Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence

Joint Inference for Extracting Text Descriptors from Triage Images of Mass Disaster Victims

Niyati Chhaya
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250
niyatic1@umbc.edu

The major contributions of this work include a set of biographical feature extractors brought together by a probabilistic graphical model. The graphical model acts as a communication medium between different feature extractors, together resulting in a text descriptor to describe triage images of disaster victims. The model is built using domain information gathered from data and literature.

Motivation

Earthquakes, hurricanes, terrorist attacks, and other such events can cause tremendous harm to people and infrastructure, often leaving those outside the areas impacted with little or no information on the state of their friends and relatives who may have been affected. Our goal is to automatically process images of patients taken as part of the intake process at emergency medical care centers to extract searchable, text-based descriptors of patients that can be accessed remotely (e.g., via the web) to facilitate identification of victims. Lost Person Finder (LPF) is a research project which aims at providing immediate post disaster communication information about disaster victims. As a part of the project, triage images of the victims are gathered and the aim is to extract a text descriptor that describes the person in the image. Achieving this goal requires finding the person in the image, and in particular their face, to drive the process of extracting features such as hair color and style, clothing color and style, the presence or absence of glasses, and other similar features.

Face detection for such non–standard images is a difficult task and was the motivation behind starting work in this domain. As we dealt with each feature detector one at a time, we realized that the different biographical features are dependent on each other. The target is to exploit this inter–dependency between individual feature detectors and improve the efficiency of the complete system.

Related Work

Extraction of biographical features has been studied to a limited extent by different computer vision groups. There is individual work in age detection (Kwon and da Vitoria Lobo April 1999), gender identification (Graf and Wichmann 2002), ethnicity recognition (Lu and Jain 2004), eyeglasses identification (Jiang, Binkert, and Achermann 2000), hair detection and hair style recognition (Yacoob and Davis 2006). Most of this work is limited and addressed to a particular set of images and tends to do poorly with disaster victim images. At the same time, there is no particular work that brings together all these features and tries to leverage relationships between them. The computer vision community has done some work in topics such as scene completion where information about different parts of the image is available, but this is a completely different problem than the one we are dealing with.

Approach

Feature extraction as introduced above needs to be preceded by first locating the person in the image, particularly the face. Face detection has been extensively studied for standard template–like face images, but for disaster victim images standard techniques do not perform well. In this work face detection (Chhaya and Oates 2011) is done as a pre–processing step. Individual feature extractors are studied and then the extractors link with each other into a probabilistic graphical model. The individual feature detectors give values for some biographical feature as a output. For example, gender detector output is about whether the person in the image is a male or a female. Inference algorithms applied over this model help come to an agreement between the different extractors and, on convergence of this algorithm, a text descriptor is built using output from different extractors, which gives identification information about the person in the image. Figure 1 shows an overview of this approach.

Current Status

As stated earlier, we initially developed an ensemble–based face detection algorithm that would perform well for images of mass disaster victims. This algorithm is a combination of skin detection, face detection and a template–based eye detection algorithm. Our dataset is a set of triage images of disaster victims, gathered during mass disaster drills conducted for Lost Person Finder. This algorithm gave a F–Score of 0.8 as against standard algorithms like Viola–Jones (Viola and Jones 2001), which gave a performance of 0.45. Experiments for each stage of this algorithm, i.e. skin detection, face detection and face detection with a combination

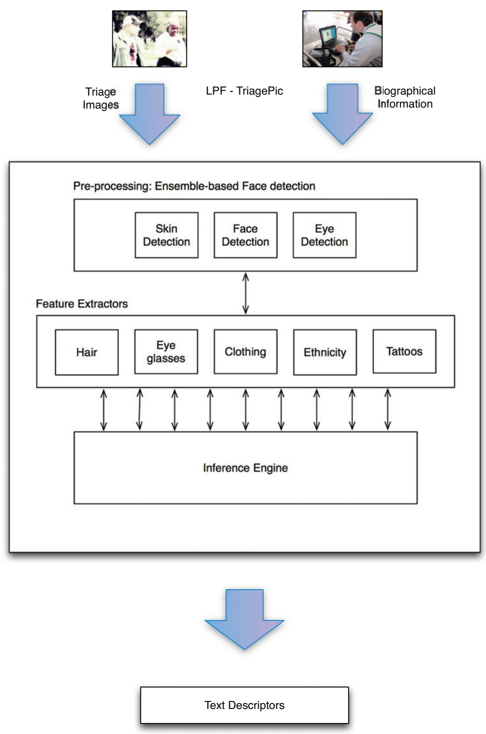

Figure 1: Project Architecture

of skin and eye detection, show that the ensemble performs best. The next step is addressing the various different feature extractors. So far, we have done work in eyeglasses detection, hair region detection using a combination of color and texture analysis, hair color and skin color identification and preliminary work in ethnicity identification. Preliminary experiments show these developed approaches perform better than existing algorithms for the given set of images.

The main target of this work is to bring these different extractors together. We designed an ad–hoc Markov net using information available in the literature. That is, information such as gender and facial hair tend to be related to each

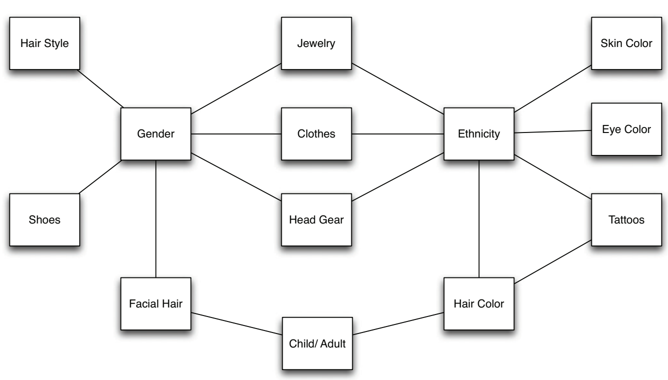

Figure 2: Markov Net

other and so do age and hair color. Figure 2 shows a diagram of this ad–hoc graph. We further developed a factor graph for this net. We are also working on gathering data to validate the different edges in the graph and to help define the domains of different nodes is a more complete manner using person datasets (Gallagher and Chen 2008) (Berg et al. 2004) and the Amazon mechanical turk.

Future Work

Having defined the graphical model, we are currently working on gathering data to validate the relationships. Another main task is to define the potential functions that communicate between the adjoining nodes and indicate whether the nodes agree with each other or not. The potential function will follow the energy minimization concept where the higher the value of the potential function, the smaller is the agreement between the nodes. Once the potential function is established and the different feature detectors are responding as needed, the next stage will be to have the individual feature detectors learn from the values of the potential function. If the potential function indicates that there is a disagreement with an adjoining node, the feature detector should adapt from information available and change its output if necessary. The final aim is to have the feature extractors agree with each other and the potential functions to converge at a minimum value. The output would be an ensemble of values from different extractors put together in the form of a text descriptor.

References

Berg, T. L.; Berg, A. C.; Edwards, J.; and Forsyth, D. A. 2004. Whos in the picture. In *NIPS*.

Chhaya, N., and Oates, T. 2011. Robust face detection for patient triage images. In *VISAPP 2011- Proceedings of the Sixth International Conference on Computer Vision Theory and Applications, to appear*. INSTICC Press.

Gallagher, A., and Chen, T. 2008. Clothing cosegmentation for recognizing people. In *Proceeding of CVPR*.

Graf, A. B. A., and Wichmann, F. A. 2002. Gender classification of human faces. In *Biologically Motivated Computer Vision 2002, LNCS 2525*, 491–501.

Jiang, X.; Binkert, M.; and Achermann, B. 2000. Towards detection of glasses in facial images. *Pattern Analysis Applications* 3:9–18.

Kwon, Y. H., and da Vitoria Lobo, N. April 1999. Age classification from facial images. *Computer Vision and Image Understanding* 74:1–21.

Lost person finder. http://archive.nlm.nih.gov/proj/lpf.php.

Lu, X., and Jain, A. K. 2004. Ethnicity identification from face images. In *SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, 114–123.

Viola, P., and Jones, M. 2001. Robust real-time object detection. In *International Journal of Computer Vision*.

Yacoob, Y., and Davis, L. S. 2006. Detection and analysis of hair. *IEEE Transactions Pattern Analysis Machine Intelligence* 28(7):1164–1169.