Automatic target image detection for morphing

Jaladhi Vyas

DA-IICT, Gandhinagar, India

Guided By
Manjunath V. Joshi

Table of content

Problem Statement

- **Problem Statement:** Given a human face image as a source, automatically identify the best target animal face image and to detect the features of both these images using the same approach so that entire morphing process can be done automatically.

Motivation

- Face morphing is widely use in making of animation movies
- Our approach solves the problem of manual selection of target image as it done by the researchers in morphing community.
- In recent animation movies, human facial features are added in cartoon face by using morphing between them to make the face more natural.

Proposed approach

- A block diagram of the proposed approach is shown below:

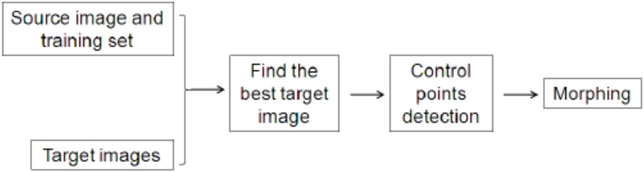


Figure 1 : Proposed approach.

Target detection [2-3]

We propose following two approaches for target detection:

- Gabor based approach for target detection.
- Target detection using textons and contrast.

Each of these approaches include of following three steps:

- Generating the texton dictionary.
- Histogram based model generation.
- Target detection.

Target detection

Gabor based approach I

- **Stage 1:** Generating the texton Dictionary
  1-Pass image through set Gabor filters.
  2-Extract 3×3 patch at every location of filter response images.
  3-Convert 3×3 patch into one dimensional vector and concatenate filter responses using set of 1-dimensional vectors
  4-Apply K-means clustering to generate the texton dictionary.
- **Stage 2:** Model generation
  Generate a model using histogram of textons frequency.
- **Stage 3:** Target detection
  Find a proper target image by using chi-square distance between two histograms.

Target detection
Gabor based approach II

Step 1: Generating the texton dictionary:

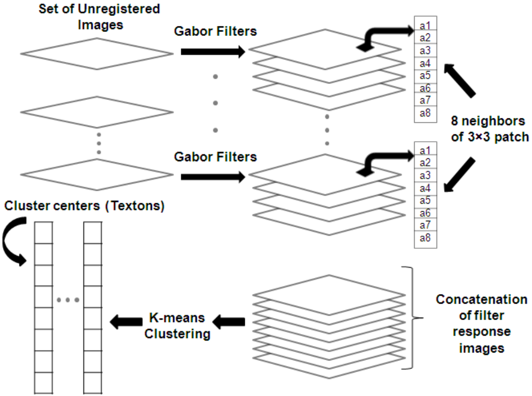


Figure 2 : Generating the texton dictionary.

Target detection
Gabor based approach III

Step 2: Histogram based model generation:

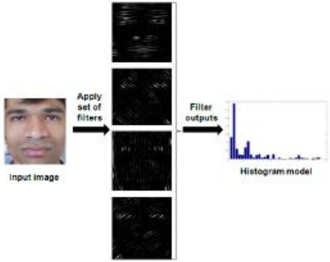


Figure 3 : Histogram based model generation.

Target detection
Gabor based approach IV

- Example of texton splitting is shown below.

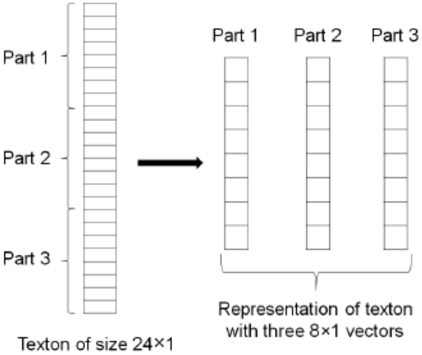


Figure 4 : Representation of textons using set of vectors to keep the size of texton and vector same.

Target detection

Gabor based approach V

Step 3: Target detection

- Best target image is found by using minimum chi-square distance between source and set of target image models.
- Equation to calculate chi-square distance between two histograms (h1 and h2) is shown below.

$$\chi^2(h_1, h_2) = \frac{1}{2}\sum_{n=1}^{bins}\frac{(h_1(n) - h_2(n))^2}{h_1(n) + h_2(n)} \tag{1}$$

Here *bins* corresponds to number of bins for textons. Note that for textons the number of bins = total number of textons in the texton dictionary.

Target detection

Based on textons and contrast I

- **Stage 1:** Generating the texton dictionary
  1-Extract 3×3 patch at every location of each image of a class.
  2-Convert each 3×3 patch excluding the center pixel into one dimensional vector of length 8×1.
  3-Sort these vectors either in ascending (or descending) order.
  4-Concatenate the vectors of all images of one class.
  5-Apply K-means clustering to generate the textons. Textons from all classes are combined to form a texton dictionary.
- **Stage 2:** Model generation
  Generate a model using joint histogram of textons and contrast.
- **Stage 3:** Target Detection
  Find a proper target image by chi-square distance between two histograms.

Target detection
Based on textons and contrast II

Step 1: Generating the texton dictionary:

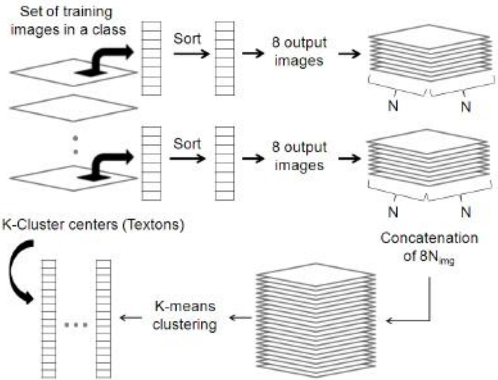


Figure 5 : Generating the texton dictionary.

Target detection
Based on textons and contrast III

- Input image and its corresponding eight output images are shown below.

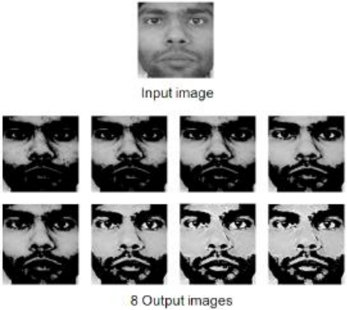


Figure 6 : Input image and its corresponding eight output images generated by sorting eight neighbors of every 3×3 patch.

Target detection

Based on textons and contrast IV

Step 2: Histogram based model generation:

- The first step to generate the histogram model of a given image, is to generate the contrast image from it. The procedure to generate the contrast image is as follows:
- Considering 3×3 image patch, calculate the average of the gray levels above the center pixel value. Let us call it A_a.
- Calculate the average of the gray levels below the center pixel value and call it as A_b. Then $A_a - A_b$ gives the contrast of that patch.
- This is repeated for all the patches considering every location of an image.

Target detection

Based on textons and contrast V

- One simple example to calculate the contrast is shown below. Here, in the end, the center pixel value is replaced by 11.25.

5	7	11
20	10	20
15	2	7

$$\mathbf{C}=\frac{20+20+15+11}{4}-\frac{5+7+2+7}{4}=16.5-5.25=11.25$$

Target detection

Based on textons and contrast VI

Let N_{bin} represents number of contrast values generated by binning the entire contrast set and N_{tex} represents total number of textons in the dictionary, then the model generation is as follows:

- For each of the vectors generated for the considered image, Euclidean distance between them and the textons in the dictionary is found.
- The count in the histogram with minimum distance is increased by one.
- The model generated using the above approach is the joint histogram of textons and contrast with $N_{tex} \times N_{bin}$ number of bins.

Target detection
Based on textons and contrast VII

- Figure shows an input image and its corresponding joint histogram model, where N_{tex}=20 and N_{bin}=13.

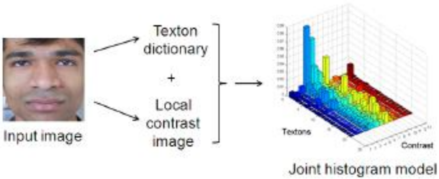


Figure 7 : 2D joint histogram model.

Target detection

Based on textons and contrast VIII

Step 3: Target detection

- Target detection step is same as explained in Gabor based approach only the model get changed from single dimension to two dimension.
- Thus, to compare two histogram models, chi-square measure with two variables is used.
- The equation of chi-square measure is shown below.

$$\chi^2(h_1, h_2) = \frac{1}{2} \sum_{n_1=1}^{N_{tex}} \sum_{n_2=1}^{N_{bin}} \frac{(h_1(n_1, n_2) - h_2(n_1, n_2))^2}{h_1(n_1, n_2) + h_2(n_1, n_2)} \qquad (2)$$

Target detection

Effect of contrast, scale and rotation on the histogram model I

Local contrast:

- The local contrast depends on gray scale values of an image patch and its range is in between [-255,255].
- The distribution of local contrast is not uniform. Majority of their values lie nearer to zero.
- However, the distribution of local contrast is not uniform in many of the images. Majority of their values lie nearer to zero.
- Because of this fixed length binning for obtaining the joint histogram will not take care of contrast variations.
- This motivate us to use variable length binning that spreads the joint histogram.

Target detection

Effect of contrast, scale and rotation on the histogram model II

- Histograms of local contrast using fixed and variable length binning are shown below:

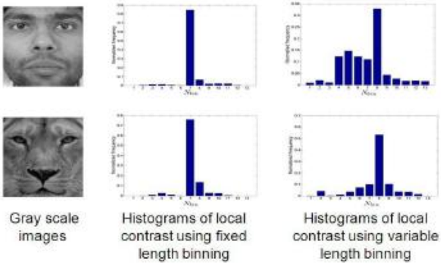


Figure 8 : Input gray scale images and their histograms of local contrast generated using fixed and variable length binning.

Target detection

Effect of contrast, scale and rotation on the histogram model III

Rotation invariance:

- In the proposed approach, a rotation invariant histogram model is obtained for a given image.
- The rotation invariance is due to the sorting of the vectors generated from every 3×3 patch of an image.
- Therefore, the histogram of textons frequency (generated by increasing the count of texton having smaller Euclidean distance to each of these vectors) is invariant of image rotation.

Target detection

Effect of contrast, scale and rotation on the histogram model IV

Rotation invariance:

- The local contrast which is the difference between average value of the gray levels above and below the center pixel value is also invariant of image rotation. Therefore, the joint histogram which is built using the histogram of texton frequency and local contrast is also an invariant of image rotation.
- The figure indicates that, the same model is obtained when the input image is rotated by an angle of 90% clockwise.

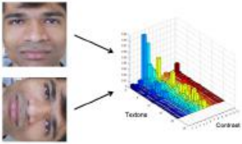


Figure 9 : Rotation invariant histogram model.

Target detection

Effect of contrast, scale and rotation on the histogram model V

Partial scale invariance:

- To deal with the effect of scale change, we use the extended training set by including same images with three different scales.
- These three scales correspond to: an original, down-sampled by a factor of two and up-sampled by a factor of two, respectively.
- Textons are generated using all these three scales of the training images.
- One may note that, when the image is up-sampled or down-sampled, the contrast may change due to interpolation. Thus the model is not totally scale invariant.

Target detection

Effect of contrast, scale and rotation on the histogram model VI

Partial scale invariance:

- Figure shows the expanded training set to achieve partial scale invariance

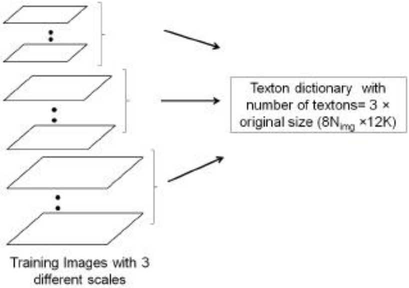


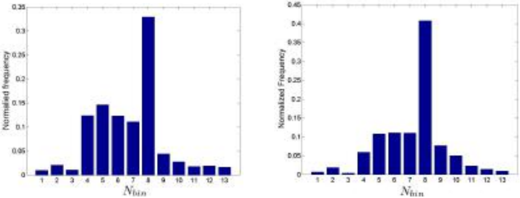
Figure 10 : Including images with different scales in training set to achieve partial scale invariance.

Target detection

Effect of contrast, scale and rotation on the histogram model VII

- Figure show histograms of local contrast for two different size of the same image.

Partial scale invariance:



(a) (b)

Figure 11 : Distribution of local contrast using variable length binning for (a) 200×200 and (b) 100×100 size images, respectively.

Experimental setup I

- To conduct the experiments, set of frontal face images are collected for four different classes with different illumination conditions.
- Our selected class of images are human, rhesus-monkey, cheetah and lion.
- Frontal face images are used because the dominant facial regions such as eyes, nose and lips can be more accurately detected compared to images with pose changes which may include external features e.g. hair along with the dominant regions.

Experimental setup II

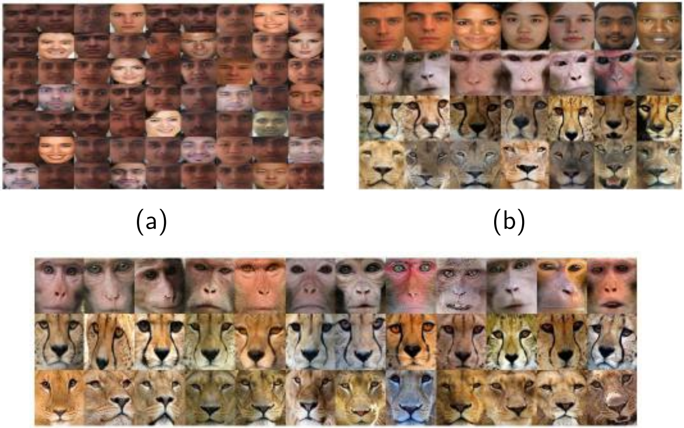


Figure 12 : (a) Source human face database. (b) Training set with four different classes. (c) Target animal face database.

Results of target detection I

- The example of source and target images with their histogram models with 5 textons per class i.e. K=5 is shown below.

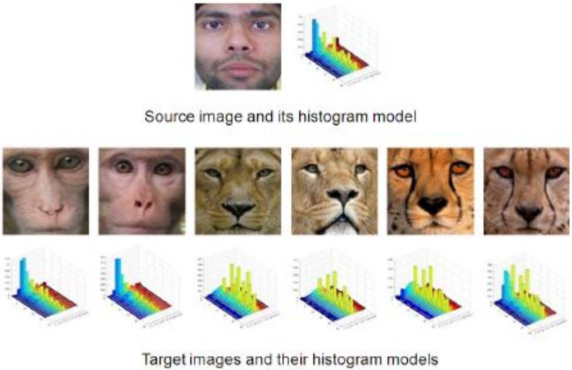


Figure 13 : Source and target images with their histogram models.

Results of target detection II

- Table shows the chi-square distance between source and three different target images for three different cases.
- First both source and target images are kept the same size, the target image is rotated by 45 degree, and lastly target image is down-sampled by a factor two.

Approach	**Chi-square distance**		
Original size (200×200)	0.1094	0.2428	0.2606
Rotation 45^o	0.1217	0.2528	0.2809
Down-sampled by a factor two	0.0814	0.2535	0.2890

Table 1 : Chi-square distance between source image and three different target images for three different cases.

Automatic control points detection I

- Once, the target image is detected, the next step is to detect the control points automatically for both source and target image.
- The common region of interest for human and animal faces is the eye.
- The eyes are detected using eye map operator and K-means clustering.

Automatic control points detection [4] II

Eye region extraction

- The block diagram of eye region extraction is shown below.

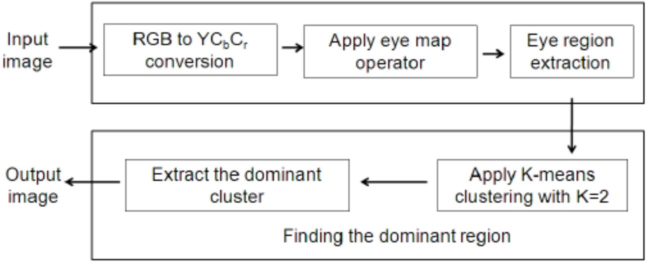


Figure 14 : Block diagram of eye region extraction.

Automatic control points detection III

Eye region extraction

- First the input RGB image is converted into YC_bC_r space.
- Two eye maps one for *luminance* component and other for *crominance* component are built.
- These two eye maps are later used to build a single eye map.

Automatic control points detection IV

Eye region extraction

- The eye map from *chrominance* and *luminance* component are given by

$$EyemapC = \frac{1}{3}\left\{(C_b^2) + (\overline{C_r})^2 + \frac{C_b}{C_r}\right\}, \qquad (3)$$

$$EyemapL = \frac{Y(x,y) \oplus g(x,y)}{Y(x,y) \ominus g(x,y) + 1}, \qquad (4)$$

- Here C_b^2,$(\overline{C_r})^2$,$\frac{C_b}{C_r}$ all lie in the range [0,1] and $\overline{C_r}$ is negative of C_r.
- *Y(x,y)* is a luminance component of a face region and *g(x,y)* is a structuring element
- $Eyemap = (EyemapC) AND (EyemapL)$.

Automatic control points detection V

Eye region extraction

- The eye map operator is used to extract the eye region. However, in many images some extra regions are also extracted which may have different intensity values.
- To remove these unwanted regions K-means clustering with K=2 is used. Which is applied on the output image generated using the eye map operator.
- The bigger cluster containing more number of pixels is then selected.

Figure 15 : Output of eye region extraction for human face image.

Automatic control points detection VI

- Once the eyes are detected, other control points are found using geometric distance relationship of different regions and invariant signature of face.
- Geometric distance relation of different regions for human and animal face images are shown below.

Distances	Human face	Animal faces
Eye-Nose tip	$0.60 \times D$	$0.65 \times D$
Eye-Lips	D	$1.30 \times D$
End points of lips	D	D

Table 2 : Geometric distance relationship between control points for human and animal face images with respect to the distance D between two eye control points.

Results of eye region extraction

- Figure show the results of eye region extraction for four different class of images.

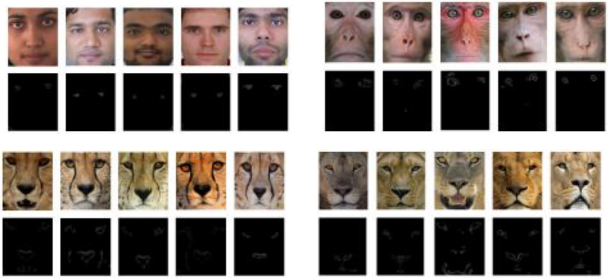
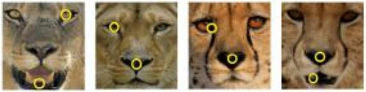


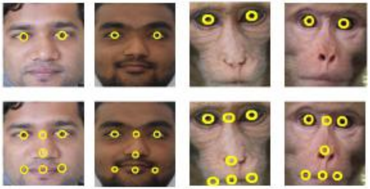
Figure 16 : Eye region extraction for different class of images.

Results of control points detection

- Results of control points detection for human and target animal face images are shown below:



(a)



(b)

Figure 17 : (a) Results of eye detection for animal face images except Rhesus-monkey. (b) Results of eye and other five control points detection for human and detected target face images.

Morphing [5]

- Once target image and control points for both source and target image are detected, the final step is the morphing.
- This is done by using triangulation mash based approach explained in [5].

Results of morphing I

- Results of morphing between human and three different target animal face images detected using the proposed approach.

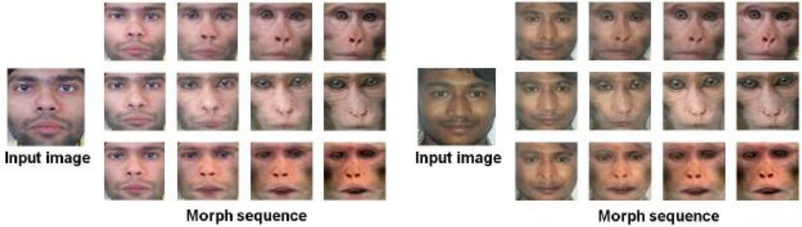


Figure 18 : Morphing results using automatic control points detection.

Comparision I

- The proposed approach for target detection is compared with JCBIR (Content Based Image Retrieval System using wavelet transform and K-means clustering developed in Java) which is MIT license software.
- In that, 63 human faces are used as query images and 36 animal face images are used as target images.
- The best three target images detected using JCBIR and our proposed approaches are same, however their retrieval ranks are different.

Comparision II

- Three target images detected using the proposed approach and JCBIR are shown below.

Figure 19 : Best three target images retrieved using JCBIR and the proposed approach.

Method	K=5	K=8	K=10
JCBIR	93.65%	93.65%	92.06%
Proposed approach	96.82%	96.82%	95.23%

Table 3 : Comparison of accuracy of perfect matching for two different methods with three different number of clusters (K) per class.

Comparision III

- To perceptually verify the results of target detection, SSIM between source and morphed images are found.
- An intermediate morphed image generated by morphing between human and rhesus-monkey face image is shown below.

Figure 20 : Source and target image with their intermediate morphed image.

Comparision V

- Table shows that SSIM between source and morphed target image detected using the proposed approach is high compare to target detected using JCBIR approach.

Source images	Target morphed images	
	Detected target	JCBIR detected target
1	0.4822	0.4384
2	0.4254	0.3857
3	0.5500	0.5063
4	0.5091	0.4705
5	0.5418	0.5081

Table 4 : SSIM between source and morphed images.

Conclusion

- Our proposed approach shows that given a human face image as a source and set of target animal face images best suitable animal face for morphing is rhesus monkey. We detect the common regions of both type of faces using same algorithm. Results of our proposed approach are verified using SSIM measure. One can easily see that the SSIM is high for the proposed method indicating that auto detected target animal face image maps well with the given source image.

Future work I

- The SIFT features of eye regions of human and rhesus-monkey face images are matches well.
- We are trying to find the matching accuracy of the eye regions of human and rest of the animal faces of the database using SIFT.
- If the matching accuracy is less for the animal faces except rhesus-monkey, then we can more strongly say about the similarity between human and rhesus-monkey face images.

Future work II

- The matching results of eye regions between human and rhesus-monkey face images are shown below.

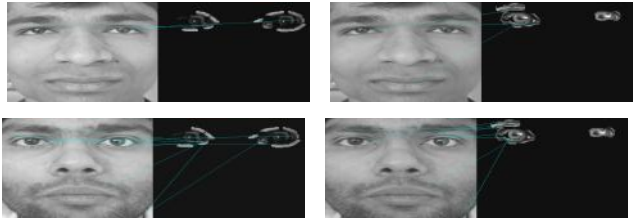


Figure 21 : Results of SIFT for matching the eye regions of human and rhesus monkey face images.

References I

[1] George Wolberg, *Skeleton-based image warping*. The Visual Computer, volume 5, 1989

[2] M Varma and A.Zisserman , *A Statistical Approach to Texture Classification from Single Images*. International Journal of Computer Vision, volume 62, 2005

[3] M Varma and A.Zisserman, *A statistical approach to material classification using image patch exemplars*.IEEE transactions on pattern analysis and machine intelligence, volume 31, 2009

[4] Rein-Lien Hsu and Mohamed Abdel-Mottaleb and Anil K. Jain, *Face Detection in Color Images*.IEEE transactions on pattern analysis and machine intelligence, volume 24, 2002

References II

[5] Stephen Karungaru and Minoru Fukumi and Norio Akamatsu, *Automatic human faces morphing using genetic algorithms based control points selection*. International Journal of Innovative Computing, Information and Control, volume 3, 2007

[6] Zhou Wang and Alan C. Bovik and Hamid R. Sheikh and Eero P. Simoncelli, *Image Quality Assessment: From Error Visibility to Structural Similarity*.IEEE Transection on Image Processing, volume 13, 2004

[7] "Google Images". *http://images.google.com/*

[8] "JCBIR". *https://code.google.com/p/jcbir/*