Automatic Segmentation of Clothing for the Identification of Fashion Trends Using K-Means Clustering

Abstract:

In this paper we propose a new method for the automatic segmentation of clothing, and its subsequent classification based on color, shape, texture and outfit complexity. The basic technique for segmentation leverages a combination of machine learning and computer vision algorithms. These algorithms include: Haar classification, skin and face detection, Canny edge detection, k-means clustering, lighting reduction and probabilistic modeling. While the automatic segmentation was successful with the majority of images from our training set, the breadth of images that the algorithm properly segments can be significantly increased by using reinforcement learning to individually adjust the tolerances and thresholds used in feature extraction. After segmenting the images, fashion trends were identified based on: color, texture (outfit complexity) and shape. While these trends represent true clusters of fashion, further research should be completed that follows a supervised learning model.

1. Introduction

21st century society has been inundated with a litany of text based recommendation services that leverage movie rental history and music preferences. However, consumers have few options in the area of automatically finding desirable clothing options. A prominent example, Like.com, features a method for doing visual search based on pattern, shape and color, however they constrain image uploads to: a single piece of clothing, not containing people and not having shadows or noise. Such a system does not reflect the current capabilities of machine learning. Thus we are interested in looking at segmentation of clothing images to construct an artificial intelligent-rich application. This application would be *Pandora*-esque in that it automatically recommends clothing items based on clothing that the user currently deems fashionable.

Motivation for Developing a Custom Algorithm

Image segmentation is a problem in computer vision that researchers have been exploring for over a decade (Mena, 2003). As such, significant developments have been made in the isolation of foreground and background. However, some gaps remain in the literature surrounding effective segmentation of individual clothing items in still imaging. Researchers have been able to differentiate clothing from skin (Gallagher, 2008) in order to extract clothing as a whole. Furthermore, work by Hu et al.(2008) demonstrated the ability to extract clothes from a person's torso. Despite the presence of these two capabilities, little research has been undertaken to segment individual pieces of clothing without the use of a precompiled database of clothing models. Borràs et al. (2003) developed a high-level description of clothing based on a combination of color-texture and structural design. However, this algorithm does not offer sufficient resolution of artifacts, nor proper separation of color versus texture. Instead of grouping different pieces of clothing, Borràs et al. grouped items based on their being plain or textured. This constrained form of clothing extraction is insufficient for our intended purpose of using these clothing features to establish an understanding of different fashion trends. Accordingly, we propose an algorithm that not only partitions individual from clothing, but also identifies individual pieces of clothing.

2. Data Description

2500 images were collected from two large US clothing retail stores. This dataset was reduced to 153 images in order to fit our constraints of an image of a single individual that displayed the face, torso and, at least, a portion of the leg. The resultant images contained individuals wearing, one to four pieces of clothing. 35 images were of men. The remaining 118 were of women. The types of clothes that individuals wore varied from formal to casual. Subjects in the pictures were from multiple ethnicities and had varied physical characteristics. Some of the images were full body, while others simply contained the thighs and above. All pictured individuals appeared on relatively monochromatic backgrounds. Close examination of the images indicated that several of the images had been altered. Finally, image sizes ranged between 6 and 12 kilobytes, with approximate dimensions of 167 x 204.

3. Research Method

Automatic Image Segmentation

Simple Segmentation Utilizing Global K-means Analysis

Global K-means was executed on each image using the Matlab Image Processing Toolkit. Following K-means segmentation, the algorithm used face and skin detection, in conjunction with segment pixel ranges to resolve the layer that contained clothing, background and skin.

K-means Color Analysis Process

Images were resized by a factor of 2 and converted to grayscale before being run through the K-means algorithm with a cluster size of 3. The K-means algorithm used the default settings of square Euclidean distance as the optimization objective, and initialized used randomly selected centroids. The segmented images were subjected to a face detection algorithm. The segment containing a face was determined to be the layer for skin. The two remaining layers, the background and the clothes, were differentiated by determining the range of the pixel locations for each segment. Clothing segments have a smaller range than background. This resulted in clearly labeled segments for the clothes, background and skin.

<u>Segmentation through Combinatorial Computer Vision and Machine Learning Algorithms</u>

This method uses knowledge of the input image criteria to optimize the segmentation. First, Canny edge detection is applied to the image, creating a separate image composed of the component boundaries. The most external boundary will typically correspond to the outline of the person because of image focus. Dilation is applied to the image to connect gaps in the external edge. A background mask is created by setting all pixels inside the external edge to white. The detected background is sampled to create a representative RGB histogram. Each pixel in the remaining image is compared to the background histogram. All high occurrence pixels are included in the background mask, a process known as back-projection. This step is done to eliminate background that is not discovered in the scanning step. Morphological operations are done to remove small pieces of clothing that may have been identified as background. The result is stored in a foreground image.

Next, the face is located using Haar classification. The pixels of the facial region are sampled and put into an H-S histogram for the skin, with the V component omitted to reduce the influence of lighting variations. Foreground image pixels are compared to the skin histogram, and pixels with high occurrences are marked as skin. Morphological operators are then applied to remove misclassified pixels

The background and skin masks are subtracted from the original image, leaving behind the clothes. The areas of all the items in the remaining mask are calculated, and removed if they are below a certain threshold. This step ensures that any noise due to the image subtraction is removed. Finally, the remaining items in the mask go through a polygon approximation algorithm. This eliminates jagged contours along the clothing. Finally, the images and masks are sent to the clustering algorithm.

<u>Segmentation through Image-Specific Combinatorial Computer Vision and Machine Learning Algorithms</u>

Here we identified 5 critical values that can be modified to improve the quality of an individual image's segmentation: 1. background probability threshold; 2. background histogram bin size; 3. skin probability threshold; 4. skin histogram bin size; 5. background sample size.

The probability thresholds are used in the back-projection step. Any pixel above a certain probability is grouped into that category. For example, if the Skin Probability Threshold is set to .90 then any pixels identified as having a 90% probability of being skin will be included in the mask. Changing these parameters is useful if the background or skin in the image has similar coloration as the clothing.

The histogram bin sizes are also used in the back-projection step and determine the resolution of the histogram comparison. A small bin size means that close intensity values (e.g. 245-255) will be grouped together as one color while a large bin size means a more accurate grouping. In essence changing the bin size allows you to vary the amount of data "smoothing" in the comparison. Changing these parameters is useful when you have small intensity variations in an image due to lighting, noise, etc.

The final value is the background sample size, which determines how much of the background is used to construct the histogram during back-projection. The larger the sample area is, the more accurate the background segmentation. The size is limited by the proportion of the image occupied by the subject. For the purposes of this project, a sample area of (image height)x(0.5*image width) was deemed ideal.

<u>Segmentation using a combination of our algorithms</u>

Using the combination of the two aforementioned methods enabled our team to effectively segment clothing tops from clothing bottoms. Thus each individual's clothing could be decomposed into its two primary components.

Clustering of Clothing Segments

To get an idea of clothing type, we extracted shape approximations. One of the visual cues for differentiating outfits is by their general shapes. For example, a shirt and pants set will have a near-rectangular shape while a gown will have a more triangular shape. To capture this information as a feature vector, we use the 7 Hu moments. Regardless of the orientation or size of the clothing, the moments allow the algorithm to make comparisons across different images.

We chose to extract the average standard deviation of the red, green, and blue channels as a way to represent the complexity of the outfit. Outfits with patterns and multiple-pieces had higher standard deviations than clothing with solid colors and fewer pieces. An extension of this feature would be to keep the R,G,B standard deviations as separate entries in the feature vector, allowing the algorithm to differentiate between different levels of complexity across the primary colors.

To allow differentiation based on the colors in the clothing, we used H and S histograms as a feature vector. The V component was omitted, since it is primarily affected by variations in lighting in an image. The H and S histograms each had bin sizes of 256, giving a total of 512 points. Using the entire histogram allows the algorithm to segment based on the occurrence of color combinations in an outfit, versus just discriminating based on primary color. Future work may involve using PCA to reduce the number of points, or using smaller bin sizes to reduce the number of points while also reducing the sensitivity of the histogram comparisons due to an "averaging" of color. Similarly clothing color between tops and bottoms was used as a feature vector to understand contrasts. These values were represented as R,G,B.

4. Results

Automatic Image Segmentation

Simple Segmenting Utilizing Global K-means Analysis

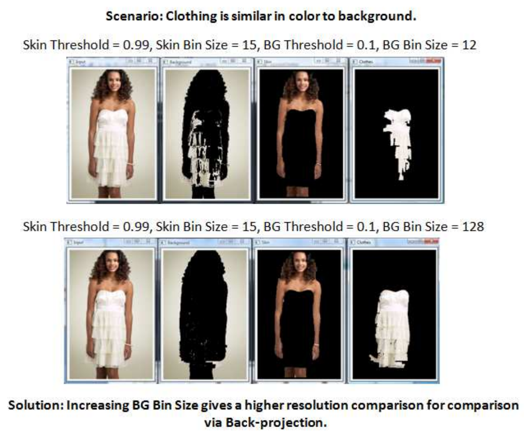
102 of the 153 images were segmented properly when using this method. A sample was deemed successfully segmented if the algorithm was capable of extracting a primary piece of clothing from the image (i.e. a top or bottom). Of the images that were not segmented properly, 16 contained black clothing and in 10 images the algorithm could not resolve the face. The remaining 25 images suffered from blending of clothing and background.

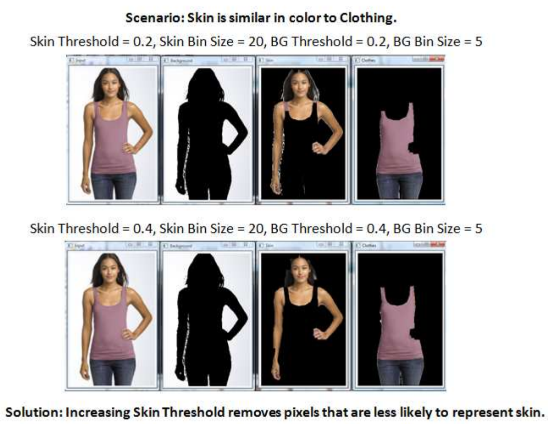
Segmentation through Generalized Combinatorial Computer Vision and Machine Learning Algorithms

75 of the 153 images were segmented properly when using this algorithm. An image was determined to be properly segmented if clothes were properly isolated.

Segmentation through Image-Specific Combinatorial Computer Vision and Machine Learning Algorithms

By adjusting one of the 4 critical features we were able to improve the quality of the segmentation.





Clustering Extracted Features (subset of results presented here for sake of brevity)

Single Feature Clustering

Results from cluster analysis based on standard deviation of clothing using k=2:

Figure 1- Low Complexity Clothing

Figure 2- Low Complexity Clothing

Results from cluster analysis based on standard deviation of color (histogram) using k=4:

Figure 3- Dark Shades

Figure 4- Blue

Figure 5- Red

Figure 6- Pink

Multiple Feature Clustering

Results from cluster analysis based on standard deviation of color (histogram) and color using k=4:

Figure 7 – High Complexity & Dark

Figure 8 – Low Complexity & Red

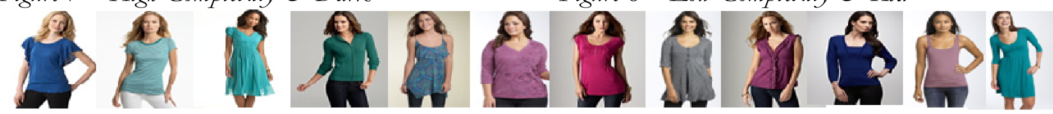


Figure 9- Low Complexity & Light Shades

5. Analysis

Efficacy of Automatic Image Segmentation

Global K-means

Using global K-means for determining clothing segments worked relatively well given its simplicity. For our dataset, this form of analysis yielded 67% accuracy. Furthermore, when coupled with a region detection algorithm, this method proved to be useful in providing reasonable results. However, as implemented, this method did not provide sufficient consistency, generalizability or precision to be used as our final method for automatic image segmentation. In terms of consistency, on more than one occasion a single sample produced no results the first time followed by good results the second time. Generalizability was in question because the algorithm did not apply well to certain colors, and certain color combinations. Precision was a concern because without additional modifications for executing connected components on the different regions of the image, it is difficult to consistently acquire whole pieces of clothing. Instead, patches of multicolor clothing would be erroneously be grouped with the background, for example.

Nonetheless, given its ease of use, global k-means provided significant benefit, and when coupled with our other algorithm, enabled for the segmentation of individual pieces of clothing.

Segmentation through Combinatorial Computer Vision and Machine Learning Algorithms

Using our learning algorithm we were able to successfully resolve clothing from nearly half of the images in our data set. However, once we adjusted the thresholds and bin sizes we were able to resolve clothing from all of the images in our training sample. Accordingly, we believe that by modifying our learning algorithm to automatically determine the best parameters, this process can be extended to include very large training sets.

Clustering Extracted Features

Through the use of K-means analysis on single and multiple features we grouped images that had similarly colored tops, and can also group together images that have similar color histograms. Similarly we can establish a grouping for images that have some level of complexity in them. This complexity can be in the form of differently colored tops and bottoms when taking the standard deviation of the histogram. Complexity can also be a proxy for highly textured clothes, or clothes with extensive designs. We see that through our analysis these types of items can be clustered together. Furthermore, the multiple feature analysis justified that by taking into account several features, we can group together clothing items that have similarities on several different levels.

Based on these clusterings we can make some initial conjectures about trends within the fashion industry. Firstly, by observing clustering based on color complexity, it is apparent that fashion tends to steer in the direction of simplicity. Complex designs are used in select cases, but the majority of the outfits are plain. Similarly, in our findings related to clothing shape, it was revealed that the majority of outfits (or at least models) tend to be more slim than wide. This may also be a case of certain types of outfits being better suited for certain body sizes, but that question is outside the scope of this project.

The result of the above feature clustering was the creation of reasonably meaningful groups. However, several feature clusterings that completed revealed results that don't have easily perceivable physical manifestation. Future work may venture to make more sense of these features.

6. Future Work

Refining the Image Segmentation Process

Though our combined algorithms were able to successfully resolve individual clothing items from the majority of the training samples, we recognize that there are actions that can be taken to improve this process. We have demonstrated that several of the samples which were not amenable to the generalized combinatorial approach could be solved by varying the parameters for the thresholds and bin sizes. Based on this finding we propose the use of reinforcement learning in order to train our system to automatically determine the appropriate parameters for segmenting each image. Determining these values would be based on a number of characteristics of the image. Furthermore, we surmise that completing a global k-means or Gaussian mixture model may help predict which images will undergo poor segmentation in our custom algorithm.

Extending Feature Extraction Capability

Additional work may also benefit from the inclusion of more extensive feature extraction. For example, we currently examine color of primary clothing objects. Extending our algorithm to also segment clothing accessories could offer novel information for training machines to recognize fashion.

Towards A Model for Learning Fashion

Finally, though we were able to identify small trends in society's conception of fashion, a more steadfast conclusion could be established using a supervised learning model. One way for creating such a model could be by taking segmented pieces of clothing, combining them at random and soliciting fashion experts to assess the fashionability of each pair. These training examples could then be analyzed using support vector machines to create a more concrete model of fashion.

7. Concluding Remarks

Through this study we have developed an algorithm for automatically segmenting clothing from still images. Using the features that we extracted from the clothes, we were able to draw some initial conclusions about fashion trends. Moreover, we have laid a framework for continued research in support of our desired end result: a machine that can make fashion recommendations based on automatic segmentation and meaningful clustering of clothing features.

8. Works Cited

Borràs Agnès, Tous Francesc, Lladós Josep, Vanrell Maria, "High-Level Clothes Description Based on Colour-Texture and Structural Features", 1st. Iberian Conference on Pattern Recognition and Image Analysis IbPRIA 2003.

Hu, Z., Yan, H., and Lin, X. Clothing segmentation using foreground and background estimation based on the constrained Delaunay triangulation. *Pattern Recogn.* 41, 5 May. 2008, Pages 1581-1592. DOI= http://dx.doi.org/10.1016/j.patcog.2007.10.005

Mena, J.B., Malpica, J.A., 2003. Color image segmentation using the Dempster sharer theory of evidence for the fusion of texture. Pattern Recognition. Internat. Arch. Photogrammet. Remote Sensing 34, Part 3/W8.