Chapter 1

Multi-Path Marginal Space Learning for Object Detection

1.1 Abstract

This chapter introduces a novel method for fast detection of objects with a large number of parameters. The method is based on Marginal Space Learning (MSL), which is a learning-based optimization technique that approaches the search for objects in images as a particle filter in a chain of subspaces of increasing dimensions, using trained detectors to prune the particles in the subspaces. MSL has been used extensively in Medical Imaging for detecting organs and landmarks in 2D and 3D data and for detecting and tracking curve-like structures such as guidewires and catethers. This chapter brings three contributions. Firstly, it introduces multiple computational paths for MSL, which can improve the detection performance compared to a single MSL path. Second, it presents an application of multi-path MSL to four parameter face detection from grayscale images. Thirdly, it observes experimentally that multiple-path MSL obtains a compact classifier with good generalization abilities. Consequently, the number of training examples can be reduced to half compared to other methods with similar performance.

1.2 Glossary

- classifier. A function that takes one or more variables called features as input and returns a discrete class label (e.g. object/non object) or a probability over the possible class labels.
- detection rate. The percentage of true positives (e.g. faces) that were correctly detected by an object detection algorithm.

- false positive rate. The percentage of detections of an algorithm that are not close to true positives (e.g. faces).
- Haar feature. A function that takes an input image and returns a linear combination of sums of intensities inside different rectangles. The computation of sums inside a rectangle can be done in constant time using the integral image.
- integral image. An image as large as the original image, containing at location (x, y) the sum of all pixel intensities in the rectangle from $(0, 0)$ to (x, y).
- Receiver Operating Characteristic (ROC) Curve. A curve displaying the detection rate vs. the false positive rate of an algorithm when a detection parameter (usually a detection threshold) is varied.
- strong classifier. A classifier that is constructed from a number of weak classifiers or weak regressors and is much better than any of them.
- weak classifier. A classifier that has a non-zero correlation with the class label, hence it is better than random guessing.
- weak regressor. A function that takes one or more variables as input and returns a continuous value that has a non-zero correlation with the value of interest (e.g. age or GPA score).

1.3 Introduction

One of the main computational challenges in object detection is dealing with the size and position variability of the objects in real-world images. There are even more challenges, however, since the objects also exhibit different rotations, and other parameters (out of plane rotation, illumination, internal parameters such as limb locations for pedestrians, etc). Previous works deal with this curse of dimensionality in different ways, which can be grouped in two main approaches.

The first main approach is to use features that are invariant to the object parameters (Forsyth, Mundy, Zisserman, Coelho, Heller & Rothwell 1991, Wood 1996, Zisserman, Forsyth, Mundy, Rothwell, Liu & Pillow 1995), such as rotation/scale invariant features (Fergus, Perona & Zisserman 2003) or illumination invariant features (Chen, Belhumeur & Jacobs n.d., Slater & Healey 1996). With this approach, some of the object parameters are ignored and a computational gain is obtained. The main challenge with this approach is to find features that are invariant yet discriminative. In some cases (Fergus et al. 2003), these features are more computationally expensive than the simple and non-invariant ones, such as the Haar features (Viola & Jones 2004), that are used in the second main approach.

The second main approach is to exhaustively search the object parameters using fast classifiers based on simple features (Fleuret & Geman 2001, Heisele, Serre, Prentice & Poggio 2003, Li, Zhu, Zhang, Blake, Zhang & Shum 2002, Roth, Yang & Ahuja 2001, Sung & Poggio 1998). Different approaches are used for speeding up the detection (Viola & Jones 2004) and for obtaining more discriminative features (Schneiderman & Kanade 2000, Wu, Rehg & Mullin 2003). For computational efficiency, most exhaustive search approaches are based on a cascade of increasingly complex classifiers, and still make a compromise by ignoring some of the object parameters, e.g. rotation.

Recently, a new computational approach named *Marginal Space Learning* (MSL) permits the use of an arbitrarily large number of parameters in the object of interest. The method was applied to many medical imaging problems (Barbu, Athitsos, Georgescu, Boehm, Durlak & Comaniciu 2007, Barbu, Suehling, Xu, Liu, Zhou & Comaniciu 2011, Feng, Zhou, Good & Comaniciu 2009, Feulner, Zhou, Huber, Hornegger, Comaniciu & Cavallaro 2010, Ling, Zhou, Zheng, Georgescu, Suehling & Comaniciu 2008, Seifert, Barbu, Zhou, Liu, Feulner, Huber, Suehling, Cavallaro & Comaniciu 2009, Zheng, Barbu, Georgescu, Scheuering & Comaniciu 2007) where the objects to be detected had between 9 and 165 parameters. The method involves training classifiers for a sequence of increasingly larger subspaces, such that the relative dimension between two consecutive spaces is small. These subspaces are Marginal Spaces, since some of the parameters of the final classifier have been ignored (marginalized). In (Zheng et al. 2007), all 9 parameters were needed to align a PCA shape model of a heart chamber for object segmentation.

The contributions of this chapter are the following:

1. It makes a connection through Marginal Space Learning between the object detection approach based on invariant features and the approach based on simple and non-invariant features with exhaustive search.

2. It introduces multiple computational paths in Marginal Space Learning, in which detections obtained through different MSL paths are aggregated into the final detection result. This new method offers improved performance over using a single MSL path.

3. It presents an application of the proposed multi-path MSL approach to four parameter face detection from grayscale images.

4. It observes experimentally that MSL and multi-path MSL have more compact classifiers than regular object detectors of comparable accuracy. Consequently, MSL and multi-path MSL need fewer manually annotated examples than other object detection algorithms (e.g. 1494 instead of 4916 in (Viola & Jones 2004)) to obtain similar or even better generalization performance.

This work does not aim to go beyond the state of the art in face detection. Instead, it shows a new approach to object detection using Marginal Space Learning that can handle a large number of parameters, obtains a more compact model and needs fewer training examples. These advantages make the approach applicable to many object detection tasks.

1.4 Related Work

The fast detection in a Marginal Space can be considered one form of *selective attention* (Amit & Geman 1999). Furthermore, the three levels of computation presented in (Amit & Geman 1999) - edge fragments, local and global groupings - represent particles in different marginal spaces, when the object of interest is modeled using edges and their spatial relationships. Similarly, in our previous work on Hierarchical Learning of Curves (Barbu et al. 2007), Marginal Space Learning was used to detect flexible curves in X-ray images using ridge fragments, short curves and long curves as marginal spaces.

Marginal Space Learning is related to the Highest Confidence First (HCF) algorithm (Chou & Brown 1990) in that it propagates the most promising partial solutions. However, HCF is greedy while MSL propagates a number of partial solutions, so it can avoid many local optima. Moreover, learning the marginal space models ensures that there will usually in the end be solutions close to the true optimum.

Learning in marginal spaces has been used for learning-based edge detection (Dollar, Tu & Belongie 2006), by learning the probability of a pixel to be on an edge while ignoring the edge direction. However, that was the final goal in (Dollar et al. 2006) and the process was not continued by increasing number of parameters to obtain a more accurate edge model.

Skin detection (Jones & Rehg 2002, Wang & Chang 1997) is used as a first step in Color Face Detection (Hsu, Abdel-Mottaleb & Jain 2002). Skin detection can be seen as a detection step in a marginal space where all other face parameters are ignored except for the skin position, obtained purely based on skin color.

More generally, Marginal Space Learning is related to part based object detection (Agarwal & Roth n.d.) and Pictorial Structures (Felzenszwalb & Huttenlocher 2005, Andriluka, Roth & Schiele 2009), since object parts are detected in smaller dimensional Marginal Spaces of the full object parameter space. There are many differences though. In these works, the full object model is constructed from the part models purely based on the geometric configuration of the parts whereas in MSL the full object model is learned using both spatial and appearance features. Moreover, in MSL many intermediate subspaces could be used between the part models and the full-object model, which could further speed-up the algorithm. One advantage of the Pictorial Structures is that they are robust to the occlusion of any of the parts, as long as sufficiently many other parts are present, while MSL relies too heavily on one of the parts. This issue is addressed by the Multi-Path MSL introduced in this work in Section 1.7.

The And-Or graph representation of the object by parts (Wu & Zhu 2011) has α and β inference processes that detect the objects directly (α process) or predict the object from a detected part (β process). These inference processes can be considered different MSL computational paths. While the And-Or graph is focused on a part-based representation, MSL can use other computational paths that are not based on parts. For example in our heart segmentation work (Zheng, Barbu, Georgescu, Scheuering & Comaniciu 2008) one of the marginal spaces for detecting the Left Ventricle (LV) was the position of the LV center, which is not a part, but a simpler representation of the object in which the orientation and scale are ignored.

Many face detection papers (Bourdev & Brandt 2005, Roth et al. 2001, Schneiderman 2004, Viola & Jones 2004) ignore some of the face parameters, training a classifier that is invariant to the ignored parameters. The same holds true for many object detection papers (Fei-Fei, Fergus & Perona 2006, Schneiderman & Kanade 2000, Torralba, Murphy & Freeman 2004). Because of the high computational burden associated with inferring shape, 3D position, deformation, and illumination, most authors disregard the large parameter spaces that would offer a more accurate characterization of the detected object. However, the benefits in terms of accuracy of the object model are multiple. For example, in Blanz & Vetter (1999) an accurate face model with more than 200 parameters is obtained in about 50 minutes using variational techniques.

Other authors describe a face detection method in which a number of more and more specialized detectors are learned in a tree structure (Huang, Ai, Li & Lao 2005). Specifically, a generic face detector is trained at the root of the tree in the marginal space of all face rotations. In subsequent layers of the tree, more and more accurate detectors are employed to prune the search space. This work can be viewed as an early MSL precursor, however the authors did not present their approach as a general learning-based optimization methodology.

Marginal Space Learning is different from a detector cascade (Schneiderman 2004, Viola & Jones 2004). In the detector cascade, all detectors work in the same parameter space, whereas in Marginal Space Learning the models (detectors) work on spaces of increasingly larger dimensionality. Marginal Space Learning draws its power from this increasing dimensionality, since rejecting one location in a marginal space virtually eliminates thousands or even millions of locations from searching in the full parameter space.

The Soft Cascade (Bourdev & Brandt 2005) could be used to train each marginal classifier instead of the regular cascade or the Probabilistic Boosting Tree (Tu 2005), methods currently used in our work. This way, each marginal classifier can be further tuned to balance speed and accuracy.

Recent work on Recursive Compositional Models (RCM) (Zhu, Chen, Ye & Yuille 2008, Zhu, Chen & Yuille 2007, Zhu, Chen & Yuille 2009, Zhu, Lin, Huang, Chen & Yuille 2008) is related to Marginal Space Learning, since partial object models are obtained in a sequence of subspaces of increasing dimension, and these models are used to efficiently propagate a set of particles. However, Marginal Space Learning is a general training and optimization methodology that can be used in many applications (object detection, segmentation, inference) and is not tied to a particular model.

Marginal Space Learning is similar to a degenerate decision tree (Kuncheva 2004), where each node is a boosted classifier trained in a marginal space. The major difference is that MSL obtains a fractional number of feature evaluations per location, (e.g. on the order of 10^{-5} for Left Ventricle detection in CT), whereas in the decision tree, the number of feature evaluations per location is at least 1 (the first node of the tree). Moreover, the MSL classifiers are tuned based on the ROC curve to certain detection rates that obtain a desired trade-off between the overall detection speed and accuracy. Because of the simplicity of the MSL chain, this tuning can be performed jointly on all the classifiers, which is impractical in a generic decision tree.

As already mentioned, the MSL optimization can be performed stochastically using sequential Monte Carlo on the marginal spaces (Doucet 2007). However, in Doucet (2007) the authors use a generic approach without learning the marginal probabilities, which is not practical for most computer vision or medical imaging applications.

1.5 Marginal Space Learning Overview

Given a trained classifier $p(\mathbf{x}|I), x \in \Omega$, the object detection problem is to find the object parameters $\mathbf{x}$ in the image I

$$\hat{\mathbf{x}} = \operatorname*{argmax}_{\mathbf{x}\in\Omega} p(\mathbf{x}|I)$$

If the parameter space Ω is high dimensional (e.g. 9D in the case of finding the Left Ventricle from a CT image), it is computationally expensive or even prohibitive to search the entire space Ω, even using a coarse-to-fine approach.

Marginal Space Learning addresses this optimization problem through learning. It is a simple and intuitive idea that can be regarded as a particle filter in a sequence $\Omega_1 \subset \Omega_2 \subset ... \subset \Omega_n = \Omega$ of increasingly larger subspaces, with the last space Ω being the parameter space of the object that needs to be detected. In each subspace $\Omega_i \subset \Omega$ some of the object parameters are ignored (marginalized) and a trained classifier is used to prune the particles. The particles are then propagated to the next space of the sequence by adding more parameters with all the possible values on a grid, and again pruning them with the associated classifier, as illustrated in Figure 1.1.

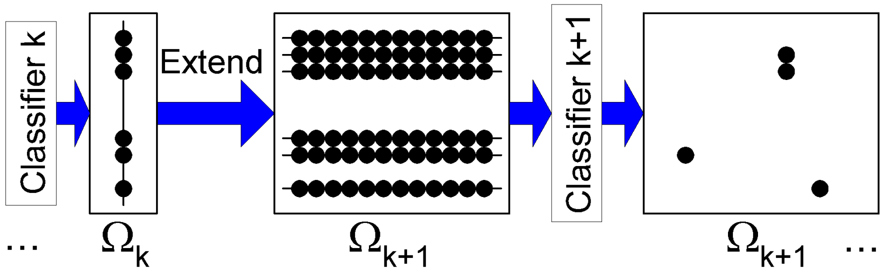

Figure 1.1: Marginal Space Learning propagates a set of particles in a sequence of increasingly larger subspaces of object parameters. In each subspace Ω_{k+1}, the particles from Ω_k are extended by adding more parameters and a trained classifier is used to prune these extended particles.

The procedure is repeated until the set of particles reaches the full space Ω of object parameters. The particles can be pruned in a deterministic fashion by always keeping the most promising particles, or stochastically by sampling them according to their probabilities given by the classifier.

The first or first few marginal spaces can be considered as following the invariant-feature based approach, since they ignore many of the object's parameters. Example of first marginal spaces could be the space of decisions whether the object is present or not in the scene, as in (Fei-Fei et al. 2006), or the space of the object's position only (Fleuret & Geman 2001).

The next marginal classifiers add more and more parameters to the object, and use features that are less invariant and usually faster and more discriminative. Directly using a classifier in large dimensional space can computationally expensive or even prohibitive. However, by using the marginal classifiers to focus the object detectors to a small number of locations, a large number of parameters can be easily handled without a computational burden.

The last classifier is based on all the parameters that are relevant to the problem at hand (scale, rotation, illumination, etc) and gives the final detection result. This last level could have complex and accurate models (e.g. generative models such as Liu (2003) and Tu (2007)) in the last classifier as a final validation step to further improve the accuracy, with minimal computational loss. For example, this last verification step for 3D lymph node detection (Barbu, Suehling, Xu, Liu, Zhou & Comaniciu 2010, Barbu et al. 2011) is performed in the 165 dimensional space of lymph node segmentations.

Marginal space learning presents some advantages and disadvantages:

- If the marginal classifiers can be trained well, speed-up by many orders of magnitude (six order of magnitude speedup in (Zheng et al. 2007)) can be obtained with virtually no loss in accuracy.

- It is easy to control the speed/accuracy trade-off through the number of particles that are propagated. A larger number of particles means an increased detection rate for a larger computational expense.

- The total size of the classifiers is smaller than in a cascaded approach, because each classifier is trained on a more representative sample set. Consequently, the MSL approach usually needs fewer training examples for the same generalization power.

- Training the marginal classifiers, especially the first one, require invariant features that are discriminative. This is the same challenge faced by the invariant-feature based approaches (Chen et al. n.d., Fergus et al. 2003, Forsyth et al. 1991, Slater & Healey 1996, Wood 1996, Zisserman et al. 1995).

1.6 Face Detection with Marginal Space Learning

Face detection is one application well suited for Marginal Space Learning. There are many different marginal subspaces that could be chosen as part of the MSL chain. Some of them correspond to different face parts such as left or right eye or ear, nose, or mouth. Other marginal spaces could be simpler representations of the whole face, such as a low resolution face patch. In this application, we will use a 2-space chain $\Omega_1 \subset \Omega$, where Ω_1 is the marginal space of possible right eye positions with a coarse scale (x^e, y^e, s^e). The space Ω is a four-dimensional space $(x, y, s, \theta) \in \Omega$ parameterizing face position (x, y), fine scale s and orientation θ. The right eye was chosen because the eye is a very salient part of the face. By training different face part detectors (eyes, nose, mouth, ears), we observed that the eye detector has a more compact classifier than the others, so it is more suited to be an intermediate subspace in the MSL chain. The diagram of the face detection application is shown in Figure 1.2.

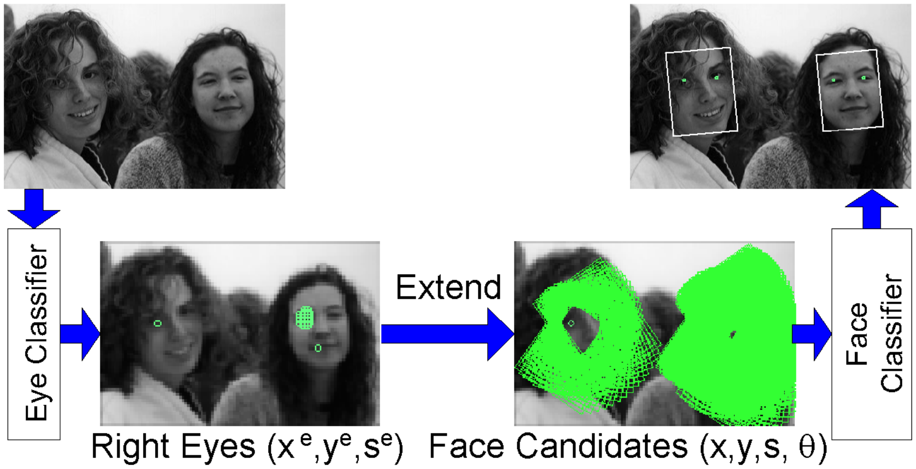
Eye Classifier
Right Eyes (x^e,y^e,s^e)
Extend
Face Candidates (x,y,s, θ)
Face Classifier

Figure 1.2: Example of MSL for face detection. The first marginal classifier detects the person's right eye (x^e, y^e, s^e), ignoring the rotation and with a rough scale. The detected eyes are transformed into face candidates (x, y, s, θ) by adding rotations and scales on a grid. These candidates are pruned by the face detector to obtain the final result.

The eye marginal space consists of the right eye locations (x^e, y^e) for faces in a certain range of scales. Because of that, the eye locations are detected on a Gaussian pyramid (Burt & Adelson 1983) hence they also have a coarse scale s^e that is a power of 2.

The final detector detects faces with four parameters (x, y, s, θ). It does so by using the candidate eye locations (x^e, y^e, s^e) returned by the the marginal classifier and adding a many possible rotation and fine scale parameters relative to the right eye. The classifier trained in this space will output a number of detected faces, which will be the result of our algorithm after non-maximal suppression.

1.6.1 Scale and Rotation Invariant Eye Detection

This first marginal space Ω_1 in our framework is the right eye location and a classifier is trained to detect the right eye, independent of the face orientation, for faces in a certain range of sizes (between 15 and 50 pixels wide). To make sure that the faces are in this range, the detector is run for images reduced by a power of 2. Thus, the detector works in the marginal space of right eye locations (x^e, y^e, s^e), where the eye is detected at position (x^e, y^e) in an image that was reduced 2^{s^e} times.

Face features have also been used for face detection before (Leung, Burl & Perona 1995), but the features were obtained as filter responses, and not by a trained classifier. Moreover, the features were combined purely based on geometry, without any use of the appearance after the feature detection step. This is not the case in this work.

The eye detector is trained as a cascade of LogitBoost (Friedman, Hastie & Tibshirani 2000) classifiers, based on Haar features (Viola & Jones 2004) and the integral image, with more details given in Section 1.8.2. Other eye detection techniques (Feng & Yuen 2001, Wang, Green, Ji & Wayman 2005, Zhou & Geng 2004, Zhu & Ji 2005) could be more appropriate than the Haar-based approach that we used.

The output of this marginal classifier is a number of detected right eyes, that are used to constrain the search in the next space.

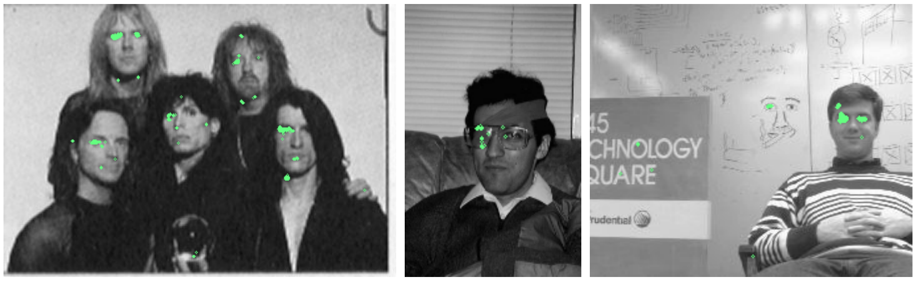

Figure 1.3: Right eye locations detected by the marginal classifier.The classifier takes advantage of the eye context (nose, glasses, etc) and detects the eye even when it is occluded, as shown in the middle image.

Examples outputs of this marginal classifier are shown in Figure 1.3. As one could see, the classifier takes advantage of the area surrounding the eye (nose, glasses, etc) and makes a correct detection even when the eye is completely occluded, as shown in the middle image in Figure 1.3.

This level eliminates more than 99% of the total number of windows that would have to be evaluated in the four dimensional space, at the cost of eliminating some of the true eye locations. If better features (e.g. rotation/illumination invariant) were used to obtain a better eye detection, the whole system's performance in both speed and accuracy could be further improved.

Algorithm 1 Face Detection By MSL

```
Input: Input Image I.
Output: Set $D$ of detected faces.
Construct a Gaussian pyramid $I^r, r = 1, n$ by reducing the image I by powers of 2 as long as the size is at least $l_{min}$.
for r=1 to n do
    Detect eye candidates $(x_i^e, y_i^e, s_i^e), i = 1, ..., d_r$ in $I^r$.
    Generate face candidates using Eq. (1.1)
        $c_{ijk} = (x_{ijk}, y_{ijk}, s_{ij}, \theta_k), i = 1, ..., d_k, j = 0, ..., n_s, k = 0, ..., n_\theta$
    where $s_{ij} = s_i^e(1 + j\delta_s), \theta_k = -\theta_{max} + k\delta_\theta$.
    Compute $p(c_{ijk}|I^r)$ and discard $c_{ijk}$ if $p(c_{ijk}|I^r) < \tau$
end for
Perform Non-Maximal Suppression (Algorithm 2) on the remaining candidates $c_{ijk}$.
```

1.6.2 Four Parameter Face Detection

The obtained particles (candidates) (x^e, y^e, s^e) from the eye marginal space Ω_1 are extended to 4-parameter face candidates (x, y, s, θ) , where (x, y) is the face center position, s is the isotropic scale and θ is the face orientation. This is done by adding to each eye candidate a set of 135 possible orientation-scale combinations, with 15 discrete orientations between -35 and 35 degrees and 9 discrete relative scales $s = s^e(1 + j\delta_s), j = 0, ..., 8$ where $\delta_s = 1/7$.

Anthropometric face measures (DeCarlo, Metaxas & Stone 1998, Horprasert, Yacoob & Davis 1996, Popovici, Thiran, Rodriguez & Marcel 2004) are used to predict the face center (x, y) given the eye location (x^e, y^e), scale s and angle θ as

$$(x, y) = (x^e, y^e) + 0.5 \cdot 7s\mathbf{d} + 0.4 \cdot 7s\mathbf{d}^\perp, \tag{1.1}$$

where $\mathbf{d} = (\cos\theta, \sin\theta)$ and $\mathbf{d}^\perp = (-\sin\theta, \cos\theta)$. An example of face candidates obtained this way at two different scales is shown in Figure 1.4.

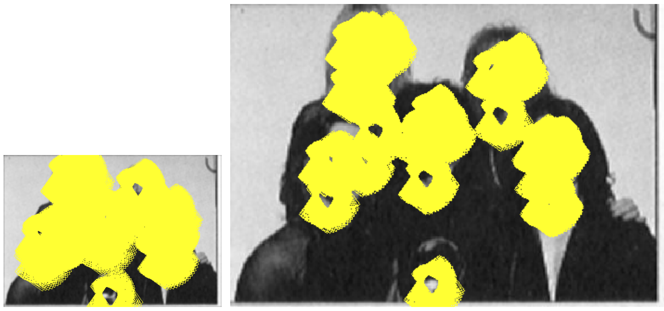

Figure 1.4: Four parameter face candidates $(x_i, y_i, s_i, \theta_i)$ at two scales generated from the detected eyes.

The face candidates close to true faces were used as positive examples while the ones that were far from faces were used as negatives. The face detector was trained using these positives and negatives, Haar features and a cascade of Logitboost classifiers, with more details given in Section 1.8.2.

The whole face detection algorithm using MSL is summarized in Algorithm 1. There are usually many overlapping detections on the same face with slightly different centers, scales and orientations, as illustrated in Figure 1.5.

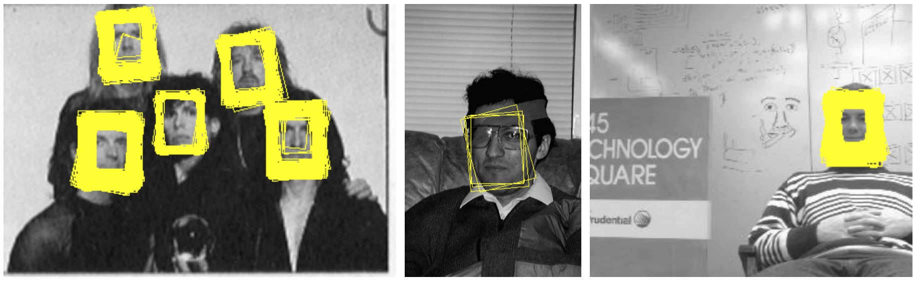

Figure 1.5: Detections rescaled to the original image size. There are usually multiple detections on the same face, which will be handled by the non-maximal suppression Algorithm 2.

To deal with this overlapping detection issue, a non-maximal suppression step is used, as described in Algorithm 2. The algorithm starts with a set of detected candidates $c_i = (x_i, y_i, s_i, \theta_i)$ with probabilities p_i above a predefined threshold τ. It keeps the highest probability candidate and removes all candidates that are close to it (i.e. have centers inside the box determined by its parameters (x, y, s, θ)). Then it keeps the best of the remaining candidates and again removes all candidates that are close to it, and so on, until no candidates are left.

Algorithm 2 Non-maximal Suppression

```
Input: Candidates $c_i = (x_i, y_i, s_i, \theta_i)$ with scores $p_i > \tau$ and bounding boxes $b_i$.
Output: Set $D$ of detected faces.
Find the candidate $c_i$ with highest score $p_i$.
if $c_i$ exists then initialize $D = \{i\}$ else $D = \emptyset$, stop.
while true do
    Remove candidates $c_j$ with centers inside any box $b_i, i \in D$.
    Find remaining candidate $c_j$ of highest score $p_j$.
    if $c_j$ exists then add $j$ to detected set: $D \leftarrow D \cup \{j\}$ else stop.
end while
```

The threshold τ can be used to control the detection and false positive rate. By varying the threshold τ, a ROC (Receiver Operating Characteristic) curve can be obtained, such as the red curve from Figure 1.10.

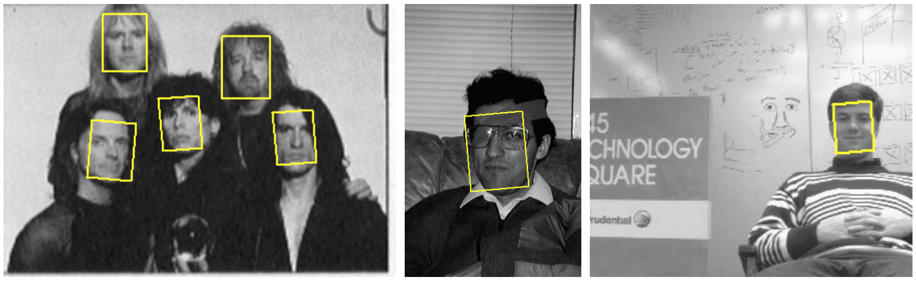

Figure 1.6: Detected faces after non-maximal suppression.

Examples of detections obtained using the face detection algorithm are shown in Figure 1.6.

1.7 Multiple Computational Paths in Marginal Space Learning

When the dimensionality of the parameter space Ω is large or when the marginal spaces correspond to parts that are occluded, it is possible that the propagated particles will not reach the final space Ω for detecting the object of interest. The Recursive Compositional Models (RCM) (Zhu, Chen, Ye & Yuille 2008, Zhu et al. 2007, Zhu et al. 2009, Zhu, Lin, Huang, Chen & Yuille 2008), which are similar in spirit to MSL, have a built-in degree of robustness to missing data that translates into a certain degree of robustness of the optimization. When failures occur in the RCMs, the obtained model parameters $\mathbf{x} \in \Omega$ have some missing parameters that are filled in with their most probable values, without taking the image into consideration. This could result in an inaccuracy of the final result, and has been addressed in the RCMs by a post-processing step of data driven segmentation using Grab-Cut (Rother, Kolmogorov & Blake 2004), using the RCM result for initialization.

To avoid the disadvantages of using a preselected MSL path, multiple MSL paths cand be used, based on different sequences of marginal spaces. However, a direct application of this idea does not show much improvement in performance because while the number of detections increases due to the multiple paths, the number of false positives also increases.

However, by aggregating the results from the different MSL paths instead of just merging them, significant performance improvements can be obtained. By using a simple aggregation scheme described below, significant improvements have been observed for the face detection with MSL application described in Section 1.6. This is because two different MSL paths could serve as independent confirmations of a detection result. This idea is similar to the approach taken by the newspapers to publish information only if it is confirmed through two different sources. Furthermore, this idea is also related to co-training (Blum & Mitchell 1998, Nigam & Ghani 2000), a method for semi-supervised learning

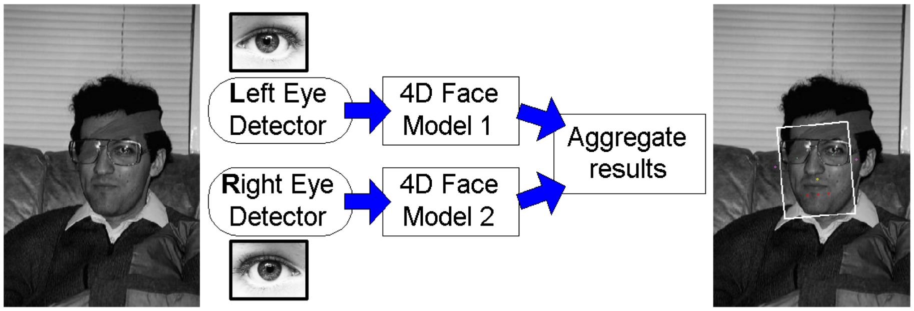

Figure 1.7: Face detection using Two Marginal Space Learning Paths, using the left and respectively right eye locations as marginal subspaces of the 4D face model. Even if one eye is occluded, the face can still be detected successfully through the other MSL path.

that uses two independent sets of features for confirmation of detection results on unseen data. In view of the co-training paradigm, semi-supervised learning based on Multi-path MSL might be feasible.

We performed an experiment using two MSL paths for face detection, using left and respectively right eye detectors as intermediate marginal spaces. The diagram of these two MSL paths is illustrated in Figure 1.7. In Figure 1.8 are shown the detected left and right eyes in cyan and yellow respectively.

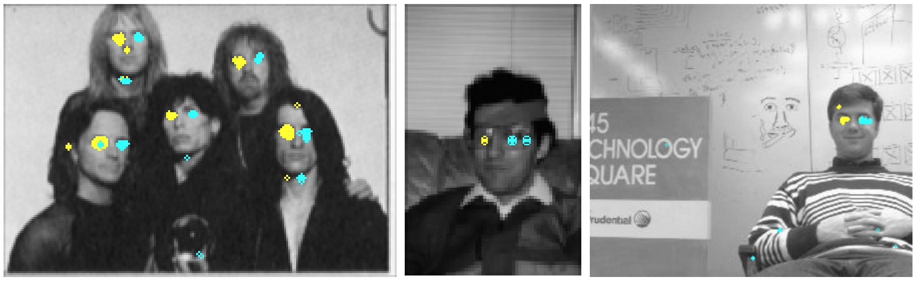

Figure 1.8: Detected right (yellow) and left (cyan) eye locations.

Two different face detectors were trained because the 4D faces are aligned differently if coming through the left or right eye path. Examples of detected faces through the left and right eye MSL paths are shown in cyan respectively yellow in Figure 1.9. As one could see, most of the time the face is detected by both MSL paths, but there are some isolated cases when only one of the detectors finds the face.

We propose the following way to aggregate the results obtained through the multiple MSL paths:

- An object detected with strong confidence from any of the paths is considered detected.

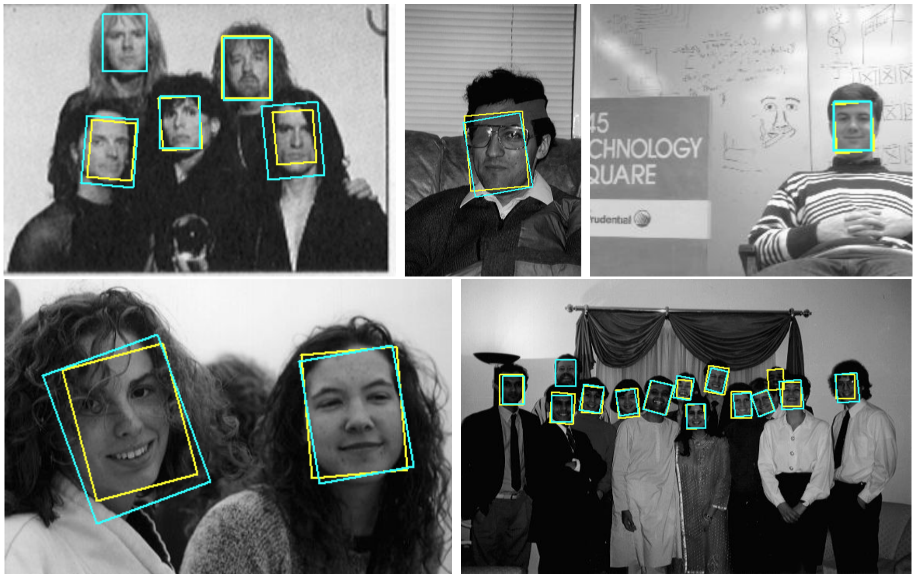

Figure 1.9: Detected face location locations through the left (cyan) and right (yello) MSL paths respectively.

- An object detected from one MSL path is considered detected if it is confirmed from another MSL path, i.e. when a sufficiently close object detected from another MSL path exists.

Hence a face is detected if either it has strong confidence through one MSL path or it is confirmed from both paths.

Using this aggregation scheme we obtained the ROC curve shown in Figure 1.10 as a black solid line. It shows that using Multiple MSL paths leads to significant improvements in accuracy, obtaining results comparable to some recent results in face detection (Brubaker, Mullin & Rehg 2006, Garcia & Delakis 2004, Xiao, Zhu, Sun & Tang 2007).

This Multi-Path MSL idea is motivated by our work in guide-wire localization (Barbu et al. 2007), where an MSL approach was used to detect and segment a thin and flexible wire in fluoroscopy (real-time X-ray). In the guide-wire work, different parts of the wire were modeled using the same discriminative classifier, even though they correspond to different possible subspaces Ω_1. It just so happened that for the guidewire, all these subspaces can use the same model, but this would not extend to other objects, e.g. a snake that have a different appearance for the head than for the body. In practice we observed that different sequences of subspaces were used in different images, depending on what parts of the wire were more visible. Moreover, the chain of subspaces could be longer or shorter depending on the length of the wire in the image. For these reasons, the guide-wire localization work (Barbu et al. 2007) served as a first proof of feasibility of the Multiple Path MSL idea.

1.8 Experimental validation

Two MSL face detectors as described in Section 1.6 were trained, one using the right eye as the marginal space and another one using the left eye. The Multipath MSL was constructed from these two detectors as described in Section 1.7 above.

1.8.1 The training dataset

The eye and face classifiers were trained on a database of 160 images obtained from the Internet, containing 2600 faces that were manually annotated. To obtain a larger set of negative examples (Sung & Poggio 1998), 215 images that don't contain any faces were added from the Berkeley dataset (Martin, Fowlkes, Tal & Malik 2001).

The faces were manually annotated by placing the two eye locations. Given the two eyes, the square window surrounding the face is aligned with the line connecting the eyes, has width twice the distance d between the eyes and height $2.6d$. The face window center is at equal distance from the eyes and at distance $0.8d$ from the line connecting the eyes. These dimensions are similar to the anthropometric measures from (Popovici et al. 2004).

1.8.2 Implementation details

The eye and the face detectors were trained as cascades of three LogitBoost classifiers (Friedman et al. 2000), with parameters given in Table 1.1.

Table 1.1: Training details for the two classifiers including the number of weak classifiers, detection rate and false positive rate.

Classifier	# Features	Clf 1	Clf 2	Clf 3	TPR Train	FPR Train
Eye	66,024	20	60	180	99.2%	0.1%
Face	124,917	50	125	312	95.6%	0.5%

Eye Detectors. The left and right eye detectors were trained as cascade of Logitboost classifiers with 20, 60 and 180 locally constant weak regressors, each weak regressor being based on only one feature. The features are based on Haar features (Papageorgiou, Oren & Poggio 1998, Viola & Jones 2004), restricted in a window of size 19 × 19 pixels centered at the location (x, y). To obtain a more robust training, the number of positive examples was increased with rotated versions by ±5 degrees. This way, 6500 positive examples were used for training.

On the training data, this level detects about 99.% of the eyes with a false alarm rate of about 0.1%.

We also performed an evaluation of this classifier on unseen data, namely the MIT + CMU frontal face test set (Rowley, Baluja & Kanade 1996), containing 130 images and 507 faces. Here, the right eye detector detects 92.7% of the right eyes with an error of at most 1 pixel, and has a false alarm of 0.15%. This shows that there were not enough training examples to capture all the variability in the test data.

Face Detectors. From the eye detector, a number of eye position candidates (x_i^e, y_i^e, s_i^e) are obtained. A total of 135 face candidates with four parameters (x, y, s, θ) are obtained from each eye candidate, as described in Section 1.6.2.

The face detector is also a three-level cascade of Logitboost classifiers with 50, 125 and 312 locally constant weak regressors respectively. Each Logitboost classifier is trained using 124,917 Haar features restricted in an image of size 21×23, working on integral images of rotated and rescaled versions of the original image. For each face candidate (x, y, s, θ), the feature parameters were scaled by s and extracted from the integral image of the image rotated by $-\theta$.

As one could see, the total number of features used in this multi-path MSL approach is 1494, smaller than the other face detection systems (4916 in Viola & Jones (2004) and 2546 in Li et al. (2002)). One of the reasons is that an eye has less variability in appearance than an entire face, at the resolution where the whole face has about 25x25 pixels. Thus an eye detector can be trained using a more compact classifier than a face detector, with good generalization power.

1.8.3 Evaluation

The three MSL versions (two one path and one multi-path) were evaluated on the MIT+CMU dataset (Rowley et al. 1996), containing 130 images and 507 faces that were not used for training.

As already mentioned in Hjelmas & Low (2001), very few papers disclose the criterion on which they report a face as detected or not. Quite similar to Osadchy, Le Cun & Miller (2007), we declare a face correctly detected if the following two criteria are satisfied:

- The distance from the detected window center to the true face window center is less than 0.3 times the true face height
- The ratio between the heights of detected window and the true face is in the interval $[0.5, 1.5]$.

Based on these evaluation criteria, the ROC curves obtained on the MIT+CMU dataset are shown in Figure 1.10. The ROC curves obtained by Viola-Jones (Viola & Jones 2004), FloatBoost(Li et al. 2002) and Convolutional Face Finder (Garcia & Delakis 2004) are also shown for comparison.

We also present in Table 1.2 a comparison with other face detection methods, including the neural-network based face detection (Rowley et al. 1996), Viola & Jones (2004), Schneiderman (2004), and the Convolutional Face Finder (Garcia & Delakis 2004). For Schneiderman (2004), we show in parantheses the actual number of false positives, since they didn't report the detection rates for the

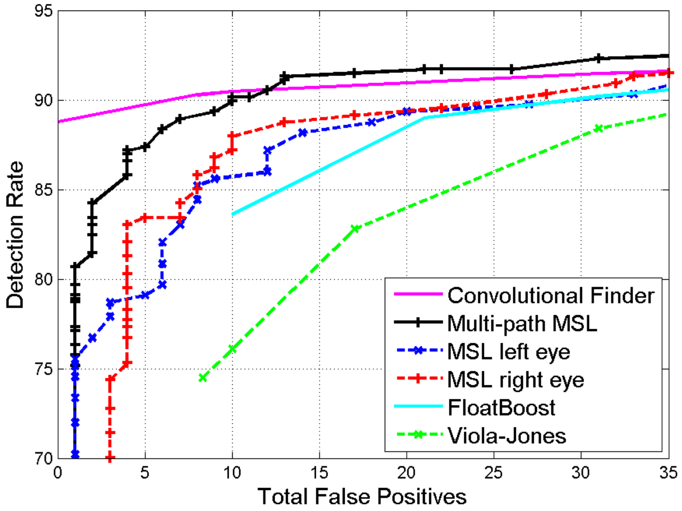

Figure 1.10: Face detection on unseen data (the MIT-CMU dataset) for Two-path MSL, MSL left eye, MSL right eye, Convolutional Face Finder , FloatBoost and Viola-Jones.

same false positives as the other papers, but at the same time, they reported detection rates out of 506 faces, not 507 as in the other papers.

From the experiments, one could see that the multi-path MSL approach outperforms the single path MSL versions. It also outperforms some face detection algorithms such as the FloatBoost (Li et al. 2002) and Viola & Jones (2004) while being trained on fewer faces. Furthermore, the Multi-path MSL obtains results comparable (at 10 false positives) to the CART based classifier from (Brubaker et al. 2006), to the Convolutional Face Finder (Garcia & Delakis 2004) trained on 3,700 faces, and to the Dynamic Cascade (Xiao et al. 2007) trained with 40,000 faces.

In the future, we plan to enlarge the number of training faces by adding images from the FERET dataset (http://face.nist.gov/colorferet/ n.d.) and using some of the smoothing and contrast reduction techniques from Garcia & Delakis (2004). Our algorithm uses only 2,600 faces for training, as opposed to the state of the art algorithms for face detection such as Bourdev & Brandt (2005) and Xiao et al. (2007) that use between 10,000 to 40,000 faces.

The detection time depends on the image complexity. For simple images, the marginal classifiers will remove most of the false positives, resulting in a fast detection. As a result, for a 384x288 image the detection time varies between 0.06 seconds and 0.30 seconds on a 2.4GHz dual core PC.

Examples of face detections on some test images are given in Figures 1.11 and 1.12.

These results can be improved in accuracy by using better features such as illumination-invariant features (Schneiderman 2004), or CART features (Brubaker et al. 2006), or by joint training of all cascade levels (Dundar & Bi 2007), learn-

Table 1.2: Face detection rates for different numbers of false positives obtained by different methods on the MIT+CMU frontal face dataset containing 130 images and 507 faces.

		False Detections		
Detector	Train faces	0	10	31
MSL Righte eye	2,600	74.8%(1)	88.5%	89.2%
MSL Left eye	2,600	64.8%	87.1%	89.9%
Multi-Path MSL	2,600	80.7%(1)	90.1%	92.3%
Rowley et al. (1996)	1,046	-	83.2%	86.0%
FloatBoost (Li et al. 2002)	6,000	-	83.6%	90.2%
Viola & Jones (2004)	4,916	-	76.1%	88.4%
Viola-Jones (voting)	4,916	-	81.1%	89.7%
Schneiderman (2004)	-	-	89.7 (6)	94.4%(29)
Convolutional face finder	3,702	88.8%	90.3%	91.5%
CART (Brubaker et al. 2006)	-	-	90.5%	93.1%
Dynamic Cascade	40,857	86.9%(1)	89.8%	92.2%

ing the features (Wang & Ji 2005) and of course by training with more faces.

1.9 Applications

The Marginal Space Learning and the Multi-Path Marginal Space Learning can be applied to most object detection problems. Furthermore, it can be applied to object segmentation, since an object segmentation is a more accurate description of an object, with more parameters than in object detection.

1.10 Open Issues and Problems

One open issue is to give a mathematical characterization of the quality of different MSL paths to help deciding which paths are the best.

Another issue that remains open for Multi-path Marginal Space Learning is how to coordinate the different MSL paths for an efficient use of computation. According to the way the results from different MSL paths are aggregated, if a detection through one path is strong enough, any detections through other paths that are close to this strong detection are not necessary. Thus redundant computation could be saved by avoiding to pursue any detections close to strong detections from one path.

1.11 Data Sets

The following publicly available datasets have been mentioned in this chapter:

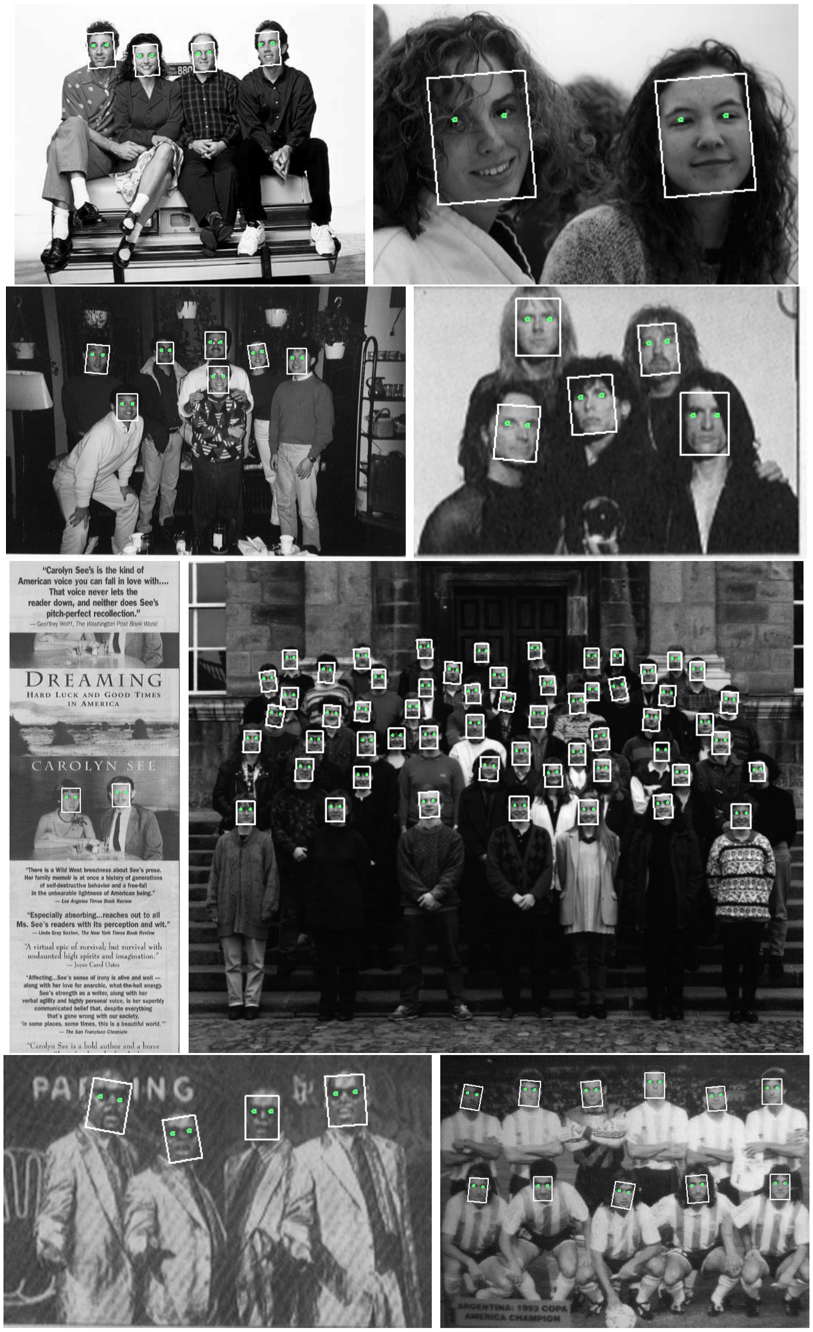
880
"Carolyn See's is the kind of American voice you can fall in love with.... That voice never lets the reader down, and neither does See's pitch-perfect recollection."
— Geoffrey Wolff, The Washington Post Book World
DREAMING
HARD LUCK AND GOOD TIMES IN AMERICA
CAROLYN SEE
"There is a Wild West breeziness about See's prose. Her family memoir is at once a history of generations of self-destructive behavior and a free-fall in the unbearable lightness of American being."
— Los Angeles Times Book Review
"Especially absorbing...reaches out to all Ms. See's readers with its perception and wit."
— Linda Gray Sexton, The New York Times Book Review
"A virtual epic of survival; but survival with undaunted high spirits and imagination."
— Joyce Carol Oates
"Affecting...See's sense of irony is alive and well — along with her love for anarchic, what-the-hell energy. See's strength as a writer, along with her verbal agility and highly personal voice, is her superbly communicated belief that, despite everything that's gone wrong with our society, 'in some places, some times, this is a beautiful world.'"
— The San Francisco Chronicle
"Carolyn See is a bold author and a brave
PAR ING
ARGENTINA: 1993 COPA AMERICA CHAMPION

Figure 1.11: Face detection results obtained by our method.

1. The CMU+MIT Face dataset (Rowley et al. 1996). A dataset of 130

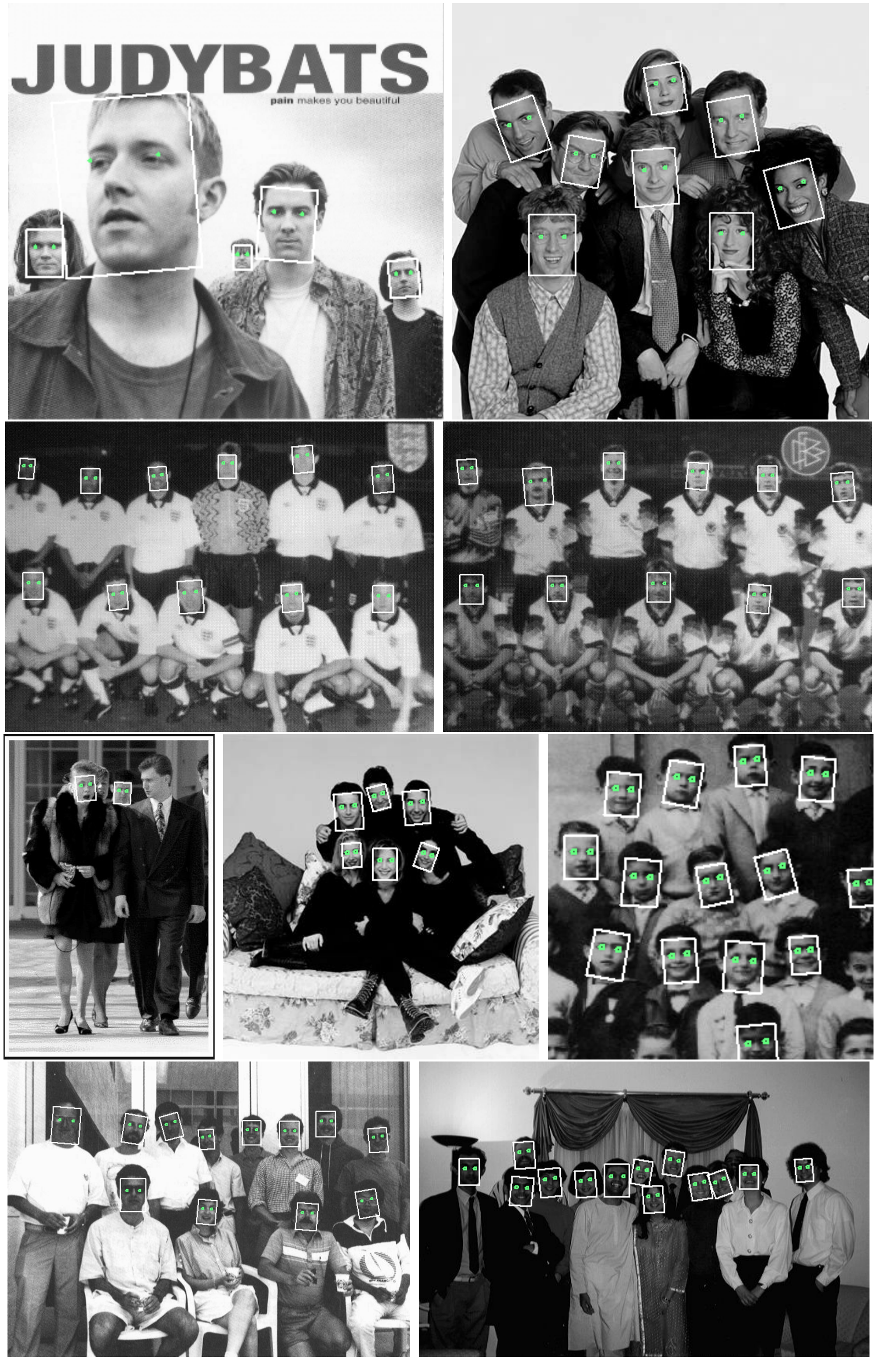

Figure 1.12: More face detection results obtained by our method.

grayscale images containing 507 manually annotated faces.

2. The FERET Dataset (http://face.nist.gov/colorferet/ n.d.). A dataset containing thousands of images of faces from different angles (frontal, half-profile, profile, etc), with different facial expressions and illuminations.

1.12 Conclusions and Future Trends

In this chapter we presented a fast and robust method for object detection that combines different computational paths of Marginal Space Learning for improved robustness against occlusion and detector failures. In this approach, invariant detectors are used to select a set of good object candidates in different marginal spaces while detectors based on non-invariant features are used to increase the number of object parameters and obtain the final solution.

We showed that this multi-path MSL method can achieve good detection rates and low false positive rates, comparable with some state of the art approaches, while at the same time using less training data than these approaches. With better features (e.g. rotation/illumination invariant), more training data and better models for the final classifier, further performance improvements are possible.

The Multi-Path MSL method can be used in the future for detecting hard-to-find objects such as boats, that pose challenges to other state of the art techniques.

1.13 Cross-References

Training Logitboost classifiers should be described in the Machine Learning section 3 of the book.

Bibliography

Agarwal, S. & Roth, D. (n.d.), 'Learning a sparse representation for object detection', *ECCV* .

Amit, Y. & Geman, D. (1999), 'A computational model for visual selection', *Neural Computation* **11**(7), 1691–1715.

Andriluka, M., Roth, S. & Schiele, B. (2009), Pictorial Structures Revisited: People Detection and Articulated Pose Estimation, *in* 'CVPR'.

Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P. & Comaniciu, D. (2007), Hierarchical learning of curves application to guidewire localization in fluoroscopy, *in* 'IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07', pp. 1–8.

Barbu, A., Suehling, M., Xu, X., Liu, D., Zhou, S. & Comaniciu, D. (2010), 'Automatic detection and segmentation of axillary lymph nodes', *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010* pp. 28–36.

Barbu, A., Suehling, M., Xu, X., Liu, D., Zhou, S. & Comaniciu, D. (2011), 'Automatic detection and segmentation of lymph nodes from ct data', *IEEE transactions on medical imaging* .

Blanz, V. & Vetter, T. (1999), A Morphable model for the synthesis of 3D faces, *in* 'SIGGRAPH', pp. 187–194.

Blum, A. & Mitchell, T. (1998), Combining labeled and unlabeled data with co-training, *in* 'Proceedings of the eleventh annual conference on Computational learning theory', ACM New York, NY, USA, pp. 92–100.

Bourdev, L. & Brandt, J. (2005), Robust object detection via soft cascade, *in* 'CVPR 2005'.

Brubaker, S., Mullin, M. & Rehg, J. (2006), 'Towards Optimal Training of Cascaded Detectors', *ECCV* .

Burt, P. & Adelson, E. (1983), 'The Laplacian pyramid as a compact image code', *IEEE Trans. Communications* **31**(4), 532–540.

Chen, H., Belhumeur, P. & Jacobs, D. (n.d.), 'In search of illumination invariants', *CVPR* .

Chou, P. & Brown, C. (1990), 'The theory and practice of Bayesian image labeling', *IJCV* **4**(3), 185–210.

DeCarlo, D., Metaxas, D. & Stone, M. (1998), An anthropometric face model using variational techniques, *in* 'Proceedings of the 25th annual conference on Computer graphics and interactive techniques', ACM New York, NY, USA, pp. 67–74.

Dollar, P., Tu, Z. & Belongie, S. (2006), Supervised learning of edges and object boundaries, *in* 'CVPR', Vol. 2.

Doucet, A. (2007), Particle markov chain monte carlo, *in* 'Third Cape Cod MCMC Workshop'.

Dundar, M. & Bi, J. (2007), 'Joint optimization of cascaded classifiers for computer aided detection', *CVPR* .

Fei-Fei, L., Fergus, R. & Perona, P. (2006), 'One-shot learning of object categories', *IEEE Trans. PAMI* **28**(4), 594–611.

Felzenszwalb, P. & Huttenlocher, D. (2005), 'Pictorial structures for object recognition', *IJCV* **61**(1), 55–79.

Feng, G. & Yuen, P. (2001), 'Multi-cues eye detection on gray intensity image', *Pattern Recognition* **34**(5), 1033–1046.

Feng, S., Zhou, S., Good, S. & Comaniciu, D. (2009), Automatic fetal face detection from ultrasound volumes via learning 3d and 2d information, *in* 'Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on', IEEE, pp. 2488–2495.

Fergus, R., Perona, P. & Zisserman, A. (2003), 'Object class recognition by unsupervised scale-invariant learning', *CVPR* **2**.

Feulner, J., Zhou, S., Huber, M., Hornegger, J., Comaniciu, D. & Cavallaro, A. (2010), Lymph node detection in 3-d chest ct using a spatial prior probability, *in* 'Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on', IEEE, pp. 2926–2932.

Fleuret, F. & Geman, D. (2001), 'Coarse-to-Fine Face Detection', *IJCV* **41**(1), 85–107.

Forsyth, D., Mundy, J., Zisserman, A., Coelho, C., Heller, A. & Rothwell, C. (1991), 'Invariant Descriptors for 3D Object Recognition and Pose', *IEEE Trans. PAMI* **13**(10), 971–991.

Friedman, J., Hastie, T. & Tibshirani, R. (2000), 'Additive logistic regression: a statistical view of boosting', *Ann. Statist* **28**(2), 337–407.

Garcia, C. & Delakis, M. (2004), 'Convolutional face finder: a neural architecture for fast and robust face detection', *IEEE Trans. PAMI* **26**(11), 1408–1423.

Heisele, B., Serre, T., Prentice, S. & Poggio, T. (2003), 'Hierarchical classification and feature reduction for fast face detection with support vector machines', *Pattern Recognition* **36**(9), 2007–2017.

Hjelmas, E. & Low, B. (2001), 'Face detection: A survey', *CVIU* **83**(3), 236–274.

Horprasert, T., Yacoob, Y. & Davis, L. (1996), Computing 3-D head orientation from a monocular image sequence, *in* 'Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on', pp. 242–247.

Hsu, R., Abdel-Mottaleb, M. & Jain, A. (2002), 'Face detection in color images', *IEEE Trans. PAMI* pp. 696–706.

http://face.nist.gov/colorferet/ (n.d.).

Huang, C., Ai, H., Li, Y. & Lao, S. (2005), Vector boosting for rotation invariant multi-view face detection, *in* 'ICCV', Vol. 1, pp. 446–453.

Jones, M. & Rehg, J. (2002), 'Statistical color models with application to skin detection', *IJCV* **46**(1), 81–96.

Kuncheva, L. (2004), *Combining pattern classifiers: methods and algorithms*, Wiley-Interscience.

Leung, T., Burl, M. & Perona, P. (1995), 'Finding faces in cluttered scenes using random labeled graph matching', *ICCV* pp. 637–644.

Li, S., Zhu, L., Zhang, Z., Blake, A., Zhang, H. & Shum, H. (2002), 'Statistical Learning of Multi-View Face Detection', *ECCV* .

Ling, H., Zhou, S., Zheng, Y., Georgescu, B., Suehling, M. & Comaniciu, D. (2008), Hierarchical, learning-based automatic liver segmentation, *in* 'Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on', IEEE, pp. 1–8.

Liu, C. (2003), 'A Bayesian discriminating features method for face detection', *IEEE Trans. PAMI* **25**(6), 725–740.

Martin, D., Fowlkes, C., Tal, D. & Malik, J. (2001), 'A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms', *ICCV* **2**, 416–425.

Nigam, K. & Ghani, R. (2000), Analyzing the effectiveness and applicability of co-training, *in* 'Proceedings of the ninth international conference on Information and knowledge management', ACM New York, NY, USA, pp. 86–93.

Osadchy, M., Le Cun, Y. & Miller, M. (2007), 'Synergistic Face Detection and Pose Estimation with Energy-Based Models', *JMLR* **8**, 1197–1215.

Papageorgiou, C., Oren, M. & Poggio, T. (1998), A general framework for object detection, *in* 'ICCV', pp. 555–562.

Popovici, V., Thiran, J., Rodriguez, Y. & Marcel, S. (2004), 'On performance evaluation of face detection and localization algorithms', *ICPR* **1**.

Roth, D., Yang, M. & Ahuja, N. (2001), A snow-based face detector, *in* 'NIPS'.

Rother, C., Kolmogorov, V. & Blake, A. (2004), '" GrabCut": interactive foreground extraction using iterated graph cuts', *ACM Transactions on Graphics (TOG)* **23**(3), 309–314.

Rowley, H., Baluja, S. & Kanade, T. (1996), Neural network-based face detection, *in* 'CVPR', pp. 203–208.

Schneiderman, H. (2004), 'Feature-centric evaluation for efficient cascaded object detection', *CVPR* **2**.

Schneiderman, H. & Kanade, T. (2000), 'Statistical method for 3D object detection applied to faces and cars', *CVPR* **1**, 746–751.

Seifert, S., Barbu, A., Zhou, S., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A. & Comaniciu, D. (2009), 'Hierarchical parsing and semantic navigation of full body ct data', *Medical Imaging* **7259**, 725902.

Slater, D. & Healey, G. (1996), 'The Illumination-Invariant Recognition of 3D Objects Using Local Color Invariants', *Illumination* **18**(2), 206–210.

Sung, K. & Poggio, T. (1998), 'Example-based learning for view-based human face detection', *IEEE Trans. PAMI* **20**(1), 39–51.

Torralba, A., Murphy, K. & Freeman, W. (2004), Sharing features: efficient boosting procedures for multiclass object detection, *in* 'CVPR', Vol. 2, IEEE Computer Society; 1999.

Tu, Z. (2005), Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering, *in* 'ICCV 2005', Vol. 2.

Tu, Z. (2007), 'Learning Generative Models via Discriminative Approaches', *CVPR* pp. 1–8.

Viola, P. & Jones, M. (2004), 'Robust Real-Time Face Detection', *IJCV* **57**(2), 137–154.

Wang, H. & Chang, S. (1997), 'A highly efficient system for automatic face region detection inMPEG video', *IEEE Transactions on Circuits and Systems for Video Technology* **7**(4), 615–628.

Wang, P., Green, M., Ji, Q. & Wayman, J. (2005), Automatic eye detection and its validation, *in* 'CVPR', pp. 164–164.

Wang, P. & Ji, Q. (2005), 'Learning discriminant features for multi-view face and eye detection', *CVPR* **1**.

Wood, J. (1996), 'Invariant pattern recognition: A review', *Pattern Recognition* **29**(1), 1–17.

Wu, J., Rehg, J. & Mullin, M. (2003), 'Learning a Rare Event Detection Cascade by Direct Feature Selection', *NIPS* .

Wu, T. & Zhu, S. (2011), 'A numerical study of the bottom-up and top-down inference processes in and-or graphs', *International journal of computer vision* **93**(2), 226–252.

Xiao, R., Zhu, H., Sun, H. & Tang, X. (2007), Dynamic cascades for face detection, *in* 'ICCV', pp. 1–8.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M. & Comaniciu, D. (2007), 'Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features', *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* pp. 1–8.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M. & Comaniciu, D. (2008), 'Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features', *IEEE Transactions on Medical Imaging* **27**(11), 1668–1681.

Zhou, Z. & Geng, X. (2004), 'Projection functions for eye detection', *Pattern Recognition* **37**(5), 1049–1056.

Zhu, L., Chen, Y., Ye, X. & Yuille, A. (2008), Structure-perceptron learning of a hierarchical log-linear model, *in* 'Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition'.

Zhu, L., Chen, Y. & Yuille, A. (2007), 'Unsupervised learning of a probabilistic grammar for object detection and parsing', *NIPS* **19**, 1617.

Zhu, L., Chen, Y. & Yuille, A. (2009), 'Unsupervised Learning of Probabilistic Grammar-Markov Models for Object Categories', *IEEE Trans. PAMI* **31**(1), 114–128.

Zhu, L., Lin, C., Huang, H., Chen, Y. & Yuille, A. (2008), Unsupervised structure learning: hierarchical recursive composition, suspicious coincidence and competitive exclusion, *in* 'Proceedings of the 10th European Conference on Computer Vision: Part II', Springer-Verlag Berlin, Heidelberg, pp. 759–773.

Zhu, Z. & Ji, Q. (2005), 'Robust real-time eye detection and tracking under variable lighting conditions and various face orientations', *CVIU* **98**(1), 124–154.

Zisserman, A., Forsyth, D., Mundy, J., Rothwell, C., Liu, J. & Pillow, N. (1995), '3D object recognition using invariance', *Artificial Intelligence* **78**(1-2), 239–288.

Index