Facial Expression Based Automatic Album Creation

Abhinav Dhall[1], Akshay Asthana[2] and Roland Goecke[3,1]

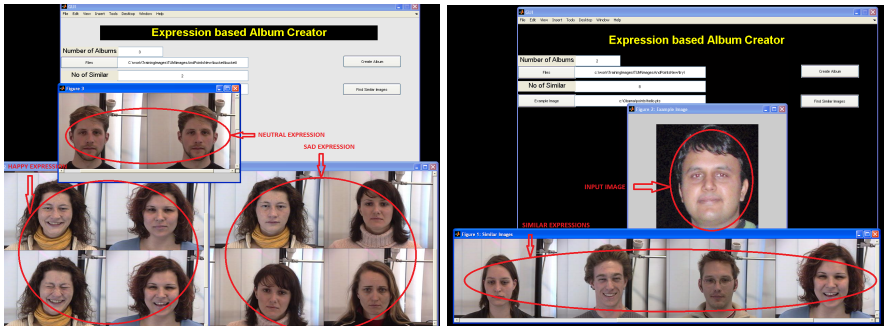
[1] School of Computer Science, CECS, Australian National University, Canberra, Australia
[2] School of Engineering, CECS, Australian National University, Canberra, Australia
[3] Vision & Sensing, Faculty of Information Sciences and Engineering, University of Canberra, Australia
Email: abhinav.dhall@anu.edu.au, aasthana@rsise.anu.edu.au, roland.goecke@ieee.org

Abstract. *With simple cost effective imaging solutions being widely available these days, there has been an enormous rise in the number of images consumers have been taking. Due to this increase, searching, browsing and managing images in multi-media systems has become more complex. One solution to this problem is to divide images into albums for meaningful and effective browsing. We propose a novel automated, expression driven image album creation for consumer image management systems. The system groups images with faces having similar expressions into albums. Facial expressions of the subjects are grouped into albums by the Structural Similarity Index measure, which is based on the theory on how easily the human visual system can extract the shape information of a scene. We also propose a search by similar expression, in which the user can create albums by providing example facial expression images. A qualitative analysis of the performance of the system is presented on the basis of a user study.*

Key words: *Automatic album creation, Facial expression analysis, Active Appearance Model, Structural Similarity Index, Image clustering.*

1 Introduction

With the advent of low-cost and easy to use consumer level imaging solutions, the number of consumer images has grown incredibly. With these increasing numbers, the management of images has become increasingly cumbersome. Classification of images into albums is a potential solution to this problem. Semantics based albums can be very helpful for effective browsing and retrieval. We propose a method for generating automatic emotion based image albums for better image management and representation. Facial expressions convey powerful discriminating information in facial images and hence form a strong criteria for image clustering. The user can group images 'based' on the emotions the faces in the image convey, such as 'happy', 'excited' or 'neutral' albums. The proposed



(a) Expression based albums (b) Album by similar expression

Fig. 1. Outputs of the system.

system uses *Active Appearance Models* (AAM) [2], which have been widely used for facial expression recognition and related applications in recent years. The *Structural Similarity Index Measure* (SSIM) [15] is used to compare similar facial expressions. The SSIM is based on the theory of the human brain noticing slight changes in the structure of a scene easily and fast. The user decides on the number of clusters/albums. New images are then added to the existing albums via facial expression comparison to the mean expression shapes of the albums. Another option for the user is to input an example image, which depicts a specific facial expression. The system then searches for images, which match the expression of the input image. We term this as 'album creation by similar expression'. In the experiment section, we present a user study on the performance of the facial expression based album creation.

1.1 Related Work

Of the manual/semi-automatic techniques, labelling has been used for long. However, as the image databases grow, managing labels becomes a complex and time consuming task. In [7], time stamping techniques are used to link photographs for effective browsing. In [3], the *Media Browser* exploits the metadata information in the images for tagging faces. Face detection and automatic labelling are used in the *FotoFile* system [9]. In [17], faces are detected and name labels are suggested based on a Gaussian framework to the user to choose from. In [11], the *AutoAlbum* uses time based clustering followed by a hidden Markov model based probabilistic approach for content-based clustering.

Recently, image editing and management tools, such as Google Picasa [6], have been used to manage and group images. The software uses robust face detection and groups all the images of one specific person together, which are then labelled by the user. In [1], face models based on AdaBoost are used to

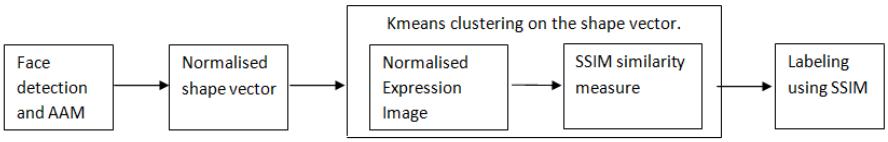


Fig. 2. Block diagram of the system.

extract facial features and semi-supervised clustering is used to group similar faces. [18] uses multiple representation spaces viz. faces, background and time of capture as input for mean-shift clustering.

Facial expression recognition is a well researched field. [5] present extensive surveys on facial expression recognition techniques. [10] use AAM for extracting facial features post fitting and machine learning techniques to classify emotions into FACS AU units [4].

Our contribution in this paper lies in creating albums from consumer images on the criteria of human expressions. Emotion based album creation is a useful feature for any image management system. Facial expressions are used to identify similarity among images. We assume that the majority of images contain faces. In existing systems, images of a specific person are grouped based on identity and labelled as an album. We want to explore the emotion/mood aspect of images, which can be a strong grouping criterion. The expression features can be used with the existing date, time and face criteria for album creation. For example a user may want to extract all the happy moments from a particular day's photographs or extracting the surprised expressions of people from an event.

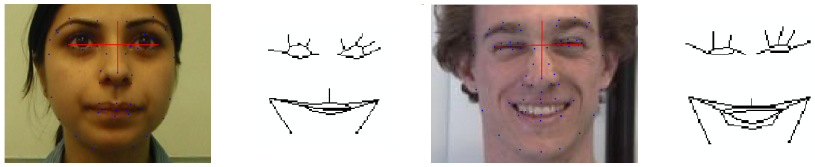
Figure 1 depicts outputs of album by clustering similar expressions. The details of the technique are discussed in depth in the following sections. The paper is divided as follows: Section 2 describes the system and its component, Section 3 shows experimental results and Section 4 provides the conclusions.

2 System

The system constitutes of four major steps, which are described in the following sub sections. Figure 2 depicts the flow of the system.

2.1 Facial feature extraction

The face is localised using the Viola Jones [13] face detector, that gives the location of the face. This is used as initialisation for AAM [2] tracking. The AAM are a powerful generative class of methods for modelling and registering non-rigid deformable objects. Their real benefit comes from its compact representation of appearance, which comprises of shape and texture, as well as its rapid fitting to unseen images. We used the AAM fitting method described in [12] for its speed and accuracy.



(a) Input (b) EI (c) Input (d) EI

Fig. 3. (a) and (c) are the input faces with red lines representing the landmark points, which are static with respect to local facial moment and donot contribute to the expression. (b) and (d) are the corresponding Expression Images (EI), which are compared for their structural similarity.

2.2 Expression image formation

The expression image is a visual map, which depicts the facial expression of the face. AAM fitting gives the shape vectors of the tracked faces, which constitute the landmark points. We then extract the eyes and mouth landmark positions from the shape vector. The new vectors obtained from the shape vectors of different images are then aligned to a common coordinate via translation, scaling and rotation for comparison.

The normalisation of the shape vector is performed by taking the horizontal Euclidean distance between the extreme end points of eye on the left and the right side. And the vertical distance is the Euclidean distance between the nose and upper eye brow. The choice for this normalisation is driven by the static nature of these points with respect to the expressions. The new EI is formed via drawing distance vectors among the new landmark points. The choice of specific landmark points and its corresponding distance vector image is derived from two motivations. One, choosing all points will bias the system towards similar faces. But our aim is different; we wish to find similar facial expressions rather than the images of the same person. Hence, a balanced number of landmark points, which represent enough information for representing the facial expression are chosen. The number and choice of landmark points was calculated with experimentation as on how much person independent the SSIM comparison can become. Two, SSIM works on images hence EI are created from the chosen landmark points. Figure 3 depicts two faces with red lines depicting the static nature of the points chosen for normalisation and their corresponding expression images.

2.3 Structural Similarity Index

We use Structural Similarity index (SSIM) [15] as the distance measure, it is a technique of calculating similarity among two images. SSIM is based on the theory that the human vision system is highly sensitive to changes in structure

of the view. Hence, a measure for calculating the structural information change can provide valuable information. In our system, SSIM is used as a distance metric of similarity among EI images. The SSIM metric between two windows w_1 and w_2 on the same size $N \times N$ is given by:

$$SSIM(w_1, w_2) = \frac{(2\mu_{w_1}\mu_{w_2} + c_1)(2\sigma_{w_1 w_2} + c_2)}{({\mu_{w_1}}^2 + {\mu_{w_2}}^2 + c_1)({\sigma_{w_1}}^2 + {\sigma_{w_2}}^2 + c_2)} \tag{1}$$

where μ_{w_1} and μ_{w_2} are the average of w_1 and w_2 respectively. $\sigma^2_{w_1}$ and $\sigma^2_{w_2}$ are the variance of w_1 and w_2 respectively. $\sigma_{w_1 w_2}$ is the covariance between w_1 and w_2. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are the variables to stabilise the division with weak denominator. L is the dynamic range of the pixel-values. We use this as a comparison of two EI.

2.4 Album creation and user options

Semi-unsupervised Album Creation For image album formation K-Means algorithm is calculated over the expression data. K-means clustering algorithm splits a set of observations into subsets by minimizing the intra-cluster variation. The numbers of image albums k serves as the initial number of clusters for K-Means clustering algorithm where the distance metric is SSIM. Therefore the clustering becomes:

$$\arg_{\mathbf{S}} \min \sum_{i=1}^{k} \sum_{x_j \in \mathbf{S}_i} \mathbf{SSIM}(\mathbf{x}_j, \mu_\mathbf{i}) \tag{2}$$

where $(x_1, x_2, .., x_n)$ are the Expression Images EI and μ_i in S_i is the mean EI. The clustering is done on the normalise landmark points of the shape vector, which are used to construct the distance vectors of EI, this is done to keep low dimensionality during clustering. Though the distance comparison is calculated on EI. The mean EI representing each clusters are then compared using the SSIM distance metric with pre-stored labeled EI. The pre-stored EI are labeled into four expressions (Happy, Neutral, Sad, and Excited). This leads to automatic labelling of the albums into the fundamental expression classes. Once a new image arrives it is added to the exiting albums via comparing its closeness to the mean image representing the respective albums.

Album by Example A user may be interested in finding images, which have the expression similar to a specific facial expression. In this case, the user provides the system with one example image. The user also specifies the number of similar images, which decides the size of this album. The system extracts the EI for the example and the group of images. The example EI is then compared using SSIM with all other images in the group. The similarity distances are then sorted and the user desired number of similar images is selected as an album with respect to the relevance. Figure 4 depicts two examples of this function.

(a) Input (b) Similar Expressions

(c) Input (d) Similar Expressions

Fig. 4. "Album by similar expression" example, (a) and (c) are the input images. (b) and (d) are the corresponding similar expressions.

3 Empirical experiment and outputs

Since different users may have different perception about an expression hence analysing the correct clustering performance is a non trivial task. To validate the performance we created a test set of sixty images from the FEEDTUM [14] and LFW [8] databases. A total of fifteen human users were asked to judge the album creation performance by figuring out the the images, which seem to have a different expression and do not belong to the album created. The average total error classification rate came out to be 13.7%. We also compare our system with fuzzy clustering algorithm. Figure 5 shows the outputs of the systems. The Figure 1 displays the experimental GUI of the system. In 1(a) the images are from the FEEDTUM database [14]. The three sub windows in the figure are the albums created after SSIM based clustering. Please note that the system groups faces of similar facial expressions into one set.

Figure 4, is the experiment on images from the LFW database [8]. Figure 4(a)is the user example input with a happy expression. Figure 4(b) is the album of top matching expressions with decreasing relevance from left to right. Similarly, Figure 4(c) is face with smiling expression and Figure 4(d) are the similar expressions. Figure 1(b) depicts album by similar expression example. The user inputs an image, which contains an example expression. The user also specifies the number of similar images desired in the album. This input serves as the number of similar images to be presented. The larger eclipse shows the searched similar images and the upper one is the user input image.

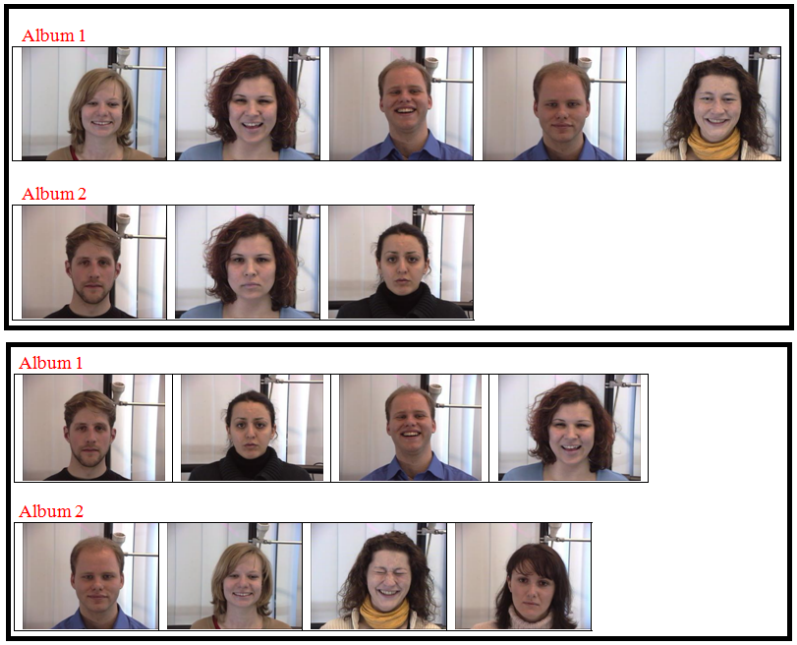


Fig. 5. Sample result on images from FEEDTUM [14] database after executing our system and Fuzzy clustering algorithm in the upper and lower box respectively.

4 Conclusions and future work

We propose a novel system, which can categorise images into albums on the basis of facial expression analysis, for effective image browsing and searching. It has applications in modern day image management systems such as Google Picasa [6] and Flickr [16]. The system uses AAM for facial feature extraction, a shape vector is extracted and normalised, and an EI is formed, which represents the facial expression of the image. Then, the SSIM is used as a distance metric for similarity, to cluster similar facial expression images together. The user also has the option to search for a particular image and form an album based on it ("creation by similar expression"). Future work is to add illumination invariance before AAM fitting, so as to have more robust fitting. Experimenting with a robust generic AAM tracker can also increase the performance of the system. Another potential area is exploring robust methods for unsupervised clustering.

References

1. W. Chu, Y. Lee, and J. Yu. Visual language model for face clustering in consumer photos. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, New York, NY, USA, 2009. ACM.

2. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. In *ECCV (2)*, 1998.
3. S. M. Drucker, C. Wong, A. Roseway, S. Glenner, and S. D. Mar. MediaBrowser: reclaiming the shoebox. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 433–436, New York, NY, USA, 2004. ACM.
4. P. Ekman and W.V. Friesen. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. In *Consulting Psychologists*, 1978.
5. B. Fasel and J. Luettin. Automatic Facial Expression Analysis: A Survey. *PR*, 36(1), 2003.
6. Google. Google Picasa. `http://picasa.google.com/`.
7. A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Extreme Temporal Photo Browsing. In *Visual Interfaces to Digital Libraries [JCDL 2002 Workshop]*. Springer-Verlag, 2002.
8. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report, 2007.
9. A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka. FotoFile: a consumer multimedia organization and retrieval system. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 1999. ACM.
10. S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. d. l. Torre, and J. Cohn. AAM Derived Face Representations for Robust Facial Action Recognition. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2006.
11. J. C. Platt. AutoAlbum: Clustering Digital Photographs using Probabilistic Model Merging. In *CBAIVL '00: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*. IEEE Computer Society, 2000.
12. J. Saragih and R. Göcke. Learning AAM fitting through simulation. *Pattern Recognition*, 42(11), 2009.
13. P. A. Viola and M. J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *CVPR (1)*, 2001.
14. F. Wallhoff. Facial Expressions and Emotion Database, 2006. `http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html`.
15. Z. Wang, A. C. Bovik, H. R. Sheikh, Student Member, E. P. Simoncelli, and Senior Member. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, pages 600–612, 2004.
16. Yahoo. Flickr. `http://www.flickr.com`.
17. L. Zhang, L. Chen, M. Li, and H. Zhang. Automated annotation of human faces in family albums. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003.
18. T. Zhang, J. Xiao, D. Wen, and X. Ding. Face based image navigation and search. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*. ACM, 2009.