Segmentation of Distinct Homogeneous Color Regions in Images

Daniel Mohr and Gabriel Zachmann

Department of Computer Science, Clausthal University, Germany,
{mohr, zach}@in.tu-clausthal.de

Abstract. In this paper, we present a novel algorithm to detect homogeneous color regions in images. We show its performance by applying it to skin detection. In contrast to previously presented methods, we use only a rough skin direction vector instead of a static skin model as a priori knowledge. Thus, higher robustness is achieved in images captured under unconstrained conditions. We formulate the segmentation as a clustering problem in color space. A homogeneous color region in image space is modeled using a 3D gaussian distribution. Parameters of the gaussians are estimated using the EM algorithm with spatial constraints. We transform the image by a whitening transform and then apply a fuzzy k-means algorithm to the hue value in order to obtain initialization parameters for the EM algorithm. A divisive hierarchical approach is used to determine the number of clusters. The stopping criterion for further subdivision is based on the edge image.
For evaluation, the proposed method is applied to skin segmentation and compared with a well known method.

1 Introduction

Image segmentation is an important task in computer vision for instance when tracking objects, it is used to identify the object to be tracked or parts of it. Other common applications can be found in medical image segmentation, for example to identify tumors or bones. Also, OCR software uses such algorithms to separate text from background and it has applications in image/video compression. One issue during initialization of a tracking system is to find the object's spatial location in the image. Especially for nonrigid objects, color is the main feature to accomplish this task. Due to unknown lighting conditions, e.g. colored light, inhomogeneously colored background, different camera hardware, and other influences, the color of an object can be very different compared to its color under controlled conditions. Thus, a static color model of a target object may fail. However, in a dynamic color model, a method is needed to initially detect at least a large part of the object. Once this has been achieved, the color model can be adapted. On application to skin segmentation, a large part of skin has to be detected, then previously presented algorithms for skin segmentation such as [1] can be used to update the color model during tracking and, thus, refine the segmentation.

The problem to detect a homogeneous colored object with a homogeneous color under controlled conditions is that we don't know the color distribution of the object under uncontrolled conditions. Using a color model obtained from images taken under different conditions can strongly differ and may fail. To address this problem our approach estimates the color model instead of using a precomputed one. Our method can basically divided into two steps. First, the image has to be segmented correctly into subsets, each representing such an object, and second, identify the correct subset representing the target object. In the following, we call those image subsets *image regions*. Our method is able to identify such a region by utilizing the separability of different regions in color space and that a target object has a certain direction relative to the image background in most cases. When applied to skin detection our approach makes no fixed assumption to skin color distribution, in contrast to other methods. To identify the image region representing the target, we only need a rough direction of skin relative to the background in color space.

Obviously, clustering should be performed in color space. We compared the RGB, HSV and Lab histograms of several images. We could not observe any noticeable improvement to distinguish color regions in HSV and Lab space. Therefore we use the RGB color space and avoid a color space conversion. The regions can be appropriately modeled by three dimensional gaussians. We determine clusters by the EM algorithm with additional spatial constraints. Appropriate initialization values have to be calculated before performing the EM algorithm. In order to determine the number of clusters we use a hierarchical approach.

2 Related Work

In the past, several image segmentation algorithms have been introduced. Most approaches can be classified as graph cutting or color space clustering method. Graph cutting methods model the image as a weighted graph and map segmentation to graph cutting with a specific cost function. Color space clustering methods approximate image segments with an appropriate model in color space. Both use criteria such that resulting segments are similar to those a human person would define. In [2] the graph cutting method with a normalized cut criterion is used. Their approach avoids cutting small sets of isolated pixels. [3] used a multilevel hypergraph in order to segment gray level images with special interest on its performance of noisy images.

Several color space clustering approaches have been presented. In [4] a morphological clustering followed by Markovian labeling is used for segmentation. [5] presented a k-mean clustering in HSI space with application to medical images. Special attention has been paid to the cyclic property of the hue component. In [6] a novel initialization scheme for fuzzy c-means clustering has been introduced. Dominant colors, defined as the most vivid and distinguishable colors, are calculated from reference colors. The colors nearest to the dominant colors are used as initial centroids.

[7] compared histogram based and mixture model representation of skin and non-skin color. They constructed the color models for skin and non-skin classes from a dataset of nearly 1 billion hand labeled pixels. They examined that the histogram based representation is superior for very large training data sets. For small training data sets, the mixture model delivers better segmentation results. They reached a detection rate of 80% at a false positive rate of 8.5% for web images. The main disadvantage is the inflexibility of a static skin color model. It may have a low performance on images captured under different conditions than their training data set. [8] improved skin detection by a variational EM algorithm with spatial constraints. For

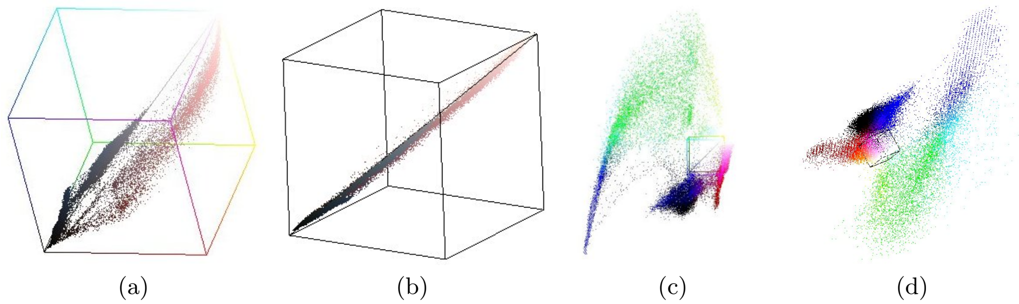

Fig. 1. Before we perform clustering, the image is transformed into a decorrelated, unit variance space. Figures (a) and (b) shows the image in RGB space. Figure (c) and (d) shows the image after transformation into a decorrelated, unit variance space. Each example is shown from two different viewpoints.

initialization, they used the skin color model of [7]. In [9] a gaussian mixture model was used to improve a skin-similar space, which was built from a rough classification with a static skin model. The skin gaussian was identified by a Support Vector Machine using spatial and shape information. [10] proposed a skin segmentation method in YCbCr space, applying Bayesian decision rules. A face detector is used in [11] to generate a skin model and then applied to images to detect skin. [1] predicted changes of skin color during tracking with a second order Markov model. Skin and non-skin color histograms are updated based on feedback from current segmentation and prediction. Skin color changes are modeled as translation, scaling and rotation in color space. Their approach requires an initial detection of skin. It possibly fails if this initial skin segmentation has a too small or high false detection rate.

3 Our Approach

We use images captured under arbitrary lighting conditions and environment as input. The goal of our method is to segment the region of an image which represents the target. In our example, we want to identify skin regions, which are typically closer to red than common backgrounds. Nevertheless, color distribution can heavily deviate from red. In contrast to previously proposed skin detection methods, we do not start with a fixed assumption for skin color distribution, which could lead to high false detection. In our input images, we have many unknown influence factors. We know by the central limit theorem that a point set, that is influenced by many small factors can be well approximated by the normal distribution. Therefore, it is useful to model the color distribution of image regions with mixture of gaussians (GMM). We have chosen a well known method to estimate the parameters of a GMM, the EM algorithm [12].

A *homogeneous color region* is a region in the image space that represents an object, which has a homogeneous color under white uniform illumination. In images captured under unconstrained conditions, histogram form of such regions can be heavily stretched. To compensate potentially negative effects on the clustering al-

Fig. 2. We use additional spatial constraints for the EM algorithm to get smoother regions. The above pictures show an image (left) in which we detect skin without constraints (middle), and with constraints (right)

gorithm we transform[1] the histogram by $\boldsymbol{y}_i = S^{-\frac{1}{2}}\ U^T\ (\boldsymbol{x}_i - \boldsymbol{m})$ where U and S are obtained from the singular value decomposition $[U, S, V^T] = \mathrm{svd}(\mathcal{C})$ of the covariance matrix $\mathcal{C}$ (see Figure 1).

3.1 Spatial Constraints

A problem of color space clustering is that we often get many isolated pixels or very small regions in image space. An example is shown in Figure 2. To address this problem, we use spatial constraints to get smoother cluster borders. Before we can explain our smoothing method, we need to introduce the edge distance image $D(\mathcal{I})$.

The Laplace edge detection operator is applied to the input image $\mathcal{I}$. The resulting image, containing the edges of the input image, is denoted with $C(\mathcal{I})$. To each pixel $\boldsymbol{x_i} \in C(\mathcal{I})$ the Filter

$$\mathcal{F}(\boldsymbol{x_i}) = \max_{\boldsymbol{x_j} \in \mathcal{N}(\boldsymbol{x_i})} \frac{C(\boldsymbol{x_j})}{\|\boldsymbol{x_i} - \boldsymbol{x_j}\| + 1} \tag{1}$$

is applied. $\mathcal{N}(\boldsymbol{x_i})$ is the $k \times k$ image neighborhood of $\boldsymbol{x_i}$. Thus, we obtain the image $\tilde{D}(\mathcal{I})$. After normalizing the pixels of $\tilde{D}(\mathcal{I})$ to $[0, 1]$, the final edge distance image $D(\mathcal{I})$ is obtained.

The idea of our spatial smoothing method is based on $D(\mathcal{I})$. In a neighborhood $\mathcal{N}(\boldsymbol{x_i})$ of a pixel $\boldsymbol{x_i}$ without an edge, all pixels in $\mathcal{N}(\boldsymbol{x_i})$ should have similar probability to belong to a particular cluster. If an edge is found in the neighborhood. If not, we have no usable information about the membership of pixels to clusters. Thus, in each iteration step of the EM algorithm, we calculate the average probability $\bar{p}(\boldsymbol{x_i}|\theta_i)$ of the neighborhood with size $l \times l$ for all pixels. Then, we use the edge distance image to interpolate between the probability of a pixel belonging to a cluster and the average neighborhood probability. The new probability $p_n(\boldsymbol{x_i}|\theta_i)$ is calculated through

$$p_n(\boldsymbol{x_i})|\theta_i) = p(\boldsymbol{x_i}|\theta_i)D(\boldsymbol{x_i}) + (1 - D(\boldsymbol{x_i}))\bar{p}(\boldsymbol{x_i}|\theta_i) \tag{2}$$

[1] This transformation is similar to the whitening transform $y = US^{-\frac{1}{2}}U^T\ (\boldsymbol{x} - \boldsymbol{m})$, but for our algorithm there is no need to perform the leftmost Matrix (U) multiplication.

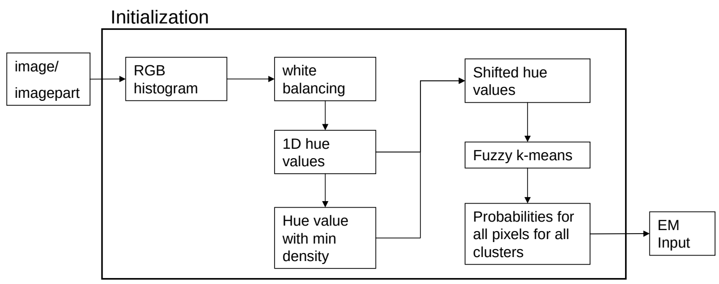

Fig. 3. Initialization: A white balancing to the image part is performed and the hue values for all pixels are calculated. Hue values are shifted, such that values around 0 have minimal density. Fuzzy k-means is applied to this hue values. The resulting probabilities are used as input for the EM algorithm

3.2 Initialization

The initialization step has a significant influence to the resulting clusters because the EM algorithm only guarantees to converge to a local minimum. Since we want to segment image regions with homogeneous color, it makes sense to initialize each cluster with a hue which differs from the hue of the other clusters as much as possible. Performing a simple hue clustering makes no sense. Consider an image part whose average color is some red value. We want to divide it into two parts. If we cluster with respect to hue, we get one big red cluster representing the whole image part. Additionally, one cannot presume that the greatest principal axis of a region with homogeneous color in color space is parallel to the gray axis. To take these issues into account, we perform a PCA based white balancing of each image part we want to segment. We have to consider the cyclic property of the hue. The handling with a metric which takes this cyclic property into account is difficult. To avoid this, we search for an hue value $\alpha_{min} \in [0, 360)$ in this cyclic color space with minimal point density and shift the point set about $-\alpha_{min}$. Finally fuzzy-k-means clustering to these hue is performed. Figure 3 illustrates the initialization steps.

3.3 Hierarchical Clustering

An important question for clustering methods is the number of clusters a dataset should be divided into. Because it is hard to answer this question prior to clustering our image we decided to use an hierarchical method. There are two main approaches for hierarchical clustering, agglomerative and divisive. We use a divisive method for two reasons. First, agglomerative clustering can have quadratic complexity. Second, the divisive approach has the advantage that we do not need to subdivide *all* clusters in our case, thus yielding a significant further speedup. This is because we are interested in a homogeneous color region of a special color, for example skin, we can skip subdivision of regions, whose mean vector direction is too far away from the color of the destination object.

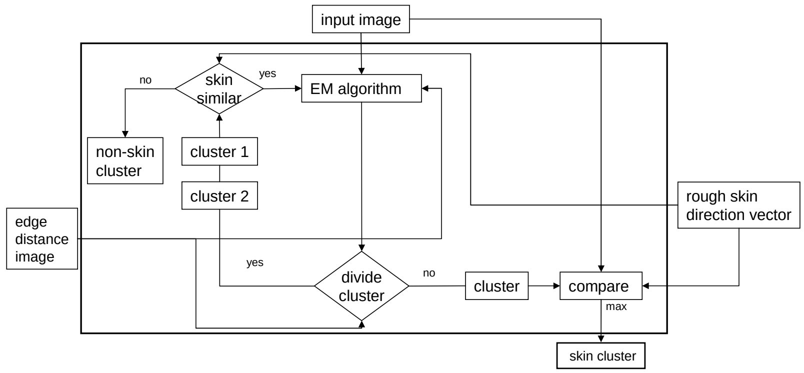

Fig. 4. Hierarchical clustering: In each iteration, EM algorithm is performed with two kernels. The edge image is used to decide if further subdivision is necessary Clusters that need no further subdivision are compared with the skin direction vector to identify the correct skin cluster.

Consider two images. The first image has a uniform illumination, the second one has strong highlights and shadows and colors are poorly saturated. Compared to the first image, the color distributions of image regions, including the target region, in the second image are closer to and stretched along the gray axis. Therefore we need to take the distribution parameters of the image into account when identifying the target region. In order to do that, we perform a whitening transform of the image colors before calculating the mean value of the target region. Let $\boldsymbol{m}$ be the mean value and $[U, S, V^t] = \mathrm{svd}(\mathcal{C})$ the SVD of the covariance matrix of the image, and $\boldsymbol{m}_p$ the mean value of a image part. Then the transformed mean value $\tilde{\boldsymbol{m}}_{\boldsymbol{p}}$ is calculated through

$$\tilde{\boldsymbol{m}}_{\boldsymbol{p}} = \left(U \cdot S^{-\frac{1}{2}} \cdot U^T\right)(\boldsymbol{m}_{\boldsymbol{p}} - \boldsymbol{m}) \tag{3}$$

The transformed mean value of a cluster has to be compared with the transformed mean vector $\tilde{\boldsymbol{m}}_{\boldsymbol{S}}$ characterizing the target region. This vector has to be calculated in a preprocessing step. We do this with a small image data set for skin detection. The images are captured under different illumination conditions. The skin regions are hand labeled.

During clustering we calculate the angle, weighted by the inverse distance, of the transformed mean vector $\tilde{\boldsymbol{m}}_{\boldsymbol{i}}$ of the ith cluster and $\tilde{\boldsymbol{m}}_{\boldsymbol{S}}$

$$\alpha_i = \frac{\tilde{\boldsymbol{m}}_{\boldsymbol{S}} \cdot \tilde{\boldsymbol{m}}_{\boldsymbol{i}}}{\parallel \tilde{\boldsymbol{m}}_{\boldsymbol{S}} \parallel \cdot \parallel \tilde{\boldsymbol{m}}_{\boldsymbol{i}} \parallel \cdot \parallel \tilde{\boldsymbol{m}}_{\boldsymbol{S}} - \tilde{\boldsymbol{m}}_{\boldsymbol{i}} \parallel} \tag{4}$$

If $\alpha_i < \varepsilon$ for some user defined ε, the cluster will be classified as a region that does not contain the region representing the target object. After clustering, the cluster with the largest α_i represents the color distribution of the target object.

The idea for our stopping criterion is based on the edge distance image $D(\boldsymbol{x_i})$. The better a subdivision calculated through the clustering of an image part in color space into two parts represents a useful segmentation in image space, the nearer pixels on

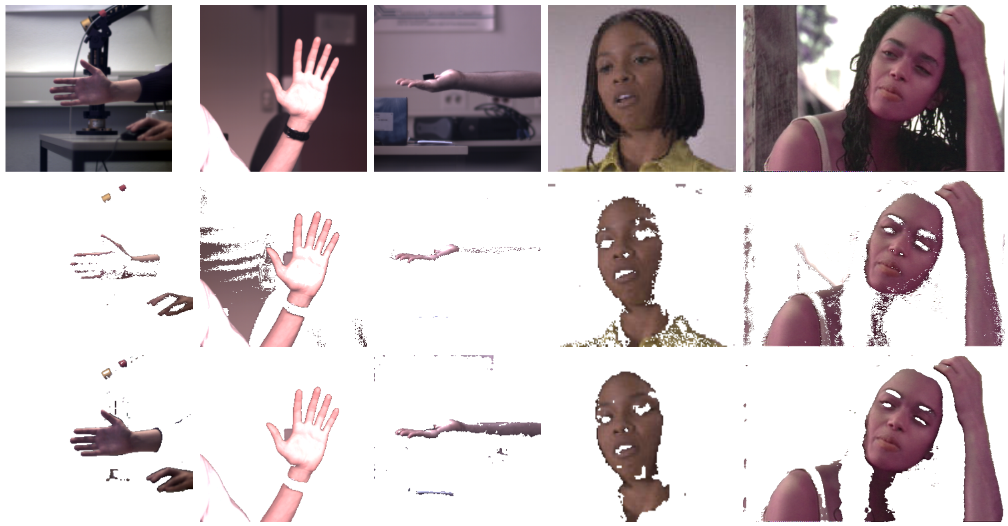

Fig. 5. Results: The original images (first row), result from [7] (second row) and our approach (third row)

the region borders in image space should lie. Therefore, we evaluate clustering quality through the average edge intensity value on this region borders. If

$$\frac{1}{|\mathcal{B}_1| + |\mathcal{B}_2|} \sum_{x_i \in \mathcal{B}_1 \cup \mathcal{B}_2} D(\boldsymbol{x_i}) > \delta \tag{5}$$

the clusters are split, otherwise not. $\mathcal{B}_1$ and $\mathcal{B}_2$ denote the region borders in image space.

4 Experimental Results

We applied our method for segmenting skin in images. Images with different illumination conditions and background have been selected. For parameter l, used in Section 3.1 to determine the neighborhood size for pixel probability averaging, we chose the value 3. We observed no further smoothing improvement for higher values of l and a smaller value would mean no or an asymmetric neighborhood. The parameter k to determine the neighborhood size to calculate the edge distance image depends on l because at a pixel we need to know if an edge in the $l \times l$ neighborhood exists. Thus, we need $k \geq l$. The edge distance map is also used to calculate the stopping criterion. Because normally we do not find the region boundaries determined by color space clustering exactly at the edge pixels, we need some tolerance. Therefore, a higher value of k would be better. But the higher k, the higher the computation cost for the edge distance map. As a compromise we set $k = 5$. For the parameter δ used in Section 3.3 for stopping criterion, $\delta = 0.23$ seems to work best for out test images. Only for the last shown image, where the edges are very poor between the black hairs and the neck due to very dark skin, $\delta = 0.17$ works better.

To our knowledge, previously presented skin segmentation methods used a static skin model or other information e.g. face detection for initialization. Our approach only uses the information of a rough direction of skin relative to background. We compare our method with the well know approach [7], because both can be used as initialization for finer (skin) segmentation. We used the Matlab source code provided by [1]. They used the method from [7] for initialization of their algorithm. To make a fair comparison, we disabled the morphological filter. It is clear, that on both methods a morphological or other filters could be applied as post-processing step, but this is not content of this paper. Figure 5 shows some results obtained with [7] and our approach. The first three images shown have a resolution of 250×250. On a Athlon 64 X2 Dual machine the algorithm processed each of the images in about 0.5 seconds. The last two images are obtained from [1] and demonstrates the performance on dark skin. The examples show that we can obtain a better detection rate. False positives occur only as small regions. More results can be found soon at our web page[2]. In images in which skin can not be well approximated with a gaussian distribution, our algorithm will detect only a smaller part of the skin. If an image has a very uncommon background, for example a saturated red background our algorithm would have problems to identify the correct region.

5 Conclusions and Future Work

In this paper, we have proposed a new method for homogeneous color region segmentation in images. The method itself can be applied to any kind of homogeneous colored surface. In this paper we show its application and performance with skin detection. Our approach is based on a divisive hierarchical clustering in color space with spatial constraints that combines global color with local edge information. Robustness and accuracy are gained especially by using the input image itself to extract color distribution of the target region, and not a fixed distribution. Homogeneous color regions are modeled as 3D gaussians and parameters estimated by the EM algorithm. The cluster representing the target region, for example skin, is identified by comparing the mean value of each cluster with a vector obtained in a preprocessing step. For this comparison, the image color distribution is taken into account.

In the future, we plan to extend our method to model-based approaches, too. This can also be used to improve the stopping criterion of the subdivision algorithm. Furthermore, we want to extend the color distribution model to handle warped clusters.

References

1. L. Sigal, S.Sclaroff and V.Athitsos: Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. IEEE Conf. on Computer Vision and Pattern Recognition (2000)
2. J. Shi and J. Malik: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2000)

[2] http://cg.in.tu-clausthal.de/research/skinseg

3. S. Rital, H. Cherifi and S. Miguet: A segmentation algorithm for noisy images. Computer Analysis of Images and Patterns, 11th International Conference (2005)
4. T. Geraud, P.-Y. Strub and J. Darbon: Color image segmentation based on automatic morphological clustering. International Conference on Image Processing (2001) 70–73
5. C. Zhang and P. Wang: A New Method of Color Image Segmentation Based on Intensity and Hue Clustering. Proceedings of the International Conference on Pattern Recognition (2000) 3617
6. D.-W. Kim, K. Hyung Lee and D. Lee: A novel initialization scheme for the fuzzy c-means algorithm for color clustering. Pattern Recognition Letters **25** (2004) 227–237
7. M. J. Jones and J. M. Rehg: Statistical Color Models with Application to Skin Detection. International Journal of Computer Vision **46(1)** (2002) 81-96
8. A. Diplaros, T. Gevers and N.Vlassis: Skin detection using the EM algorithm with spatial constraints. IEEE International Conference on Systems, Man and Cybernetics (2004) 3071 - 3075
9. Q. Zhu, K.-T. Cheng, C.T. Wu, Y.L. Wu: Adaptive Learning of an Accurate Skin-Color Model. IEEE International Conference on Automatic Face and Gesture Recognition (2004)
10. D. Chai, A. Bouzerdoum: A Bayesian approach to skin color classification in YCbCr color space. Theme, Intelligent Systems and Technologies for the New Millennium (2000)
11. M. Wimmer, B. Radig: Adaptive Skin Color Classificator. International Conference on Graphics, Vision and Image Processing (2005)
12. J. A. Bilmes: A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report ICSI-TR-97-021