Automatic Age Estimation and Interactive Museum Exhibits

Kyle Patterson

March 20, 2014

Abstract

As a result of the advent of smaller and cheaper displays and speakers, museums have the ability to add an increasing amount of multimedia to their exhibits. Unfortunately, instead of being used in conjunction with information plaques to provide a deeper understanding of an artifact, these installations are generally used on their own. As a result they are often not an integral part of the overall exhibit and can easily be overlooked by visitors.

Through the use of video sensing equipment it is possible to create exhibits that can leverage multimedia to provide a more rich learning experience. By enabling museum exhibits to determine the approximate age of those viewing them, it is possible to provide additional interactivity that is tailored to meet their demographic and provide a more engaging experience.

Contents

1 Introduction

Traditionally, the information found in museums is located on small plaques near an artifact. This approach limits not only the information available to visitors, but also has an impact on the narrative possibilities open to curators. As technological advancements have led to both cheaper and thinner displays for digital media, some museums have begun to use them to provide supplemental audio and video for an exhibit. Unfortunately, current implementations are often displayed on their own, away from other artifacts because of the space constraints associated with creating viewing or listening areas that don't interfere with other artifacts. By combining multiple forms of media along with the artifact on display through small screens and speakers, it becomes possible to portray a deeper narrative about an individual artifact, as well as the exhibit as a whole to visitors than previously possible. Furthermore, this approach has the advantage of allowing museums to place more information within a smaller area opening up space for additional artifacts in an exhibit.

Through the use of interactivity with the viewer additional media as well as the content presented in an exhibit can be altered to best fit the situation. For example, the goal of some exhibits is not only to teach, but to also serve as a way to record the experiences or memories the viewer has on a particular subject. This can be done by asking the viewer questions related to the exhibit's topic. If the exhibit is on memories of Pearl Harbor, the relevance of certain questions can be directly related to the viewer's age. If the viewer appears to be older than 80 then asking what they remember about the event and the aftermath will result in a better response than if you asked someone who would not have been alive at the time. Likewise, asking the viewer what they learned about Pearl Harbor in school would result in a better answer from someone who is younger and would therefore not have had first hand experience. The benefit of this is that not only can an interactive exhibit create a more natural relationship with the viewer but, in the previous example, allows the museum to see how the memory of an event has shifted over time.

Another situation where altering presented content based on the viewer's age would be useful is to censor content inappropriate for a young audience. For example, an exhibit that examined war crimes in World War II would display different images based on the viewer's age.

This dynamic approach to tailoring content to the viewer can be achieved in multiple ways, both actively and passively. One way is to require the user to make some sort of input when they arrive at an exhibit. This could take the form of having the viewer input their age directly on a keypad or choose what age group they fit into. Another approach might involve having the exhibit audibly ask the viewer their age and listen for their vocal response. However, because these approaches require manual input from the viewer there may be circumstances where the requested action is either misinterpreted or where the user fails to make any input. The result of this is that the exhibit is either unable to make any changes or makes the wrong one due to either user error or malintent.

The use of computer vision allows an exhibit to shift for the user while avoiding the pitfalls of requiring their input. By using images to estimate the age of viewers as they approach the exhibit, changes to content can be made seamlessly to provide a better experience. While this approach may have cases of error in age estimation, because it is done without the input of the user, any changes to the content may be considered to simply be part of the original exhibit and not tailored to them.

2 Background and Related Work

Kwon and Lobo developed an approach to classify images into the categories of baby, young adult, and senior adult. The approach involved analyzing the ratios between facial features as well as skin wrinkles. The work done here however would have a large impact on later research as many later approaches would involve some combination of wrinkle detection and feature ratios. [8]

Horng, et al. follow a similar approach to that done by Kwon. A combination of facial features and wrinkles are used to classify the general age of the subject. Where this approach differs is in the focus on having a robust system that uses minimal resources. In order to minimize computational cost Sobel edge detection, which looks for sharp contrast in the gradient of an image to find an object's edges, was used alongside two neural networks, one to look at facial features to determine if the subject is an infant and the second to examine skin wrinkles. [6]

Bauckhage, et al. also use the detection of wrinkles in the classification process, but the manner in which this is achieved is unique. They extract features by dividing the facial image into a series of small grids which then have their local features detected. From these calculations features such as extreme local intensity indicating wrinkles and the gradients of these local features were used to classify the subjects age. [3]

Beyond using facial recognition techniques, analysis of a subject's gait can also be used to classify age. The work done by Davis sought to differentiate between the walking styles of adults and children. Using reflective markers to locate the head and feet of subjects as they walked, recordings were made of their stride. To differentiate between adults and children the amount of time between strides was calculated and the approximate number of strides/min. was used to discriminate between the age groups. [5] This approach has limited application outside of experimental environments because of the reliance on reflective markers to accurately track the head and legs of a subject.

The approach by Makihara, et al. used multiple video angles to classify male and female subjects as either children, adults, or elderly. Recordings of the subjects walking from multiple angles were normalized and silhouettes created to for classifications to be made on. Their system examined multiple factors including the stride frequency, posture, and relative size of body frame and head to differentiate between genders and the different age categories. [9]

In another work, Makihara, et al. sought to estimate age based on gait using video from in front and behind a walking subject. In order to estimate the subjects' age the researchers used Gaussian Process Regression, which is used to find a distribution of the gait feature from a normal distribution of ages, and then estimates the age based on the given classifying parameters and training data. The researchers tried using a series of different parameters with GPR and found that using the period of the subjects gate is only effective at differentiating between children and adults but not between different age groups of adults. [10] The work done by Makihara, et al. involving gait analysis presents a challenge for use in museums and other crowded environments. Because these approaches require the use of multiple cameras the volume of people present in a museum moving in different directions would make it difficult for accurate measurement of a single person's gait from different angles.

3 Approach

To solve the problem of altering an exhibit based upon the audience, we will need to estimate the viewer's age. This can be performed through either gate analysis or facial recognition. However, for the proposed application, facial recognition is a more appropriate approach. The reason for this is because the viewer's age does not have to be determined until they are at the point of interaction with the exhibit. Furthermore, choosing facial recognition over gait analysis frees the system from having to locate and analyze every person that walks within range of the system, regardless of whether or not they intend to approach and interact with the exhibit.

3.1 Method

In order to perform age estimation through facial recognition, the approach outlined in Figure 1 will be used.

1. The system will be divided into two parts that each take an input. The first part takes an input of training images with known ages, while the second part takes the image that we want to classify the age of as input.

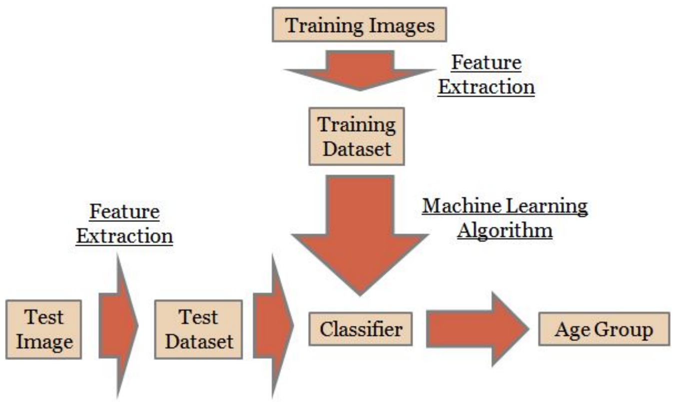

Figure 1: Model of Age Estimation Approach

2. Both sets of input are run through the same feature extraction algorithm to produce a set of usable data for each image.

3. The training images are run through a machine learning algorithm to produce a classifier for each age range. This is what determines which image features are most important in determining whether an unknown image resides in a particular age group.

4. The test image is compared to the classifier.

5. The estimated age group for the test image is the output.

3.2 Data Collection

1. In order to collect data the system must use video input or stills to provide images of the user.

2. Usable data must then be taken from the images using a feature extraction

algorithm.

3. This data must then be sent to a machine learning algorithm to be classified.

3.3 Principle Component Analysis

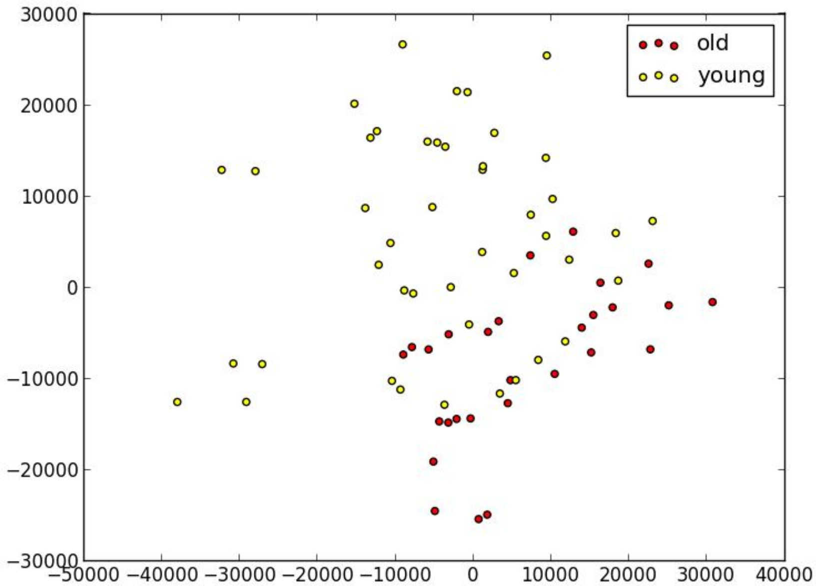

Figure 2: Graph of Vector Values for Images of Old and Young Males Over Two Dimensions

Principle Component Analysis, or PCA, is one of the procedures for extracting information from images that we will be using. PCA is an algorithm used in data extraction that is designed to reduce the complexity of that data to a manageable size.

This is done by first reducing the overall complexity of the data in an image from the total number of pixels, to a pre-defined number of dimensions, in the case of this implementation five. These components are created from correlations in the data that are most important to maintaining the overall integrity of the data contained in the image, while still reducing its complexity.

Figure 2 shows the output of PCA over two dimensions on a set of data containing images of old males, the red points, and young males, the yellow points. As you can see, the data representing each image has been reduced from the total number of pixels, to two vectors that compose the direction where the image has the most variance. Furthermore, the skewing of points representing young males towards the top of the graph, while the points representing old males are skewed towards the bottom illustrates how this reduction of data complexity can lead to differentiating between groups of images.

3.4 FisherFaces

Figure 3: Example Output of FisherFaces Detailing Class Features [1]

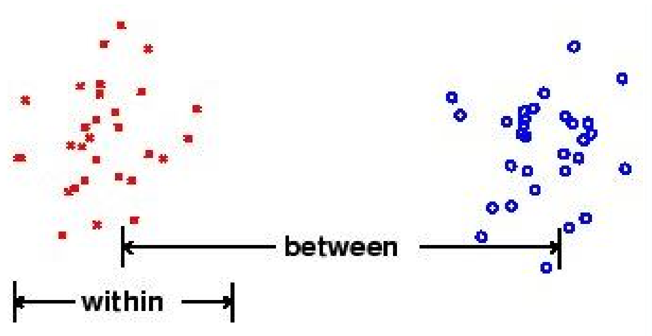

Figure 4: Sample Graph of Ideal Results from LDA [2]

FisherFaces is another procedure for extracting information from images that we will be using. FisherFaces is an algorithm designed to perform facial recognition between defined classes, in this case age group. Instead of only reducing the complexity of the data, like PCA, FisherFaces finds the facial features that differentiate between the defined classes through Linear Discriminant Analysis.. [4] Figure 3 shows an example image created by Fisherfaces. This image is a colormap of the features separating between two classes, which is gender in the case of this example. The more red an area is on the image, the more important that feature was in distinguishing between the classes.

The Linear Discriminant Analysis algorithm performed by FisherFaces performs in a similar way as PCA. However, after projecting the extracted data onto a set of axes, LDA then rotates the axes in order to minimize the distance between members of the same class, while maximizing the distance between classes. Figure 4 demonstrates an ideal graph created by LDA. Unlike Figure 2 created by PCA where there is a large amount of overlap between the two classes, this graph shows a very distinct separation between the two classes.

4 Evaluation

In order to test the efficacy of the data extraction approaches it is necessary for them to be tested on a set of data. Because gathering a large group of subjects from various age groups in order to test the system's accuracy was impractical, a database of facial images was used for training and testing instead.

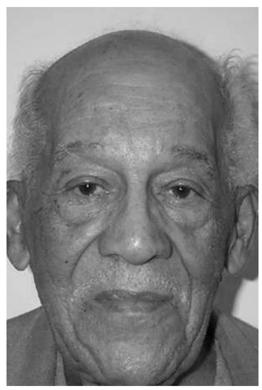

Figure 5: Sample Images from the Center for Vital Longevity Face Database. [7]

Training and testing is performed using facial images from the Center for Vital Longevity Face Database, an example of which is shown in Figure 5. [7] From this database, the subset of data we used contains 180 pictures of men and women between the ages of 18 and 94, providing a wide range of images to test. The set has an equal distribution between subjects in the age groups above and below 50 years old, as well as an equal distribution of men and women. Furthermore, this set contains only front facing facial grayscale images which have been cropped and normalized. The benefit of this is that the effects of lighting and skin tone differences can be mitigated, allowing the software to detect features of aging that are independent of these variables. Other facial databases were investigated for use in testing. However, many of these lacked either readily available documentation of subjects ages, contained less subjects over a smaller age span, or used images from angles that were not from straight forward.

Originally we aimed to differentiate between children and the three age groups of adults. However, no facial database containing images of children was readily available and creating one for this experiment was not a practical option. This does have some effect on museum applications, for example, using this system to censor graphic images in a museum exhibit from children is no longer possible. However, this does not remove all utility from the system.

The actual testing is carried out by running the results of the data extraction algorithm through a machine learning algorithm to generate a classifier for each age group against which the test images can be checked.

These tests are carried out using 10-fold cross-validation for the best results. This means that the entire set of data is first broken down into ten subsets. The first nine subsets will be used for training the machine learning algorithm to create a classifier and the final one is used for testing. This is then repeated ten times so that each subset is the test set at one time to avoid any anomalies caused by a particular combination of images. The results are then formed by combining the data from each of the ten tests.

The success of each series of tests will be evaluated on the accuracy of the systems performance. Because there is no base line accuracy that computer vision can achieve in this situation, the evaluation of an approach's success will be made in comparison to the results of other approaches. For example, if Fisherfaces produces results with

60 percent accuracy and using Principle Component Analysis produces results with 80 percent accuracy, PCA would be the standard by which future comparisons would need to be made.

5 Results

PCA Performed Over 5 Dimensions			
	Successful Classification Percentage		
Age Groups	Only Males	Only Females	Both Sexes
18 - 33	90.9	78.0	68.8
34 - 64	16.6	22.2	25.0
65 - 92	78.5	64.5	72.8
All Ages	72.2	62.2	60.2

Figure 6: Estimation Accuracy with PCA

The table in figure 6 shows the results for testing using Principle Component Analysis. The testing was performed with three different classifiers represented by the columns in the table. The first two classifiers were created from subsets of facial database composed of only the male and only the female images, respectively. The third classifier was created using both the male and female images.

5.1 Discussion

There a few points of interest in these results. The first is that the system performs best when only using male images, achieving the correct age estimation 72.2 percent of the time. Even more important is that the 18 - 33 year old age group was correctly estimated 90.9 percent of the time. These results are even more impressive when compared to the results achieved by humans in the experiment carried out by Horng, et. al. In their testing humans had an accuracy of 78.49 percent when classifying adults into three age groups. [6] Although our testing uses a different set of images it provides a general barometer for comparing our approach to human results.

Another important point that can be gathered from the data is that the overall accuracy of the system performs at its worst when the image data used includes both males and females. When testing with both genders the estimation accuracy across all ages drops by 10 percent for males and 2 percent for females.

A final important aspect of these results is the consistently low accuracy within the 34 - 64 age group, achieving success rates that are drastically lower than the other age categories within any test group at 16.6 percent for only males, 22.2 percent for only females, and 25.0 percent with both genders. One potential reason for these low results could be that the range of facial changes people tend to undergo within this age range prevent the machine learning algorithm from creating a classifier that satisfactorily represent the group. Furthermore, because of this, the images from either end of this age range likely correlate more closely with the classifiers of the other age groups.

6 Conclusion

The deployment of interactive museum exhibits has become more common as the technology associated with them has decreased. However, despite wider use, the methods of interactivity in these exhibits currently remain limited in scope. The goal of this research was to explore methods of age estimation that could be used by an interactive exhibit to cater content to the user's demographic without the requirement of user input.

This paper presented two systems, Principle Component Analysis and FisherFaces, for estimating the general age group of an individual based on facial features. Our system was run on a facial database containing 180 images from ages 18 to 94 and results were verified using 10-fold cross-validation.

The results achieved with the use of Principle Component Analysis show potential, despite the fact that no results have been achieved for FisherFaces against which to make a comparison. The accuracy of PCA when testing male faces, 72.2 percent, show promise for implementation in interactive museum exhibits after further research and refinement.

6.1 Future Work

With this approach to age estimation, there appear a number of avenues for future work. These avenues allow for both improvement an system accuracy as well as additional implementations for the system. One area of future work would be to test the same facial database with humans that is being used for our system testing.

This would provide both an additional metric to compare the accuracy of the system against, as well as a goal to reach a level of parity with human vision.

An area of future work that would provide an increase in the accuracy of the system based on current results would be a way to determine the gender of a user from their facial image. By first determining the users gender, their face could then be compared only to a classifier trained on the same gender. When testing with males this would provide a much higher level of accuracy compared to using a mixed gender classifier. Furthermore, this would allow exhibits to generate different interactive displays based on both gender and age, providing an increased level of personalization to the user.

Another area of future work that could have a positive impact on the accuracy of the system would be the inclusion of multimodal support while estimating age. For example, in addition to using facial images, the system could listen to the user as they speak and perform pitch analysis to supplement the age estimation result from images.

One example of an additional implementation would be a plugin for museum websites. The plugin would allow users to grant the website access to input from their webcam which could then be used for age estimation. This would allow museum websites to tailor the content of digital exhibits in the same way as those in physical museums. Furthermore, the fact that there is generally only a single user at a time on a computer can be used to the systems advantage. The entire website's collection could therefore be tailored to meet the users age group at the beginning of their visit, instead of having to recalculate their age every time they navigate between exhibits.

References

[1] Fisherface jet colormap. `http://docs.opencv.org/trunk/_images/fisherface_0.png`. Last Accessed: 2014-03-19.

[2] Lda separation. `http://www.dtreg.com/LdaSepGood.jpg`. Last Accessed: 2014-03-19.

[3] C. Bauckhage, A. Jahanbekam, and C. Thurau. Age Recognition in the Wild. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 392–395. IEEE, August 2010.

[4] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[5] James W Davis. Visual categorization of children and adult walking styles. In *Audio-and Video-Based Biometric Person Authentication*, pages 295–300. Springer, 2001.

[6] Wen-Bing Horng, Cheng-Ping Lee, and Chun-Wen Chen. Classification of age groups based on facial features. *Tamkang Journal of Science and Engineering*, 4(3):183–192, 2001.

[7] K. M. Kennedy, K. Hope, and N. Raz. Lifespan adult faces: Norms for age, familiarity, memorability, mood, and picture quality. In *Experimental Aging Research*, volume 35(2), pages 268 – 275, 2009.

[8] Young H. Kwon and N. da Vitoria Lobo. Age classification from facial images. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 762–767. IEEE, June 1994.

[9] Yasushi Makihara, Hidetoshi Mannami, and Yasushi Yagi. Gait analysis of gender and age using a large-scale multi-view gait database. In Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, *Computer Vision ACCV 2010*, volume 6493 of *Lecture Notes in Computer Science*, pages 440–451. Springer Berlin Heidelberg, 2011.

[10] Yasushi Makihara, Mayu Okumura, Haruyuki Iwama, and Yasushi Yagi. Gait-based age estimation using a whole-generation gait database. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6. IEEE, 2011.