Surveillance, crime scenes, and mass disasters require us to understand the appearance of the persons involved. Biometric features such as fingerprints and face recognition have been successfully used to identify suspects and victims. Emergency situations such as mass disasters create a need to provide information about the affected people to their friends and families but biometric features and scene photographs cannot be broadcast to the media due to privacy concerns. Soft biometric features such as hair color and style, ethnicity, gender, age and clothing can be used to describe a person without revealing the identity and this information can be used to search and identify disaster victims. Extracting this appearance-related information using feature detectors on person images to form a text descriptor would avoid privacy concerns and allow searching for a person in a dataset using the text descriptor. A text-descriptor that resembles missing persons report would fulfill the requirement of providing person information while at the same time addressing privacy concerns.

Individual soft biometric features have been studied to a limited extent with the aim of person identification. Some of these soft biometric features, e.g. facial hair and gender, are correlated. Our work deals with images of mass disaster victims taken during triage at emergency medical care centers. Traditional approaches for soft biometric feature extraction fail for these real world triage images due to their cluttered graphic nature (presence of injuries, blood, debris, bandages and disfigurement). In this work, we present a Markov network to exploit correlated soft biometric features and relate individual feature detectors with each other to build a coherent set of feature detectors. Our work is to individually extract features from triage images of disaster victims, check correlation and coherence between the various features using joint inference over the probabilistic graphical model and use this information to get a set of feature values that do not have a conflict within their values. For example, for the features facial hair and gender, if facial hair is present when the subject is a female, there is a conflict. Here, the graphical model indicates a relationship between these two features and inference is used to indicate the conflict and eventually converge when the feature detectors reach a coherent set of values (here facial hair absent, gender female or facial hair present, gender male).

We designed a set of potential functions at the node and the edge levels to represent the correlation and the agreement criteria. The functions consist of three principal terms: agreement potential (interaction potential, edge level, binary), self potential (node level, unary) and the change potential (to force convergence). The first two terms are defined for the graphical model. Each pair of nodes is associated with an agreement table that defines values for which the feature value pairs agree or disagree. The agreement tables are defined based on how people describe other people and also from available information in the anthropology literature. The node potentials are defined on the basis of whether the feature detector has a valid output. It defines how correct and confident the feature detector is about its own value. The third term, which is the change penalty, is used during the inference process. This term is used to force convergence of the algorithm if the nodes take long to converge. The change penalty penalizes the graphical model configuration to reduce changes in feature values and come to a consensus regarding feature nodes.
Using this setup, we implement the Markov network for two feature nodes and show that the resultant text descriptor has higher accuracy when inference is used against when inference is not used. We get a statistically significant result for two different datasets: standard - FRGC image dataset (p=0.006) and triage image data (p=0.052).

The utility of this work is for people to use when looking for others and hence to understand what values of soft biometric features different people use and how people are described we conducted an Amazon’s mechanical Turk survey where the task was to describe people shown in an image. This information is useful while building the graphical structure, i.e., defining which features are related and which ones are not, and also in understanding how people describe other people. The annotations through this survey show that people tend to refer to the same features while describing a person but use different vocabulary. For example, ‘Male, 9-18 months, Caucasian’ and ‘White, toddler, boy’ were two annotations for the same person. We are currently looking at methods to cluster these different annotations and understand the domain space for various features.

To summarize, in this work we present a set of soft biometric features combined together using a probabilistic graphical model to extract text descriptors from triage images of disaster victims and a Amazon mechanical Turk survey to understand how people describe other people to support the text descriptor.