accepted EANN'97, Stockholm, 1997

Invariant classification of image parts using a dynamic grid of point representations

R. Herpers[1] D. Müller[1] M. Michaelis[1] G. Sommer[2]

[1] GSF – Institut für Medizinische Informatik und Systemforschung, MEDIS
Ingolstädter Landstr. 1, 85764 Oberschleißheim, Germany, Email: herpers@gsf.de
[2] Institut für Informatik, Christian-Albrechts-Universität
Preußerstr. 1-9, 24105 Kiel, Germany

Abstract

We present an application of an elastic graph matching approach to classify facial image regions. In contrast to the dynamic link architecture introduced by the Malsburg group in our application not an identification task has to be solved but a classification task. Therefore, our approach differs in many important aspects: (1) the choice of the filter set, and (2) the selection of the positions of the nodes of the graph to represent the characteristic image information, (3) the generation of a representative reference pattern needed for the calculation of the classifications and (4) a new two step graph matching approach based on the simulated annealing technique. The approach is tested on facial regions taking the eye region as an example. A classification performance for the verification of eye regions of more than 97 % can be achieved.

1 Introduction

Face processing and in particular face recognition became an important research field in computer vision. A popular example, which applies artificial neural network techniques, is the elastic graph matching approach of the Malsburg group [1, 5].

In the contribution presented here a verification module is proposed, which is partially motivated by von der Malsburgs' approach. In contrast to their identification task the goal presented in this contribution is the classification of different facial image parts such as the eye region. Therefore, our approach differs in many important aspects: in the choice of the filters used to represent the characteristic image information, in the selection of the positions for the nodes, in the generation of a reference pattern, and in the application of a matching algorithm.

The design of the model graph used is the result of several investigations comparing the performance of the classification using different positions. The image positions of the nodes of the graph used for our object representation are located only at specific keypoint positions of the objects (e.g anatomical landmarks). This results in a special graph, which is adapted to the characteristic structures of the objects to be classified. This procedure avoids the drawback of homogeneous grids, where many nodes may lie at positions with unspecific local image structures.

For the class of region to be classified a synthetic reference pattern is generated. This is necessary because for a classification task an unknown image region cannot be compared with an individual reference pattern as it may be possible for an identification task.

The classification is performed by matching an individual pattern to a reference pattern using a two step graph matching approach. In the first step the whole image region considered is scanned using an undistorted graph to find an optimal starting position. In the second step all nodes are moved independently, restricted only by some constraints which avoid too strong distortions. The optimal match is obtained applying a simulated annealing approach. Applying these methods a classification performance for the verification of eye regions in a set of arbitrary facial regions of more than 97 % can be achieved.

[1]This work is partially supported by DFG grant So 320/2-1.

2 Labeled graph

The facial regions which are to be classified are represented by a labeled graph. The labeled graph $\mathcal{G} := \{(\mathcal{N}, \mathcal{K})\}$ is defined as a set of nodes $\mathcal{N} := \{N_1, \ldots, N_J\}$ and a set $\mathcal{K}$ of edges, where each member or element $K_{j,k}$ denotes the edge between node N_j and node N_k $(j \neq k)$. Each node N is labeled by its position $\vec{x}$ and a S-dimensional label L, $N := \{(\vec{x}, L) | \vec{x} = (x, y), L = l_1, \ldots l_S\}$. Each component $l_i, i = 1, \ldots, S$ of the label L is the response (projection coefficient) of the i'th filter from a given set of filters (fig. 1).

Several polar and cartesian separable filters based on Gaussian derivatives have been investigated to find an optimal filter set for the representation of local image structures. It turned out that polar separable filters were superior to cartesian ones. Also filters with too many radial or angular oscillations decrease the representation performance.

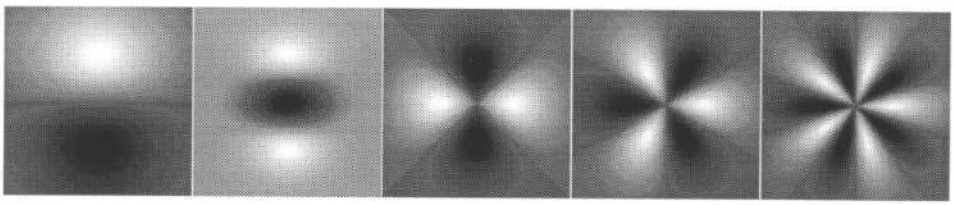

Figure 1: Set of filters which are applied to describe the local image structure at the position of a node. Left: edge and line detection filters ($\sigma = 6$, template size = 27×27 pixel). Right: polar separable filters with different angular oscillations ($m = 2$, $m = 3$ and $m = 5$ respectively). For each filter also the orthogonal partners are applied (not depicted).

The fundamental idea of our approach is that the nodes are positioned only at image positions in the image domain, for which evidence in the local structure exists (fig. 2). The underlying ensemble of edge and line structures are represented applying the filter set just mentioned. The underlying assumption is that the encoded edge and line structures are characteristic for the entire class of objects to be classified.

Several nodes of the graph are connected whereby multiple edges between the nodes are necessary to ensure the maintenance of the spatial relationship between the nodes (fig. 2). The distance between a node N_j and a node N_k is defined as the euclidean distance between their positions $\vec{x}_j$ and $\vec{x}_k$; $d_{j,k} := |\vec{x}_j - \vec{x}_k|$.

A synthetic reference pattern $\mathcal{G}^R$, which is calculated from an average of several examples, is introduced to provide a representative model. We found, that superpositions of several individual local structures such as eye corners result in a more robust reference pattern than a single individual one. Therefore, a number of examples are selected to generate a representative reference pattern, which describes the common structure of all of the class members.

$$\hat{L}_j^R(\vec{x}) := \frac{1}{n} \sum_{h=1}^{n} L_{j,h}^R(\vec{x}) \tag{1}$$

Here n is the number of examples of corresponding labels $L_{j,h}^R(\vec{x})$ of the reference pattern related to a particular kind of landmark j (fixed). To establish also reliable reference values for the edges between the nodes of the model graph, distance ranges with a mean value and a standard deviation are introduced, which are also based on the same set of n examples.

To quantify the similarity between corresponding labels $\hat{L}_j^R$ and L_j^T the normed scalar product of the two feature vectors is calculated as follows:

$$(\hat{L}_j^R, L_j^T) := \; 1 - \frac{\langle \hat{L}_j^R, L_j^T \rangle}{\|\hat{L}_j^R\| \cdot \|L_j^T\|} \tag{2}$$

where the index R stands for the representation of the reference pattern and the index T marks a representation of a test pattern.

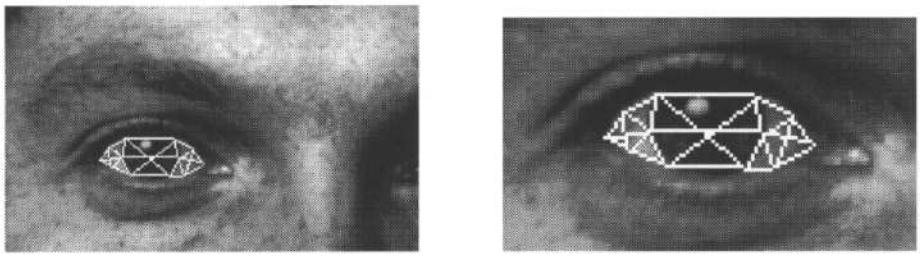

Figure 2: Adapted graph for the class of eye regions demonstrated at one sample eye region. *Left:* the original image part with the model graph. *Right:* the enlarged graph of the eye to document the spatial relationship between the nodes.

3 Graph matching

To match an individual pattern G^T to the reference pattern G^R a two step graph matching approach has been developed. In the first step the initial position of the undistorted graph is calculated using the entire image region as search area. Beside the compensation of the translation of the pattern to be located an adaptation of the scale of the model graph is also realized. The computation of a reliable starting position for the subsequent fine-matching of the model graph to the individual image structures is necessary to ensure a high quality classification result (fig. 3). In a second step each node of the graph is moved independently to achieve an optimal correspondence to the reference node. For this optimization problem, a simulated annealing approach [4] is applied to compute the best match of the model graph to a test pattern. For each node N_j an alternative position $\vec{x}_j(t+1)$ is selected randomly and the costs of the entire graph $C_g(\vec{x}_j(t+1))$ are calculated. The difference $\Delta C(t) := \hat{C}_g(\vec{x}_j(t+1)) - C_g(\vec{x}_j(t))$ between the expected costs $\hat{C}_g(t+1)$ of a suggested position $\vec{x}(t+1)$ and the most recent previous position $\vec{x}(t)$ determines the acceptance of the suggested position. The introduced cost function $C_g(t) := C_n + \lambda C_c$ is a quantitative measure of the quality of the calculation of the match, where λ is a weighting factor ($\lambda \geq 0$) to control the quality of the matching process. Lower values for λ will result in strongly distorted graphs while higher values will maintain the global relationship between the nodes. The total cost $C_g(t)$ of a graph of a test pattern at a particular time step t has two components. The first component $C_n(t)$ quantifies the similarity of each label of a node of a graph in comparison to the reference description based on the similarity function (see formula 2). The second component $C_c(t)$ quantifies the changes in the spatial relationship between the nodes. The calculation of the cost function $C_c(t)$ is based on the distance of the actual edge $d^T_{j,k}(t)$ of the test pattern G^R relative to the distance $d^R_{j,k}$ of the corresponding edge in the reference graph G^T.

$$C_n(t) := \sum_{j=1}^{J} S(L^R_j(\vec{x}), L^T_j(\vec{x}, t)) \quad , \quad C_c(t) := \sum_{j=1}^{J} \frac{1}{U_j} \sum_{k=1}^{U_j} f\left(\frac{d^R_{j,k} - d^T_{j,k}(t)}{d^R_{j,k}}\right) \tag{3}$$

where U_j is the number of edges concerning node j and f is a monotonous increasing function (e.g. $f(z) := z^2$).

To enable also an acceptance of states with higher energies during the adaptation procedure, the decision of acceptance of a suggested position is realized using a probability function. For $\Delta C > 0$ the probability of acceptance should be lower (but not equal to 0) than for $\Delta C \leq 0$. The probability of acceptance p is realized by applying a sigmoid function, which depends on the cost difference ΔC and the temperature τ.

$$p(\Delta C(t), \tau) := 1 - \frac{1}{1 + e^{-\frac{\Delta C(t)}{k\tau}}} \tag{4}$$

where t is a particular time value and k is a constant factor to control the relationship between the cost difference ΔC and the temperature τ.

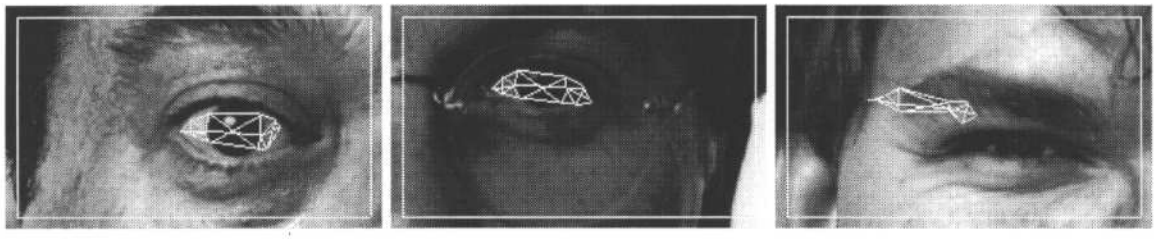

Figure 3: Results of the local matching algorithm. For eye regions the local adaptation enhances the similarity between the several point representations if the graph is positioned correctly on the eye. For non eye regions the spatial relations will be distorted more severely, which causes high costs for the connections $C_c(t)$.

4 Discussion and conclusion

Classifications of particular keypoint positions were obtained by calculating point descriptions as reported in [3]. Based on these results in this contribution a **set** of image positions are taken to represent an extended complex image region containing an entire facial part of interest [2]. The presented approach is partially motivated by the dynamic link architecture developed by the Malsburg group to identify individuals [1, 5]. In contrast to their approach in our application not an identification task has to be solved but a classification task, which is more complex concerning the variability of patterns to be considered. Therefore, a new processing strategy has been introduced to represent the characteristic information of the entire class of objects. A representation of a class of images must have a more global character than in the case of a representation of particular images of several individuals.

Another novelty of our approach is that the characteristic image structure is represented by applying a special filter set, which encodes more reliably the underlying ensemble of edge and line structures to be represented. In the Malsburg approach Gabor wavelets are applied, which are in particular suited to represent textured image structures. In contrast to this procedure, we apply a set of Gaussian filters, which are more suited to represent line and edge structures [6] and which provide a more invariant representation of the possible appearances of individual patterns of the class. Caused by the special preconditions given by our task several methodical developments of the referred approach were necessary to tackle successfully our application. This has concerned the selection of filters used for the point representations, the topology of the graph adapted to the object and the selection of the reference pattern, which has to represent the entire class of images. Finally, the graph matching approach has been modified to meet the requirements given by the distinct precondition.

References

[1] J. Buhmann, J. Lange, and C.v.d. Malsburg, *Distortion invariant object recognition by matching hierarchically labeled graphs*, In: Proc. Int. Joint Conf. on Neural Networks, IJCNN, IEEE, pp. 155-159, 1989.

[2] R. Herpers et al., *GAZE: An attentional processing strategy to detect and analyze the prominent facial regions*, In: Proc. of the Int. Workshop on Autom. Face- and Gesture-Rec., M. Bichsel (ed.), Zurich, Switzerland, pp. 214-220, 1995.

[3] R. Herpers et al., *Evaluation of local image structures applying a DCS network*, In: Solving Engineering Problems with Neural Networks, Proc. of the 2nd int. Conf. EANN96, London, A.B. Bulsari et al. (eds.), pp. 305-312, 1996.

[4] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Optimization by simulated annealing*, Science, Vol. 220, pp. 671-680, 1983.

[5] Lades M. et al., *Distortion invariant object recognition in the dynamic link architecture*, IEEE Trans. on Computers, Vol. 42, pp. 300-311, 1993.

[6] M. Michaelis, *Low level image processing using steerable filters*, PhD thesis, Christian-Albrechts-Universität, D-24105 Kiel, Germany, 1995.