination, de personne et d'expression faciale. Pour obtenir un détecteur de visage fiable, les relations spatiales entre les caractéristiques détectées doivent être prises en compte. Nous proposons d'apprendre les relations des caractéristiques les plus remarquables par un histogramme en espace logarithmique angulaire. La reconnaissance est obtenue par évaluation de la mesure de divergence entre l'histogramme modèle et l'histogramme observé. Nos exemples montrent des détections fiables dans des cas qui représentent des défis pour notre approche. Nos experiences montrent qu'un modèle de n'importe quelle classe d'images peut être généré à partir d'un faible nombre d'exemples.

Mots Clef

Extraction et représentation des connaissances, classification, reconnaissance de formes dans l'image

Abstract

In this article we propose a method for learning generic and robust features from a visual image class and apply it to face images. The resulting feature detectors are robust to illumination, person identity, gender, and facial expressions. In order to obtain a powerful class detector, we learn the spatial relations of the most stable class features by computing a histogram in log-polar space. Detection is then performed by computing the histogram divergence between query and model histogram. The target objects, in our case faces, are detected under challenging conditions, even in the case of unconstrained images. The proposed method is general, and can be applied to learn any visual image class.

Keywords

Knowledge representation, classification, appearance based pattern recognition.

1 Introduction

In this article we propose a method for the automatic computation of robust and generic features from images forming a visual class. We have chosen faces to demonstrate the performance of our method. The goal of our research is to develop a system that can learn any visual class. For this reason we avoid to use any constraints that simplify the face detection task and make the system more specialised. We focus on the fast and simple training phase and on the possibility to apply our method to any set of images. The system presented here is not meant to compete with the very specialised face detection systems summarised in [18]. We use the face detection example to validate our approach.

In our example application, the features are learned from face images acquired under controlled illumination conditions and can be applied to unconstrained face images. This experiment demonstrates the ability of our generic feature detectors to generalise from few examples to unknown images of the same class. A key point to the robustness to illumination changes is the detection of facial features from luminance. The robustness to the changes in acquisition conditions of the generic features is demonstrated in the experiments.

In the second part of this article, we propose a method to learn the spatial relations of the generic features. The obtained model in form of a log-polar histogram serves for detection. This two stage learning system has the advantage that it combines the properties of the low level feature extraction and the higher level spatial relation context. The resulting model inherits the robustness from the feature extraction and the discriminance from the spatial relation context. A model computed from few images produces good detection results. The system can be applied to any type of images and it requires little supervision during the learning stage.

The remaining article is organised as follows. Section 2 explains the extraction of the raw appearance features using scale normalised Gaussian receptive fields. Section 3 describes the clustering approach for computing the generic and robust feature detectors from the raw features. In Section 4 we describe how to judge the quality of the generic feature detectors. The developed measure allows to select high quality feature detectors. Section 5 describes an approach to learn the spatial relations of local measurements. The proposed log-polar histograms are a means to model the data and to avoid over-fitting. Experimental results are given in Section 6. The experiments show examples of successful detection in challenging cases.

2 Feature Selection

Gaussian derivative receptive fields are used by many researchers for the description of local feature appearance [2, 8, 11, 12, 14]. Low order derivatives measure the basic geometries of features [5]. Local features are represented by the response to a bank of Gaussian derivative receptive fields centered on the image position. Scale invariant receptive fields are obtained by normalization to intrinsic scale at each pixel, where the intrinsic scale is determined from extrema in the normalised Laplacian over scale [7].

Many popular face detection methods use chrominance to detect skin regions [15, 17]. However, the chrominance information perceived by the camera is the product of the objects pigment and the color of the illumination according to the dichromatic reflection model described by Klinker [4]. By restricting the feature space to the luminance component, we obtain a facial feature detector that is not sensitive to changes in illumination color.

For the description of local appearance features we use first and second order Gaussian derivatives of the luminance channel. The restriction to the luminance channel and the suppression of the derivative of order zero makes the feature vector less sensitive to illumination variations. In our previous work [3], experiments with feature spaces up to third order derivatives showed no increase in discrimination quality. The higher dimension of the feature space increases the average distance between associated features which augments the error rate.

Features situated at positions of a dense pixel-wise grid are extracted at the specific intrinsic scale. This produces a large number of data points from a small number of images which is good for the subsequent learning algorithm. The data is normalised to compensate for the dynamic of receptive fields of different orders such that the distribution has 0.0 mean and 1.0 standard deviation.

Traditional methods use scale invariant local feature descriptors. This has the advantage that features that occur at different scales due to perspective transformation are associated. Scale invariant feature description allows matching invariant to the feature scale. If such features are used for modelling, the model does not contain information of the relative scale between features.

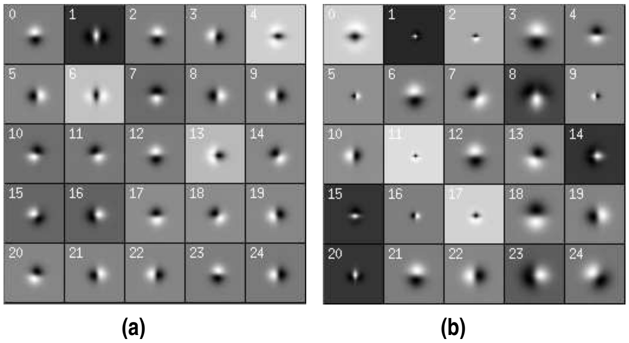

Figure 1: Cluster examples. (a) scale invariant feature space up to order 2. (b) feature space up to order 2 with scale.

The relative scale relations are discriminant for the object and should be preserved. For this reason we have developed a feature space that we refer to as scale feature space which allows the generation of a globally scale invariant model that preserves the internal scale relations. We propose to use $(L_x, L_y, L_{xx}, L_{xy}, L_{yy}, \sigma)$, where L_x denotes the first derivative of image I in direction x. This corresponds to an extension of the Gaussian derivative feature space by an additional dimension containing scale. This additional scale dimension allows to take into account the local scale for clustering.

3 Computation of generic features

The idea of vector quantization or clustering of the outputs of linear filter sets has been applied by Leung and Malik for texture recognition and image segmentation [6, 9]. They define texture as entity with spatially repeating properties. Zhu and his collaborators obtain clusters robust to rotation and scale changes by applying a transform component analysis to image patches before clustering [20]. The textons that represent the texture clusters allow the efficient modeling of textures. Schmid has applied the same k-means clustering scheme to compose generic features for image indexing [13].

Faces are composed of facial features, that consist of particular local appearances. Facial features of different faces have similar appearance such that all face images can be considered to form a visual class. The local feature appearance is captured by an appropriate feature space such as the scale feature space described in Section 2. A visual class has spatially repeating properties over the elements of the class. Clustering as applied by Malik for texture classification finds these repeating properties and learns their variations. The result is a set of associated point clouds which we refer to as generic features or classtons. The choice of this name is an analogy to Malik's texton prototypes.

We use k-means to associate nearby points and find the classton clusters. K-means is an iterative algorithm that converges to a local minimum. To avoid the problem of

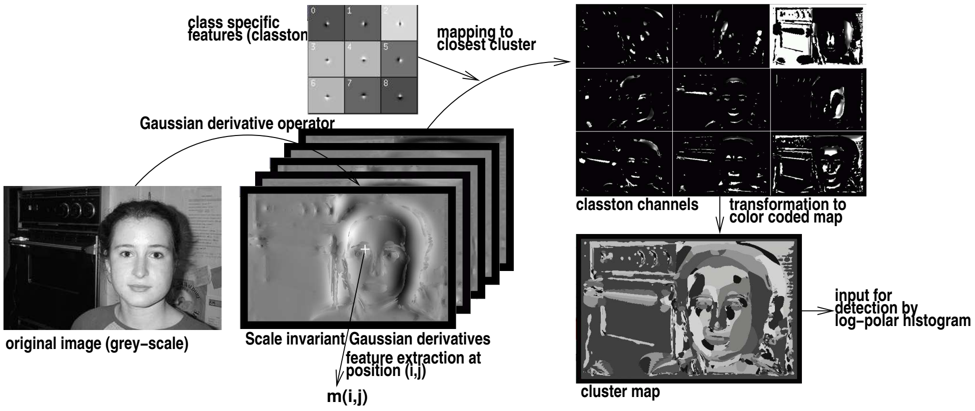

Figure 2: Feature extraction algorithm.

finding a suboptimal solution, k-means is performed several times and the best solution in terms of overall error is kept. We experimented with k in the range of 10 to 50. Figure 1 shows examples of clustering results. The clusters are represented as the center of gravity of the cluster elements. The vectorial representation of the gravity centers provide the weights of the linear combination of the impulse responses of the Gaussian derivative receptive fields. This linear combination is only used to visualize the mean feature of the associated cluster points. It contains no information about the shape of the point cluster in feature space. Figure 1 (a) shows the cluster centers using a scale normalised feature space, Figure 1 (b) shows the centers of a feature space with additional dimension for scale.

The classtons are a set of descriptors that model the repeating properties of any image class. For an observed image, we can compute classton channels in the same way as Malik. In a classton channel those image positions are marked whose underlying feature is mapped to the particular classton by evaluating a distance measure such as the one described in equation 2. The classton channels provide a partition of the image. It is therefore possible to display several classton channels in a single image (cluster map) coded as different colors (pixels marked by the same grey-level correspond to the same classton channel). The mapping from the Gaussian derivative features, $\vec{m}_{xy}$, to the cluster map representation, $M(x,y)$, can be formalised as follows.

$$M(x,y) \quad = \quad \arg \max_{j=1,..,k} d_j(\vec{m}_{xy}) \tag{1}$$

where

$$d_j(\vec{m}_{xy}) \quad = \quad |\vec{m}_{xy} - \vec{\mu}_j| \tag{2}$$

$d_j(\vec{m}_{xy})$ is the Euclidean distance of the measurement and the cluster center $\vec{\mu}_j$ of cluster C_j in a Gaussian derivative feature space with 0.0 mean and 1.0 standard deviation.

Figure 2 illustrates the feature extraction process. The top right graph shows the classton channels. Features associated to the particular channel are marked light grey. We observe maps which mark uniform regions, bar like regions or more complex regions such as the eyes. The corresponding clusters are the class-specific feature detectors. Many neighboring pixels are assigned to the same cluster and form connected regions. This is natural, because the local neighborhood of close pixels have a strong overlap, with a high probability that the image neighborhoods are assigned to the same cluster.

4 Cluster quality

Clusters are dense collections of data points. They are useful for classification because they represent a collection of highly similar features. Under the condition that the training images are visually similar, those dense clusters represent the most significant features for the trained image class and allow to learn the variations of these features.

Anyhow, there is an incongruity between the clusters that are automatically computed using density criteria and the feature detectors that we wish to obtain. This is natural, since the clustering associates points only based on distance in feature space. As a consequence, clusters may emerge that group similar features having no semantic meaning. In the following we have formalised additional selection criteria to judge the quality of a cluster. Application of these criteria allow to select the feature detectors that correspond to those features that are focussed by human saccades when presented to a face image as described by Yarbus [19].

In order to judge the quality of a cluster, we discuss the following measures: the compactness and the density in feature space and the average size of the regions in the classton channel (ACCS).

$$\text{Compact}(C_k) \quad = \quad \frac{A}{V} \approx \frac{\prod_{i=1}^{D} \sigma_i}{\max_i(\sigma_i)^D} \tag{3}$$

$$\text{Density}(C_k) \quad = \quad \frac{\text{\# points}}{A} \tag{4}$$

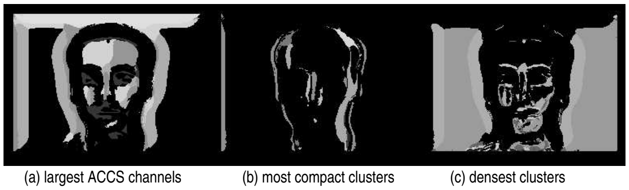

Figure 3: (a) Classton channels producing the largest average components (ACCS measure). (b) The 6 most compact classton channels. (c) The 6 densest classton channels.

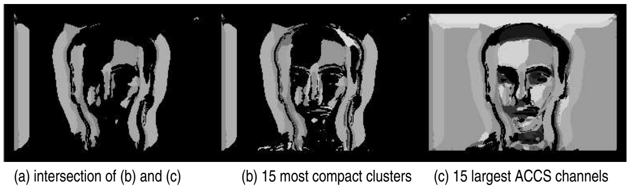
(a) intersection of (b) and (c)
(b) 15 most compact clusters
(c) 15 largest ACCS channels

Figure 4: (a) Combination of compactness and image-based measure. (b) The 15 most compact classton channels. (c) Classton channels with large ACCS.

$$ACCS(I) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} F_i(I) \qquad (5)$$

with $F_i(I)$ size of connected component in image I. ACCS is obtained by averaging the connected regions within the training images. Compactness is defined as the ratio of a volume and the enclosing sphere. In order to compute the compactness of a cluster C_k, we modify the geometrical definition of compactness as follows. The D-dimensional volume A of a cluster is approximated by the product of standard deviation σ_i of its members in each dimension $i = 1, \ldots, D$. The volume of the enclosing sphere V is computed as the maximum standard deviation to the power of D. Cluster density is computed as average number of points per volume unit.

The density of clusters depends on the total number of feature points. For this reason, a threshold for reliable detection of dense clusters can not be found. Compactness has the advantage that it is independent from the number of features, because it takes into account the cluster shape. A generic feature with good generalisation ability produces large connected components in the classton channels and has therefore a large ACCS measure. Figure 3 shows an example of compact clusters producing large connected components that specify forehead, hair, eyes, nose, and mouth region as significant features of faces. For this reason, ACCS is a good measure for the generalisation ability of a cluster.

Examples of classton channels of the different selection criteria are shown in Figure 3. The densest classtons are composed by many points and little variation such as the background. This type of classtons can be useful for figure ground segmentation on uniform background. Compact clusters form nearly spherical point clouds in feature space. Typical examples are very distinct features such as the nose and other bar-like facial features. On the other hand, the form of the eye region cluster has a more complex form in the feature space due to the complex appearances. This motivates the use of an image based measure such as ACCS. The eye region is only detected by this image-based measure. A meaningful facial feature detector is therefore characterised by high compactness and large ACCS. The combination of connected and compact clusters is shown in Figure 4.

5 Modeling spatial relations

The generic features have the property that they can be computed at any image position. They do not contain any notion of image position. The generic features are detected robustly to intra-class variability because clustering is a means to learn this variability from examples. They provide the local measurements needed for a higher level recognition process. Due to their locality, they respond to cluttered background. By taking into account the spatial relations between features, faces can be detected reliably despite background clutter. In this section we propose an automatic model generation that learns spatial relations of generic features. This model is inspired by Belongie's shape context [1].

5.1 Log-polar histograms

A log-polar histogram has bins that are uniform in log-polar space. This corresponds to a linearly increasing positional uncertainty with distance from the reference position $\vec{p} = (x_0, y_0)^T$. This means that the descriptor is more sensitive to measurements at nearby positions than to measurements at image positions farther away. This makes the log-polar description appropriate for applications where the object undergoes affine transformations. It is appropriate for the description of face images and other non-rigid objects, that often have small deformations.

The computation of the log-polar representation is performed in two steps. First, the region around the query position $\vec{p_i}$ is transformed into polar and then log-polar representation according to:

$$\rho = \sqrt{x^2 + y^2}, \;\; \eta = \tan^{-1}\left(\frac{x}{y}\right) \qquad (6)$$

$$\chi = \log_2(\rho), \;\; \gamma = \frac{N_a}{2\pi}\eta \qquad (7)$$

with $(x, y) = (x_0 + \Delta x, y_0 + \Delta y)$ Cartesian coordinates, (ρ, η) polar coordinates, N_a angular resolution and (χ, γ) log-polar coordinates.

The polar representation contains the pixel values of the transformed original image. The so obtained polar image is then sampled uniformly in log space according to Equation (7) in order to fill the histogram. Each histogram cell contains the ratio of the surface covered by the query pixel and the total surface of the histogram bin. The construction

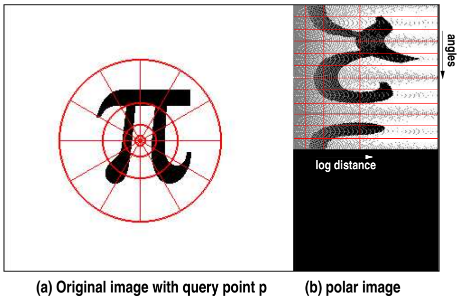

Figure 5: (a) Image in Cartesian coordinates with query point. The range of the log-polar histogram is marked by the large circle. (b) Image (a) transformed to polar coordinates.

of the log-polar histogram is illustrated in Figure 5. A fast implementation uses a lookup table of the direct transformation.

5.2 Learning the spatial relations

Applying a set of k high quality clusters, an image is transformed to k binary images, the classton channels. This is an enormous data reduction, but it preserves the type and the position of the local image feature. This is exactly the information needed for detection. The classton channels serves as input for the log-polar histogram. The log-polar histogram measures the relative positions of the detected classton regions and provides a signature of the spatial relations within a particular range.

For learning the spatial relations of the target object, the user selects a reference position within a set of training images. This is the only user interaction required for training. A model histogram is constructed as the average log-polar histogram of the training histograms extracted at the reference position. In the face image example, we choose the center between the eyes as reference position and computed the average model histogram from 10 training images.

For measuring the similarity between any query histogram Q and the model histogram H we use the χ^2 divergence measure.

$$\chi^2(H,Q) = \sum_i \frac{(q_i - h_i)^2}{q_i + h_i} \tag{8}$$

where h_i is the content of bin i of model histogram H. If the divergence is sufficiently small, the face is detected at the current position. Such detections are marked as circles in the Figures 6 to 10.

The sampling in log distance from the reference point provides increasing robustness in position with increasing distance from the reference point. Scale and pose changes or facial expressions typically produce variations in position of facial features which is compensated by log-polar sampling. An example is shown in Figure 6. The robust modeling is the major advantage of the log-polar histogram approach over other direct modeling methods such as cartesian histograms or learning the spatial relation of a set of facial feature points as described by Wiskott [16] that requires the precise extraction of eyes, nose and mouth corners.

6 Experiments

6.1 The Databases

We use two public face databases, the AR face database [10] and the Caltech face database[1]. Additional experiments are performed on home made digital images of our research group. The images of the AR face database have a resolution of 256×192, and contain a large number of individuals, men and women from different ethnic groups, with and without glasses, different hairstyle or beard. The images show different facial expressions, lighting changes and occlusions. To demonstrate the performance of our method on images with cluttered background, we use the Caltech face database that consists of 435 images of 30 individuals with various background, indoor and outdoor illumination. Some images are underexposed. The images are rescaled such that the head size approximately corresponds to the head size of the AR face database.

6.2 Robustness of generic feature detectors

For constructing the generic facial feature detectors, we use the first 15 neutral faces of men from the AR face database. We use segmentation maps to focus on the object features and speed up the learning process. The segmentation maps are not used for testing. From the 37 k-means detectors we select the 5 classtons that score highest according to ACCS and compactness. Those detectors form 5 classton channels that are combined into a single cluster map representation according to equation 1, where each channels is marked by a different grey value (black means that none of the 5 classton features has been detected). Figure 7 illustrates the responses of the different detectors to faces on cluttered background. We observe only few false positive detections due to background clutter.

Among the 5 channels, we obtain a detector for left side of forehead, cheek and chin, a second symmetric detector for the right side of forehead, cheek and chin. A detector for the regions between the eyes and center part of the chin and forehead. A detector for eyes, that responds also for the mouth region. The last detector is sensitive to bar like structures as the nose.

Figure 8 shows the results of robustness to significant illumination changes. It can be observed, that some facial features are stable even in images with significant illumination changes, and others are not.

Typically, classtons that mark facial features with significant local structure such as the eyes, and mouth region are stable under changing lighting conditions. Those fea-

[1] available at http://www.vision.caltech.edu/html-files/archive.html

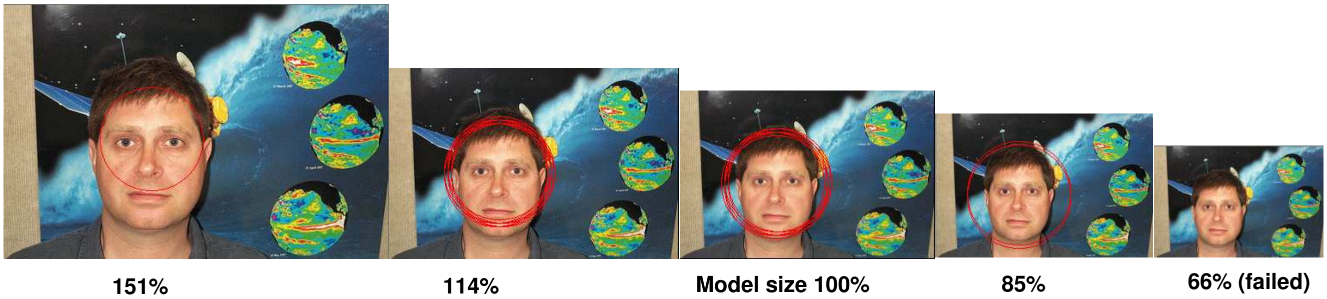

Figure 6: Detection results for artificially scaled images. The detection approach is robust to limited scale changes due to the flexibility introduced by the log-polar histogram.

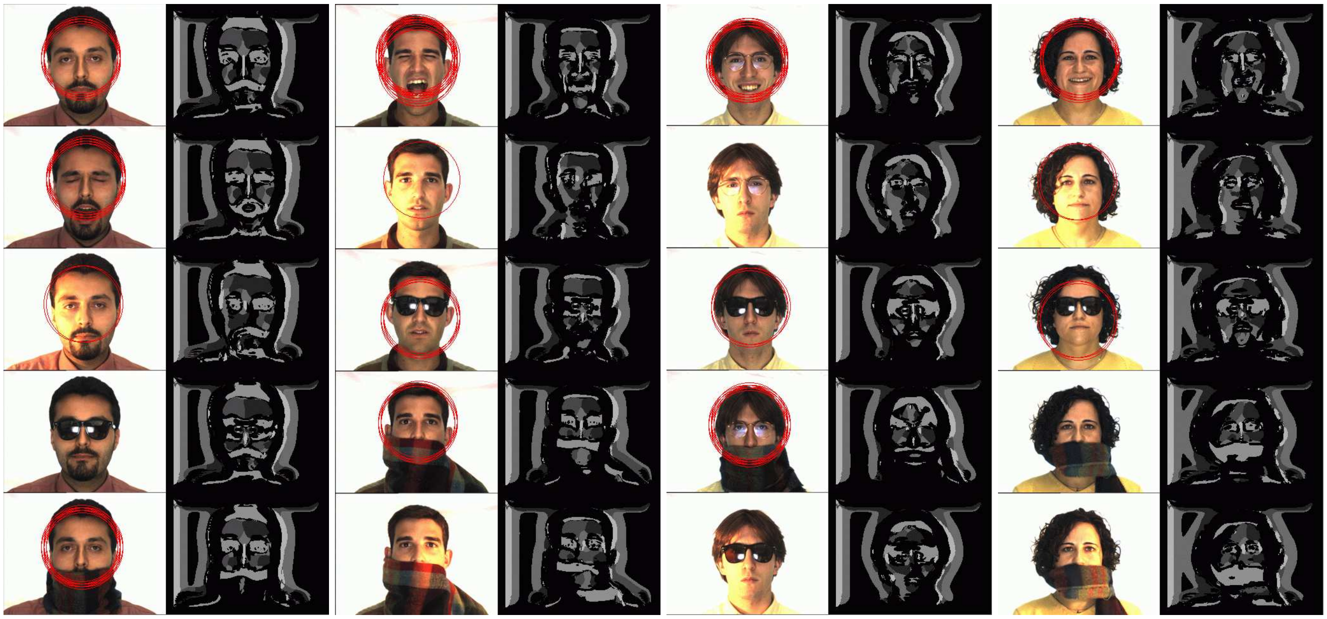

Figure 8: Classton channels of images under significant illumination changes. The classtons are obtained from 15 frontal faces under uniform lighting such as the image in the top row. The facial features are detected reliably. Face detection is successful in images with various facial expressions, strong illumination, and occlusions. It fails in cases with combined strong illumination and occlusions.

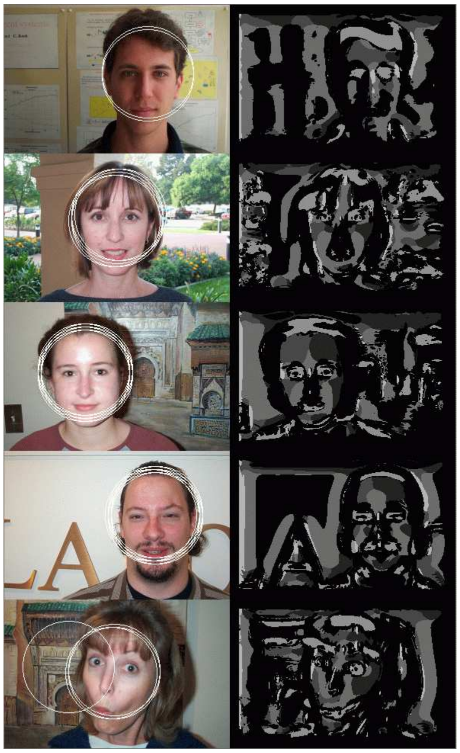

Figure 7: Classton channels of individuals other than the 15 neutral male training faces. All faces are reliably detected using a log-polar histogram, insertions due to background clutter are rare.

ture that display little local structure such as the cheeks, nose and forehead are much more sensitive to illumination changes. In other words, complex features are sufficiently outstanding, such that a change in illumination does less disturb the matching.

6.3 Face detection

The classton channels of the 5 high quality clusters serve as input for the face detection system. Faces show a consistent spatial pattern within the face region. This spatial pattern is modelled by the log-polar histogram. As the feature extraction by classton channels is robust to illumination changes, head pose orientation, small changes in head size, and gender, so is the modeling by log-polar histogram. In addition, the sampling in log-polar space introduces a robustness to position of the facial features that is required for successful face detection robust to facial expressions.

For localisation, the query image is raster scanned with step size 4 pixels. At each grid node the corresponding log-polar histogram is extracted and the divergence measure of this query histogram and the model histogram is computed. If the divergence measure is below a threshold, a detection is registered. Such a successful detection is visualised by a circle in the original image. The radius of the circle corresponds to the range of the log-polar histogram.

In order to show the stability of our approach to images with cluttered background and different illumination conditions, we have performed a face detection experiment on the Caltech face database on 435 images. Examples are shown in Figure 7. A detection system is commonly characterised by two values: how many objects are detected (recall) and how many of the detections are correct (precision).

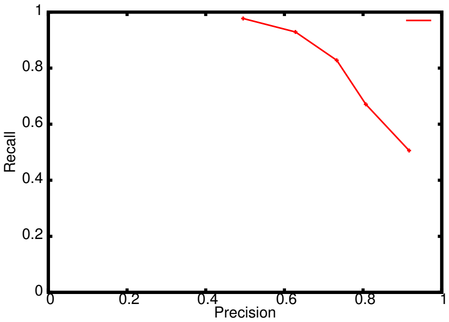

Figure 9: Evaluation of the face detection on the Caltech face database.

$$\text{Recall} = \frac{\#\ \text{correct positives}}{\text{Total}\ \#\ \text{positives in dataset}} \tag{9}$$

$$\text{Precision} = \frac{\#\ \text{correct positives}}{\#\ \text{correct positives} + \#\ \text{false positives}} \tag{10}$$

The detection results can be displayed as a precision recall curve where a parameter, in this case the maximum cost, is varied to obtain several values on the curve. A good detector has high recall and high precision. Figure 9 shows the precision recall curve for the Caltech face database. We obtain a recall of 97.7% (425 out of 435 images are detected). Figure 6 shows the robustness to scale variations without explicitly compensating for scale changes. Our approach can be made scale invariant by a small number of modifications (extraction of the raw Gaussian derivative features and adaption of the histogram range). These modifications are not yet implemented, the robustness is achieved only by the flexibility of the log-polar representation.

We perform a face detection experiment for images from the AR face database showing other individuals than those used for training. We obtained following detection rates for a set of 444 images (Table 1). We provide only recall rates because the precision is not interesting for images with uniform background. The first column shows the detection rate for Cartesian histograms that are inferior to the detection rates of the log-polar histograms. In all cases the divergence measure of Cartesian histograms is higher than the divergence measure of log-polar histograms. Both facts motivate the modeling of spatial relations in log-polar space. We have very good detection results for different facial expressions, different illumination or occlusions. The most detection errors occur for combined occlusion and illumination changes.

Figure 8 shows the robustness to illumination changes and occlusions. Detection fails in cases where significant features such as the nose or the mouth region are not detected

Detection rate using	Cartesian histograms	Log-polar histograms
Expressions	98.8%	99.3%
Illumination OR occlusion	91.4%	97.8%
Illumination AND occlusion	54.0%	69.4%
Total	84.5%	91.2%

Table 1: Face detection rate for AR face images under different conditions. We have very good detection results for different facial expressions, different illumination or occlusions. The most detection errors occur for combined occlusion and illumination changes. The precision is high, because the images have uniform background.

by the corresponding classton. This is the case in occluded images or in images with extreme lighting conditions. In cases where a high number or the facial features are detected, the divergence measure allows a successful detection.

The log-polar implementation allows the modelling of spatial relations that is sufficiently discriminant to avoid false detections and is general enough to avoid over-fitting. In the example of unconstrained images in Figure 10 all faces are correctly detected. The large amount of false detections by the generic detectors all over the background are successfully discarded by the spatial relation constraint imposed by the log-polar histogram. This is a convincing result considering the variations in scale, head pose and lighting.

7 Conclusions

In this article, we propose an approach for learning of local coefficients for the construction of detectors for common features of a visual class. In order to obtain robustness to illumination changes, the feature detectors are computed from luminance images, since the luminance channel is less affected by illumination changes than the chrominance channels. In order to obtain automatically those clusters that correspond to meaningful features, we develop a measure to judge the quality of each cluster. For our training images of the AR face database, the application of the quality measure selects facial feature detectors that correspond to those features that are preferred by humans, as observed in psychophysical experiments.

The local facial features are detected robustly to intra-class variability and serve as input to a module that measures the spatial relations. Using a log-polar histogram, the obtained model is sufficiently discriminant to provide reliable face detection. Face detection fails only in cases where important features are missed by the detectors.

In the experiments we demonstrate the stability of the facial feature detectors with respect to person identity, lighting changes, different facial expressions, occlusions and

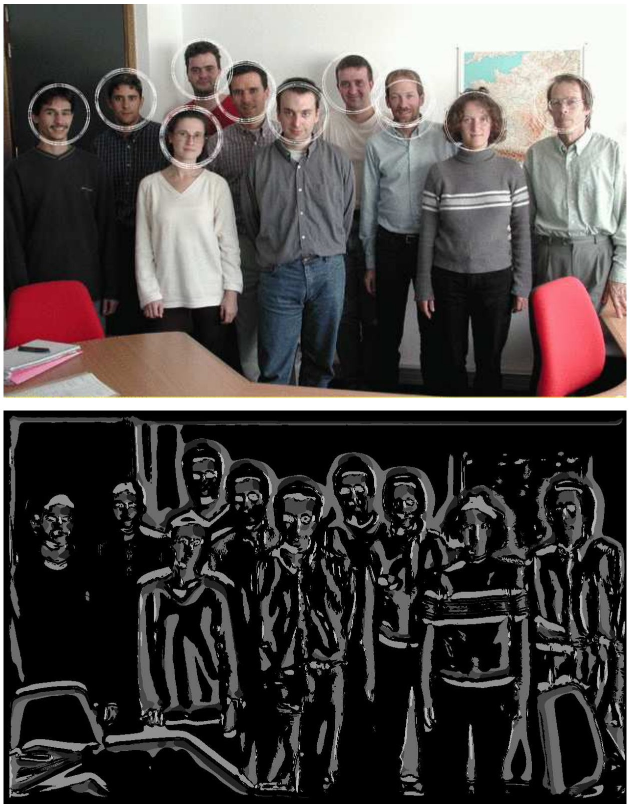

Figure 10: Classton channels of unconstrained image. The training is performed on frontal faces from the AR database. Detected faces marked by circles are characterised by the combined occurrence of facial features. No false detections are observed.

cluttered background. The detectors generalise well to unknown faces, and are robust to gender and facial expressions. The combination of strong side illumination and occlusion disturbs the characteristic face pattern which make detection more difficult. Reliable face detection is possible on images with cluttered background and small changes in head size because the log-polar histogram representation allows to be insensitive to facial feature detections in the background by taking into account the spatial relations.

These results are a step towards the construction of a robust recognition system that can learn and model any visual image class. The log-polar histogram approach is one possibility among others to learn spatial relations. The advantages are clear. The modeling avoids one to one matching, provides sufficient discriminance for reliable recognition and at the same time is robust to position changes of distant feature points. Furthermore, it is a straight forward approach that requires little supervision.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

[2] D. Hall, V. Colin de Verdière, and J.L. Crowley. Object recognition using coloured receptive fields. In *European Conference on Computer Vision*, pages I 164–177, Dublin, Ireland, June 2000.

[3] D. Hall and J.L. Crowley. Computation of generic features for object classification. In *Scale Space Methods in Computer Vision*, pages 744–756, Skye, UK, June 2003.

[4] G.J. Klinker, S.A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 1990.

[5] J.J. Koenderink and A.J. van Doorn. Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):597–605, June 1992.

[6] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. In *International Conference on Computer Vision*, Corfu, Greece, September 1999.

[7] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[8] D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

[9] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, June 2001.

[10] A.M. Martinez and R. Benavente. The ar face database. Technical Report 24, CVC, June 1998.

[11] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence*, 78(1–2):461–505, 1995.

[12] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, January 2000.

[13] C. Schmid. Constructing models for content-based image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, USA, December 2001.

[14] C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.

[15] K. Schwerdt and J.L. Crowley. Robust face tracking using color. In *International Conference on Automatic Face and Gesture Recognition*, pages 90–95, Grenoble, France, March 2000.

[16] L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Mahlsburg. *Face Recognition by Elastic Bunch Graph Matching*, chapter 11, pages 355–396. Intelligent Biometric Techniques in Fingerprint and Face Recognition. CRC Press, 1999.

[17] J. Yang and A. Waibel. A real-time face tracker. In *Workshop Applications of Computer Vision*, pages 142–147, 1996.

[18] M.-H. Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, January 2002.

[19] A.L. Yarbus. *Eye Movements*. Plenum Press, 1967.

[20] S.-C. Zhu, C. Guo, Y. Wu, and Y. Wang. What are textons? In *European Conference on Computer Vision*, pages IV 793–807, 2002.